Deep learning-empowered wavefront shaping in scattering media

Luo, Yunqi

2021

Luo, Y. (2021). Deep learning-empowered wavefront shaping in scattering media. Doctoral thesis, Nanyang Technological University, Singapore. <https://hdl.handle.net/10356/147459>

<https://hdl.handle.net/10356/147459>

<https://doi.org/10.32657/10356/147459>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 08:49:11 SGT

**DEEP LEARNING-EMPOWERED WAVEFRONT SHAPING
IN SCATTERING MEDIA**

LUO YUNQI

School of Electrical & Electronic Engineering

2021

DEEP LEARNING-EMPOWERED WAVEFRONT SHAPING IN SCATTERING MEDIA

LUO YUNQI

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

02/04/2021

.....
Date

Luo Yunqi

.....
Luo Yunqi

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

Apr. 02, 2021

.....
Date

.....
Zheng Yuanjin

Authorship Attribution Statement

This thesis contains material from 1 paper published in the following peer-reviewed journal, 1 paper accepted by the following peer-reviewed journal, 1 paper published in arXiv, and 1 paper accepted at conferences in which I am listed as the first author.

Chapter 3 is published as Luo, Y., Yan, S., Li, H., Lai, P., and Zheng, Y. Focusing light through scattering media by reinforced hybrid algorithms. *APL photonics*, 5(1), 016109 (2020). DOI: 10.1063/1.5131181.

The contributions of the co-authors are as follows:

- Prof. Yuanjin Zheng and Prof. Puxiang Lai supervised and coordinated the whole work, and revised the manuscripts.
- I did the experiments, processed the data, and prepared the manuscript.
- Dr. Suxia Yan and Mr. Huanhao Li helped with the experiments and manuscript revision.

Chapter 4 is published as Luo, Y., Yan, S., Li, H., Lai, P., and Zheng, Y. (2019). Towards smart optical focusing: Deep learning-empowered wavefront shaping in nonstationary scattering media. arXiv preprint arXiv:1909.00210.

The contributions of the co-authors are as follows:

- Prof. Yuanjin Zheng and Prof. Puxiang Lai supervised and coordinated the whole work, and revised the manuscripts.
- I did both the simulation and experiments, processed the data, and wrote the manuscript.
- Dr. Suxia Yan helped in conducting experiments and revising manuscript.

- Mr. Huanhao Li helped in simulation and manuscript revision.

Section 4.3.1 is published as Luo, Y., Li, H., Zhang, R., Lai, P., and Zheng, Y. Deep learning assisted optical wavefront shaping in disordered medium. In Adaptive Optics and Wavefront Control for Biological Systems V (Vol. 10886, p. 1088612). International Society for Optics and Photonics (2019, February).

The contributions of the co-authors are as follows:

- Prof. Yuanjin Zheng and Prof. Puxiang Lai came up with the idea, supervised and coordinated the whole work.
- I did simulation and wrote the manuscript.
- Mr. Huanhao Li helped with the simulation.
- Ms. Ruochong Zhang helped in revising manuscript.

Chapter 6 is accepted by *Photonics Research* as Luo, Y., Yan, S., Li, H., Lai, P., and Zheng, Y. (2020). Towards smart optical focusing: Deep learning-empowered dynamic wavefront shaping through nonstationary scattering media.

The contributions of the co-authors are as follows:

- Prof. Yuanjin Zheng and Prof. Puxiang Lai supervised and coordinated the whole work, and revised the manuscripts.
- I did both the simulation and experiments, processed the data, and wrote the manuscript.
- Dr. Suxia Yan helped in conducting experiments.
- Mr. Huanhao Li helped in simulation and manuscript revision.

02/04/2021

.....
Date

.....
Luo Yunqi

Acknowledgement

Throughout my four-year Ph.D. study and thesis preparation, I have experienced many struggles, but I also gained growth. I would like to appreciate a lot of people that have offered me support and help.

First of all, I would like to express my sincere thanks to my supervisor Assoc. Prof. Yuanjin Zheng. His insightful guidance has helped me a lot during the transformation from a learner to an active researcher. His research experience is invaluable in developing research topics and methodology.

I would like to acknowledge my collaborators: Asst. Prof. Puxiang Lai and Dr. Huanhao Li from The Hong Kong Polytechnic University. I am grateful for their valuable advice on my research work.

I would like to thank my seniors: Dr. Fei Gao, Dr. Xiaohua Feng, Dr. Ran Ding, Dr. Suxia Yan, Dr. Haoran Jin, and Dr. Weitao Song. I appreciate their kind and patient assistance. Many thanks to my group members: Ms. Ruochong Zhang, Mr. Siyu Liu, Mr. Chuanshi Yang, Mr. Zesheng Zheng, and Mr. Zhengyuan Zhang. They have offered me a lot of help and support.

I would like to thank all of my dearest friends. I deeply appreciate their warmest encouragement and accompany when I feel depressed.

Many thanks to technicians in VIRTUS and VALENS 2: Ms. Lim-Tan Gek Eng, Ms. Seow-Guee Geok Lian, Mr. David Robert Neubronner, and Mr. Lim Teng Kheng, Desmond. I am grateful for their kind help in resolving technical problems.

I appreciate Nanyang Technological University for providing me the research scholarship and research environment. Many thanks to reviewers for their valuable time to go through this thesis.

Last but not least, I would like to give my most sincere thanks to my parents.
Their selfless love and support always inspire me to become better.

Table of Contents

Statement of Originality	III
Supervisor Declaration Statement.....	IV
Authorship Attribution Statement.....	V
Acknowledgement	VII
Summary.....	XII
List of figures.....	XIV
List of tables.....	XX
Chapter 1. Introduction	1
1.1. Background	1
1.2. Motivation and Objectives	2
1.3. Major Contributions	3
1.4. Organization of the thesis.....	4
Chapter 2. Controlling Multiple Scattering in Disordered Media with Wavefront Shaping: A Review	6
2.1. Multiple scattering theory	6
2.2. Properties of optical speckles	7
2.3. Focusing light through scattering media with wavefront shaping	8
2.3.1. Optical wavefront shaping.....	10
2.3.2. Guidestar-assisted wavefront shaping	11
2.3.3. Deep learning-assisted wavefront shaping and image recovery.....	13
2.3.4. Key parameters	14
2.4. Applications of wavefront shaping techniques	15
2.4.1. Photoacoustic imaging.....	15
2.4.2. Optical imaging	15
2.4.3. Endoscopic imaging.....	16
2.4.4. Optical manipulation	17
2.5. Memory effect.....	18
2.6. Summary	18
Chapter 3. Focusing Light Through Scattering Media by Reinforced Hybrid Algorithms	21
3.1. Working Principle	21

3.1.1. Wavefront shaping using genetic algorithm	21
3.1.2. Wavefront shaping using deep neural networks	23
3.2. Experimental results and comparisons of algorithms	26
3.3. Discussion	33
3.4. Summary	40
Chapter 4. Deep Learning-Assisted Wavefront Shaping through Disordered Media with Perturbations	42
4.1. Theoretical analysis of implementing deep convolutional neural networks for light focusing and refocusing through scattering media with perturbations	42
4.1.1. Inverse scattering problems and deep convolutional neural networks	42
4.1.2. Speckle correlation theory	45
4.2. Working Principle	46
4.3. Simulation and experimental results	50
4.3.1. Simulation results	50
4.3.2. Experimental results	54
4.4. Discussion	59
4.5. Summary	65
Chapter 5. Deep Learning-Empowered Wavefront Shaping through Changing Media	67
5.1. Theoretical analysis of implementing deep learning framework for light focusing and refocusing through nonstatic scattering media	67
5.1.1. Inverse scattering problems in time-variant environment	67
5.1.2. ConvLSTM layer	70
5.2. Working principle	72
5.3. Experimental results	75
5.4. Discussion	82
5.5. Summary	88
Chapter 6. Deep Learning-Empowered Adaptive Recursive Framework for Dynamic Wavefront Shaping through Nonstationary Scattering Media.....	90
6.1. Working principle	91
6.2. Simulation and experimental results	93
6.2.1. Simulation results	93
6.2.2. Experimental results	103
6.2.3. Comparison of light focusing and refocusing performance with TFOTNet,	

conventional ConvLSTM, and CNN	109
6.2.4. Comparison of time cost in focusing recovery with various algorithms ...	113
6.3. Summary	116
Chapter 7. Conclusion and Future Work	118
7.1. Conclusion.....	118
7.2. Future work	120
Author's Publications	122
Bibliography	124
Appendix I	131
Appendix II	133

Summary

Optical focusing through or within disordered media is of great significance yet challenging in a lot of scenarios including biomedical imaging, optical communication, and astronomy due to the inevitable scattering. Wavefront shaping has emerged as a widely accepted approach to solve this problem. By pre-compensating the optical wavefronts before the light enters the disordered medium, wavefront distortions induced by scattering can be counteracted, leading to the realization of optical focusing. In addition, time reversal and phase conjugation of scattered light serves as another method to obtain a focal point. In recent years, wavefront shaping techniques witnessed rapid growth in terms of algorithms, efficiency, and applications, *etc.*

Dynamic control of wavefronts and light focusing through randomly changing scattering media may bring revolutionary advancements to many fields such as biomedical imaging. However, so far, most implementations to pre-compensate the optical wavefronts have only operated with static media due to the requirements of iterative optimizations or measurement of the transmission matrix, which are time-consuming. With the goal to comprehensively resolve wavefront shaping problems through nonstationary scattering media, this Ph.D. thesis comprehensively investigates the fundamental physics of scattering and inverse scattering in disordered media. Deep learning frameworks are developed based on the mathematical models, and light focusing and focusing recovery through scattering media with perturbations, media that are continually changing at constant speeds, and randomly altering media are all achieved.

In the introduction, background and motivations of the research work are described. Then state-of-art wavefront shaping techniques, including their working principles,

advantages, drawbacks, and applications are reviewed in Chapter 2. In Chapter 3, the hybrid of complementary algorithms is proposed as a reinforced approach, and it can lead to significant improvements in the optimization speed and performance. In Chapter 4, a deep learning framework implemented by a specifically designed Convolutional Neural Network is employed to deal with light focusing and focusing recovery problems through the scattering medium suffering from environmental perturbations, taking advantage of the fine-tuning engine of a neural network and the speckle correlation theory. Afterwards, in Chapter 5, a more sophisticated deep learning framework implemented by a multi-input-single-output Convolutional Long Short-Term Memory network is adopted to tackle the light focusing and refocusing problems through a scattering medium that is changing continually at a constant speed, not just limited to environmental disturbances. The aim of Chapter 6 is to resolve the ultimate challenge of wavefront shaping through a randomly altering medium. With the adaptive adjustments of hyperparameters and recursive fine-tuning system, spatiotemporal non-stationarity can be handled efficiently by the proposed deep learning framework, permitting fast light refocusing through the nonstationary medium.

In conclusion, covering the physics fundamentals, algorithm developments, simulation, and experiments verifications, this Ph.D. thesis comprehensively investigates wavefront shaping problems through nonstatic media, taking a large step towards practical applications.

List of figures

Fig 2.1. (a) A coherent light beam with unmodulated wavefront undergoes scattering when it passes through a disordered medium, and speckles are observed outside the medium. (b) A coherent light beam with modulated wavefront is focused to a point after passing through a disordered medium.....	9
Fig 3.1. Working principle of employing genetic algorithm for wavefront shaping. First of all, a specific number G of initial phase patterns are created, then they are ranked according to their fitness function scoring results. Two patterns are selected as parents to generate offspring. Several segments of the offspring are mutated (indicated by white circles), then the offspring are evaluated by the fitness function. Each generation $G/2$ offspring are reproduced to replace existing $G/2$ patterns, then all phase patterns are ranked again based on their fitness function scores. The whole breeding and mutating process is iterated multiple times until the end condition has been met, typically, a pre-defined number of generations are reproduced, or the fitness function evaluation result reaches a threshold.....	23
Fig 3.2. Working principle of applying deep neural networks for wavefront shaping. First of all, a DCNN is trained with a number of samples. DCNN inputs are speckles recorded by a camera and outputs are the corresponding incident phase patterns modulated by a SLM. After training, the DCNN can accurately map speckles to their corresponding phase patterns. Afterwards, a preset focused speckle is sent to the DCNN as the desired pattern, and the phase pattern predicted by DCNN is able to focus light through the medium when the pattern is loaded on the SLM.	24
Fig 3.3. Two methods to apply DCNN results to GA in Genetic Neural Network (GeneNN). (a) GeneNNv1. The phase pattern predicted by the DCNN is one of the initial phase patterns in GA. This pattern should have the highest ranking, so it is of great possibility of being chosen to breed the next generation. (b) GeneNNv2. The SLM pattern from DCNN works as the common base, and initial phase patterns in GA are created by augmenting various additional patterns to it. .	26
Fig 3.4. Experiment setup. Light emitted from the laser passes through a ND filter (F), then light is expanded and collimated by a telescope (L1 and L2). A half-wave plate (HWP) and a polarizer (P) are used together to adjust the polarization state of light. Light is modulated and reflected by the SLM. The following two lenses (L3 and L4) reduce the beam size so that light can enter an objective lens (OBJ1). OBJ1 focuses light on the surface of the diffuser (D), and another objective lens (OBJ2) placed behind the diffuser collects scattered light and projects it onto the camera.....	27
Fig 3.5. The structure of the proposed deep convolutional neural network (DCNN). Input 1 is speckle patterns, and their corresponding SLM patterns are noted as Input 2. Input 3 is the desired speckle, and DCNN output is its required SLM pattern corresponding to Input 3. With Input 1, Input 2, the first three convolutional layers (Conv1, Conv2, and Cov3), and the first fully-connected layer (FC1), the DCNN is supposed to learn the information of scattering processes in the medium. The last fully-connected layer (FC2) is trained to establish the mapping from speckles to SLM patterns.	29
Fig 3.6. Experiment results of GeneNNv1, GeneNNv2, randomly initialized GA, and DCNN. (a) A focused speckle obtained using the SLM pattern predicted by the pre-trained DCNN. (b) Focusing result with GeneNNv2, while the population size G is 50. (a) and (b) use the same colormap and scale. (c)-(g) Intensity enhancement factor η achieved by the three algorithms: GeneNNv1, GeneNNv2, and randomly initialized GA when population size G is set as 10, 20, 30, 40, 50, respectively. (h) Different numbers of samples are used for DCNN training, resulting in different η after light focusing. The increase percentage in η over the result achieved when	

training sample size is 10,000 is a function of training sample size. (i) To increase η to 60, the time spent by three algorithms (GeneNNv1, GeneNNv2, and randomly initialized GA) are different at different population size. (j) To increase η to 50, the time spent by three algorithms (GeneNNv1, GeneNNv2, and DCNN) are different at different population size. (k) With the same time spent $t=500$ s, the η achieved by different algorithms (GeneNNv1, GeneNNv2, and randomly initialized GA) are different at different population size. (l) With the same time spent $t=500$ s, the η achieved by different algorithms (GeneNNv1, GeneNNv2, and DCNN) are different at different population size.	31
Fig 3.7. GeneNN optimization results with different mutation rate R using GeneNNv1 and GeneNNv2.	37
Fig 3.8. The focusing performance of GeneNNv1 and GeneNNv2 with various initial η achieved by DCNN at different population size G . (a) The enhancement factor η achieved by GeneNNv1 when $G=10$ with different initial η offered by DCNN. The initial η realized by DCNN are 42 (grey line), 37 (red line), 32 (blue line), 26 (green line), and 19 (purple line), respectively. (b) The enhancement factor η achieved by GeneNNv2 when $G=10$ with different initial η . (c) The enhancement factor η achieved by GeneNNv1 when $G=20$ with different initial η . (d) The enhancement factor η achieved by GeneNNv2 when $G=20$ with different initial η . (e) The enhancement factor η achieved by GeneNNv1 when $G=30$ with different initial η . (f) The enhancement factor η achieved by GeneNNv2 when $G=30$ with different initial η	38
Fig 4.1. A brief illustration of light scattering. The phase patterns of light are modulated by a spatial light modulator (SLM), then light is focused onto a scattering layer (S) by an objective lens (OB1). Scattered light transmitted the scattering layer is collected by another objective lens (OB2) and recorded by a camera.	44
Fig 4.2. (a) Structure of the proposed RFOTNet. RFOTNet has 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 targeted speckle pattern, while RFOTNet output is the SLM pattern needed in order to get Input 3 through the scattering medium. (b) Illustration of the working principle of the proposed method. First of all, samples are collected to train RFOTNet. After training, the RFOTNet can model the inverse scattering function, and output a SLM pattern that can focus light through the scattering medium. When the medium is altered due to perturbations, the focal point can be degraded or even lost. Benefiting from the correlation of speckles before and after the change, only a small number of extra samples are sufficient to fine-tune the pre-trained RFOTNet in step 3. After the tuning, the RFOTNet can adapt to the concurrent medium state and recover the optical focus.	49
Fig 4.3. Simulation results. (a) Focused speckle pattern obtained in Step 2. (b) Speckle pattern after the medium change. (a) and (b) are recorded with the identical modulation pattern on the SLM. (c) The focused speckle pattern can be recovered after fine-tuning the pre-trained RFOTNet. (a), (b), and (c) use the same colormap. (d) Different levels of variances/perturbations result in different levels of scattering medium change. (e) The relationship between the degree of medium change and the number of required fine-tuning samples to recover the focusing. .	51
Fig 4.4. The structure of a conventional single-input-single-output CNN. Input is speckle patterns and output is corresponding SLM patterns. The structure and parameters of all the convolutional layers and fully-connected layers in the CNN are the same as those in RFOTNet.	53
Fig 4.5. Simulation results of RFOTNet and a single-input-single-output CNN in light focusing and refocusing. (a) Training performance of RFOTNet and CNN. Straight lines indicate the results of RFOTNet, while dashed lines show the results of CNN. Black and red lines show the	

training loss values (mean-squared-error) and validation loss values (mean-squared-error), respectively. (b) The performance of RFOTNet and CNN during fine-tuning. (c) Focused speckle obtained using RFOTNet after training. (d) Focusing recovery results with fine-tuned RFOTNet after medium change. (e) Focused speckle obtained using CNN after training. (f) Focusing recovery results with fine-tuned CNN after medium change. Note that (c)-(f) use the same color bar. 54

Fig 4.6. Schematic of the experimental setup. Light is expanded by two lenses (L1 and L2), and then half-waveplate (HW) and polarizer (P) adjust the polarization state of light incident onto the spatial light modulator (SLM). Light is modulated and reflected by the SLM, then passes through two lenses (L3 and L4) and is focused onto a diffuser (D) surface by an objective lens (OB1). Scattered light is collected by another objective lens (OB2) and recorded by a camera. 55

Fig 4.7. Experimental results of four different groups to achieve focusing recovery through the diffuser. (a)-(c) Experimental results of Group 1. (a) is the original focal pattern obtained using the RFOTNet trained in Step 1, (b) is the pattern after the perturbation, and (c) is the recovered focal pattern after fine-tuning the RFOTNet. (d)-(f), (g)-(i), and (j)-(l) are the speckle patterns corresponding to before change, after change, and after RFOTNet fine-tuning for Groups 2 to 4, respectively. (a) - (l) use the same colormap and scale. (m) The target speckle. (n) Intensity enhancement factor η of the focus of speckle patterns. Black squares indicate the η after Step 2, red circles indicate the η after the medium change, and blue triangles indicate the η of recovered focus after Step 4. 57

Fig 4.8. Experimental results of focusing recovery with memory effect. (a) Original focused speckle obtained using the RFOTNet. (b) Speckle pattern after moving the diffuser left by $1\mu\text{m}$ while the SLM pattern is the same as (a). (c) Speckle pattern after the SLM shifted left accordingly. (d) Correlation between the original focus (a) and speckles after medium change (with the same SLM pattern) is a function of translation distance. (e) Focal peak intensity as a function of diffuser translation distance. Figs. a-c use the same colormap and scalar bar. 61

Fig 4.9. The relationship between the intensity enhancement factor η remaining ratio and the levels of scattering medium change. The original focal point was the same, and the diffuser was moved by various distances to introduce different levels of perturbations. All results are scaled as the original η to be 1. Red dashed line illustrates the change of median values while blue dotted line shows the mean values. 62

Fig 4.10. The focusing recovery performance, as measured by the relative η , as a function of the number of samples in fine-tuning and the degree of medium change. 63

Fig 4.11. The intensity enhancement factor η achieved and training time cost as the number of training samples varies. 64

Fig 5.1. A brief illustration of light scattering through time-variant media. The phase patterns of light are modulated by a spatial light modulator (SLM), then light is focused onto a nonstatic scattering layer (S) by an objective lens (OB1). Scattered light transmitted the scattering layer is collected by another objective lens (OB2) and recorded by a camera. The time-variant scattering medium is created by a rotating diffuser. 68

Fig 5.2. Illustration of the proposed deep learning-empowered 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 passing through the scattering medium in experiments 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 a SLM pattern that can obtain the target through the

current medium. Inevitable environmental disturbance (disturbance) or the nonstationary 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. After tuning, the TFOTNet can adapt to the concurrent medium state and recover the optical focusing performance. 74

Fig 5.3. Schematic of the experimental setup. Light is expanded by two lenses (L1 and L2), then a half-waveplate (HWP) and a polarizer (P) adjust the polarization state of light incident onto the spatial light modulator (SLM). Light is modulated and reflected by the SLM, then passes through two lenses (L3 and L4) and is focused onto the surface of a mechanically rotating diffuser (D) by an objective lens (OB1). Scattered light is collected by another objective lens (OB2) and recorded by a camera. 77

Fig 5.4. Experimental results. (a) Light focusing results through continually changing media (Group 1-4) with SLM patterns precited by a pre-trained TFOTNet (upper row) and a pre-trained CNN (bottom row). Medium changing speed in Group 1-4 is listed in Table 5.1. Light focusing performance is measured by PBR. (b) Light focusing results with trained (upper row) and fine-tuned (bottom row) TFOTNet and CNN under three circumstances (Groups 5-7). (c) Left image: the original focus obtained through a continually changing medium. Middle image: The degraded speckle pattern after the diffuser is suddenly rotated by a large angle while the SLM pattern is the same as the left one. Right image: The recovered focal pattern after fine-tuning the TFOTNet. During the fine-tuning phase, the scattering medium keeps changing at the same speed before the disturbance happened. All figures in (a) – (c) use the same colormap and scale. (d) The targeted focused speckle pattern. 79

Fig 5.5. Experimental results of light focusing recovery using TFOTNet through a nonstatic medium under various conditions. (a) A focused speckle obtained through a continually changing medium. (b) and (c) The degraded speckle pattern recorded after different time while the SLM pattern is the same as (a). Speckle correlation with (a) are 0.7547 and 0.5521, respectively. (d) Light focusing recovery performance in the six scenarios with different medium changing speeds, fine-tuning initialization conditions, and fine-tuning samples. 82

Fig 5.6. Experimental results of applying GA to focus light through media that are changing continually at various speeds. (a) – (c) Light focusing results with GA. From (a) to (c), the diffuser rotating speed is increased accordingly. (a) - (c) are recorded at the position where the whole optimization process ends. The black dotted circle indicates the position of target focused speckle. (d) – (f) Light focusing results when the diffuser is rotated back to the position where the experiment starts. (a) and (d), (b) and (e), (c) and (f), are recorded with the same SLM pattern, respectively. Figs. a-f use the same colormap and scale. (g) The PBR achieved by GA under the three conditions. PBR of focused speckles obtained at positions that experiment ends (gray square) and starts (red dot) are evaluated. 85

Fig 5.7. Experimental results of applying CSA to focus light through media changing continually at various speeds. (a) – (c) Light focusing results under three conditions. From (a) to (c), the rotating speed is increased accordingly. (a) - (c) are recorded at the position where the whole optimization process ends. The black dotted circle indicates the target position of the focused speckle. (d) Focused speckle pattern recorded at the position where the experiment starts. (b) and (d) are recorded with the same SLM pattern. Figs. a-d use the same colormap and scale. (e) The PBR achieved by CSA under the three conditions. The PBR at positions that the experiment ends (gray square) and starts (red dot) are evaluated. 86

Fig 5.8. Experimental results with targets of different sizes. (a) – (c) Target focused speckles generated in Matlab. From (a) to (c), the diameter of the focused speckle keeps increasing. (a) -

(c) use the same colormap. (d) – (f) Experimental results. Focused speckles obtained with pre-trained TFOTNet using (a) (b), and (c) as the target, respectively. (d) - (f) use the same colormap and scale.	88
Fig 6.1. Flow chart of the proposed adaptive recursive algorithm for light focusing and refocusing through nonstationary media.	93
Fig 6.2. The influence of hyperparameters on fine-tuning time cost to reach a pre-defined PBR when a 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 that the scattering medium is changing at different speeds (indicated by lines of different colors, and quantified by speckle decorrelation time).	95
Fig 6.3. (a) The required number of fine-tuning samples with and without adaptive adjustments of hyperparameters as the medium changes at different speeds (characterized by SDT). (b) - (c) The relationship between PBR after fine-tuning and the fine-tuning samples amount using the adaptive algorithm as the medium changes at different speeds.	97
Fig 6.4. Fine-tuning results with ten random nonstationary processes using three different algorithms. One nonstationary process can be regarded as consisting of multiple piecewise stationary stochastic sub-processes, while the SDT and time duration of each sub-process are randomly chosen. (a) - (j) Fine-tuning results with the adaptive recursive algorithm (grey line), nonadaptive recursive algorithm (red line), and traditional fine-tuning algorithm (blue line) in the ten nonstationary process. Each process is characterized by SUM_M . Figs. (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. Figs. (k) and (l) use the same legend.	102
Fig 6.5. 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 the adaptive recursive algorithm and traditional algorithm. Figs. (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.	106
Fig 6.6. (a) - (c) Results of the three experiments without environmental disturbance. The SDT of each stationary sub-process is shown in the figure, and the PBR target is indicated by yellow dashed lines. (d) - (f) Results of the three experiments with environmental disturbance. Figs. (a) - (f) use the same legend. (g) Speckle images recorded during a nonstationary process (Fig. f) using the adaptive recursive and traditional algorithm. All speckle images use the same colormap and scale, and are interpolated to 253×253 for a better view. The color bar indicates the detected light intensity in arbitrary unit. The middle image in the bottom row (traditional algorithm) is an interpolated result using recorded speckles.	108
Fig 6.7. Comparisons about fine-tuning ability in a nonstationary process using three different networks. (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) are shown. All images use the same colormap and scale, and the color bars indicate the light intensity in arbitrary unit. (b) Experimental results. Light focusing and refocusing results using three networks with 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 unit. (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 unit. (d) Details of the nonstationary process and PBR with the three networks in simulation (Fig. a). (e) Details of the nonstationary process and PBR with the three networks in experiments (Fig. b). (f) Details of the nonstationary process and PBR with the three networks in experiment (Fig. c). These three figures (f-h) use the same legend.	110
Fig 6.8. (a) Structure of a conventional single-input-single-output ConvLSTM network. (b) Structure of a conventional single-input-single-output CNN.	111

List of tables

Table 3.1. Increase percentage in the intensity enhancement factor η comparing with randomly initialized GA.....	31
Table 3.2. The number of required measurements to reach different η with varying population size G and hybrid algorithms.	35
Table 3.3. The drop percentage of the enhancement η after the first measurement of GA against initial η offered by the DCNN with different population size G , hybrid algorithms, and initial η	40
Table 5.1. Summary of experiment scenarios.	77
Table 5.2. Summary of six fine-tuning scenarios.	82
Table 6.1. SDT dependent PBR target.	97

Chapter 1. Introduction

First of all, this chapter briefly discusses the background of the research work, including the significance as well as obstacles of light focusing and imaging through or inside disordered media. Then the motivation and objectives of the research work are introduced. Finally, the major contributions are stated, and the organization of the thesis is provided.

1.1. Background

Light focusing within or through disordered media is of great significance in a lot of applications ranging from optical communications to biomedical imaging, but it has been considered challenging for decades due to the inevitable light scattering induced by the mismatch of refractive index [1]. Traditional wisdoms in the field have assumed that image information is carried only by unscattered ballistic or quasi-ballistic light, whose intensity decays exponentially in a heterogeneous medium [2]. After propagating more than one transport mean free path (~ 0.1 mm for human skin), light will be totally scrambled, as a consequence, information carried by light will be lost. If light is coherent, scattered light along different optical paths interferes randomly, forming 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 as speckle correlation time) [3]. Inspired by this property, various approaches have been reported recently to realize optical focusing and imaging through scattering media, such as time reversal [1, 4-9], wavefront pre-compensation [3, 10-14], and memory effect [1, 15-18]. Time reversal methods, such as TRUE [4] and TROVE [6], 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. By contrast, pre-compensated wavefront shaping techniques modulate the phases of light before light enters into the scattering medium, based on the measurement of transmission matrix [11, 13, 14] or the maximization of feedback provided by the optical [10] or photoacoustic signal strength [3], with a goal to pre-compensate for the scattering-induced phase distortions. Representative algorithms to obtain the optimal compensation include continuous sequential algorithm (CSA) [19], genetic algorithm (GA) [20], and phase-conjugation of the transmission matrix [11]. As for the memory effect method, image information is encoded in the autocorrelation of the measured speckles as long as the imaging area is within the memory effect range, then images can be reconstructed from speckles using iterative phase retrieval algorithms [1, 16]. Some neural networks were recently introduced to focus light through scattering media, for example, [21]. These impressive breakthroughs open up new pathways to tackle inverse scattering problems, demonstrating great potentials in facilitating practical implementations that require light control such as deep tissue imaging.

1.2. Motivation and Objectives

Among the aforementioned approaches, pre-compensated wavefront shaping techniques have emerged recently as attractive methods due to their plain working principle and experimental setup. Algorithms to control wavefronts have already been reported a lot, including stepwise sequential algorithm [10], continuous sequential algorithm [19], phase retrieval algorithm [22], genetic algorithm [20], and neural networks [21], *etc.* Nevertheless, with most reported algorithms such as stepwise sequential algorithm and continuous sequential algorithm, many iterations are required to achieve light focusing, thus the optimization process is inherently time-consuming,

restricting wavefront shaping implementations to static media such as diffusers. Besides, employing a single algorithm towards the globally optimal solution may consume extremely long time and enormous computational source in practice.

The objective of the thesis is to push pre-compensated wavefront shaping to more general and realistic implementations. Specifically, it contains the following two parts:

1. To overcome the limitations of individual algorithms in reaching the global optimum, improving the efficiency of optimization algorithms as well as light focusing performance.

2. To analyze inverse scattering problems in both static and dynamic media physically and mathematically, then develop deep learning frameworks to achieve fast light focusing and focusing recovery through nonstatic scattering media.

1.3. Major Contributions

As the summary of my research work in the past few years, the thesis concentrates on tackling the limitations of pre-compensated wavefront shaping techniques stated above, and the major contributions can be reflected in three aspects.

First of all, in order to resolve the dilemma that employing a single optimization algorithm towards the globally optimal solution is hardly possible, a new thought is exploited that one algorithm can be reinforced by another complementary algorithm since they effectively compensate each other's weaknesses, resulting in a more efficient hybrid algorithm. The reinforced hybrid algorithm significantly improves wavefront shaping speed and performance.

Secondly, deep learning frameworks are proposed based on the understanding of inverse scattering mathematical models in static and dynamic media, realizing light focusing and rapid focusing recovery through time-variant scattering media. With the

adoption of the fine-tuning engine of deep neural networks, time cost in light refocusing can be much shorter than traditional iterative algorithms.

Last but not the least, an adaptive deep learning-empowered framework is introduced, which includes adaptive adjustments of networks' hyperparameters and the recursive fine-tuning algorithm. The adaptive adjustments of hyperparameters are based on the statistical properties of nonstationary stochastic processes, permitting the handling of the spatiotemporal non-stationarity of time-variant media. Moreover, the recursive fine-tuning algorithm makes the best use of speckle correlation, thus light refocusing can be expedited. To the best of our knowledge, this is the kind of first time to introduce an adaptive deep learning-empowered framework to tackle the challenge of optical focusing and refocusing through randomly changing media with wavefront shaping.

1.4. Organization of the thesis

The organization of this thesis is as follows:

Chapter 2 serves as theoretical foundations of the thesis work. The working principle of wavefront shaping is provided, followed by general reviews of state-of-art various wavefront shaping methods. Both merits and limitations of each approach are discussed.

In Chapter 3, a new thought is exploited that a more efficient wavefront shaping algorithm can be developed by the hybrid of two or more conventional algorithms that can effectively compensate each other's weaknesses. A systematical approach combining deep neural networks and the genetic algorithm is introduced as a proof of concept, leading to double optimization speed and 40% improvement in light focusing performance.

In Chapter 4, the implementation of a Deep Convolutional Neural Network (DCNN)

framework to resolve wavefront shaping problems is introduced. The theoretical analysis of inverse scattering problems in static situations is stated first. Inspired by the mathematical model, a specifically designed DCNN is applied to focus light through the scattering medium. Benefiting from the speckle correlation theory and DCNN's fine-tuning property, fast recovery of light focusing performance from degradations caused by environmental disturbances can be feasible.

Chapter 5 focuses on applying deep learning to more complex scenarios that the scattering medium is changing continually at a constant speed. A deep learning framework is introduced and implemented by a Convolutional Long Short-Term Memory (ConvLSTM) network to cope with the stationary variations, realizing faster optical focusing recovery when compared with traditional iterative algorithms.

Chapter 6 concentrates on tackling the ultimate challenge of light focusing and refocusing through randomly changing scattering media which are regarded as nonstationary stochastic processes. The proposed adaptive recursive fine-tuning algorithm makes the best use of the speckle correlation, assuring the timely recovery of light focusing performance from degradations. The proposed algorithm also introduces the adaptive adjustment on hyperparameters of the proposed deep neural network, allowing the handling of spatiotemporal non-stationarity, which cannot be achieved by conventional algorithms.

Finally, in Chapter 7, conclusions are drawn, and future work plans are discussed.

Chapter 2. Controlling Multiple Scattering in Disordered Media with Wavefront Shaping: A Review

2.1. Multiple scattering theory

When propagating inside an inhomogeneous medium, photons can be absorbed and re-radiated along all directions by a particle, and this physical process is described as scattering [23]. Single scattering refers to the phenomenon that light is scattered by a single particle whose dimensions are very much greater than the optical wavelength [24]. With the reduction of distance between scattering particles, multiple scattering will predominate, which suggests that light is thus scattered multiple times. In practice, multiple scattering is much more common since scattering particles are grouped together, and most real-life scenarios are multiple scattering, such as scattering within clouds, opaque glasses, stellar atmospheres, *etc.* [24] While single scattering is random, multiple scattering can be treated as a more deterministic process, and is usually modeled by diffusion theory. In the thesis, unless otherwise specified, all scattering processes are regarded as the multiple type.

The scattering properties of disordered media are generally described by four parameters: scattering mean free path l_s , absorption mean free path l_a , anisotropy factor g , and transport mean free path l_s^* [25]. In scenarios including biological tissues and ground glass diffusers, $l_s \ll l_a$, indicating that in this situation, scattering is predominant while absorption can be neglected. The transport mean free path l_s^* is usually adopted to characterize the distance that light can travel before receiving isotropic scattering, and is defined as $l_s^* = l_s (1 - g)^{-1}$ [25]. The anisotropy factor g

defines the probability that light is forwardly scattered, which is typically 0.8~1 for light scattering in tissues [26], which indicates that forward scattering is dominant.

If light is coherent, scattered light from different paths interfere with each other, and the granular structure of resultant intensity field is called speckles [27]. Although visually random, the way that light is scattered is actually deterministic within a certain time window (usually referred as speckle decorrelation time) [3]. The deterministic scattering process is characterized by a large number of parameters that can up to 10 million [27], fortunately, researchers have recently found that the disordered media could be regarded as optical waveguides that are made up of many beam splitters regardless of scattering properties and media thickness [25]. In the light of the theory, the scattering processes inside media can be simplified to linear equations, which can be described by matrices.

2.2. Properties of optical speckles

Speckles are widely observed when coherent light propagates through an inhomogeneous medium or is reflected by a rough surface. The signal at position $\mathbf{r}$ at time t is composed of a large number of independently phased additive complex components, mathematically, the optical field at this point is given as [28]:

$$A(\mathbf{r}, t) = \sum_n^N a_n(\mathbf{r}, t) e^{i\phi_n(\mathbf{r}, t)} \quad (2.1)$$

where $a_n(\mathbf{r}, t)$ and $\phi_n(\mathbf{r}, t)$ are amplitude and phase value of the n^{th} optical mode at position $\mathbf{r}$ at time t , respectively.

As stated above, although visually random, the statistic properties of speckles can be deterministic within a certain time window. Assuming light is monochromatic and speckles are fully developed, i.e. phase values of all contributing fields $\phi_n(\mathbf{r}, t)$ are

uniformly distributed over $[0, 2\pi]$ [27], the probability distribution of speckle intensity can be written as [28]:

$$p_I(I) = \frac{1}{\bar{I}} \exp\left(-\frac{I}{\bar{I}}\right) \quad (2.2)$$

where $\bar{I}$ is the mean intensity. In this ideal situation, the standard deviation of the intensity $\sigma_I = \bar{I}$. Therefore, both the contrast C and signal-to-noise ratio (SNR) of fully developed speckles should be $C = \frac{\sigma_I}{\bar{I}} = 1$ and $SNR = \frac{1}{C} = 1$, respectively, indicating that noise is severe [28]. The phenomenon of contrast equaling 1 is only valid for linearly polarized light, and for partially polarized or unpolarized light waves, the contrast will be reduced [27].

If numerous contributing phases in Eq. (2.1) do not follow a uniform distribution, speckles will be regarded as partially developed. Under these circumstances, both contrast and SNR of speckles are no longer constants. Instead, they will be related with the total number of optical modes N and standard deviation σ_ϕ of the phase of contributing electric fields (provided that all ϕ_n are identically distributed) [28].

It should note that all the above properties are deduced based on the assumption of static situations. For time-variant scattering media, those properties are still valid on the condition that the integration time of measured speckle intensity is much shorter than the characteristic time of media altering [27].

2.3. Focusing light through scattering media with wavefront shaping

In general, wavefront shaping techniques fall into two main categories: one is pre-compensating the scattering induced wavefront distortions by adjusting the phase values of light using spatial light modulators (SLM) before the light enters into turbid media,

the other is based on time-reversal and phase conjugation of the scattered light [25].

As stated in Section 2.1, the scattering process is simplified to linear equations. The m^{th} transmitted optical field E_m after scattering can be regarded as a linear combination of all N incident optical fields E_n [10], shown as Fig. 2.1(a),

$$E_m = \sum_{n=1}^N t_{mn} E_n = \sum_{n=1}^N |t_{mn}| e^{i\phi_{mn}} A_n e^{i\varphi_n}, \quad (2.3)$$

where A_n and φ_n are the amplitude and phase of the n^{th} input optical field, respectively.

ϕ_{mn} is the phase value of t_{mn} , which is an element of a complex-valued transmission matrix that can be adopted to describe the scattering process. Usually, only the intensity of transmitted speckles is recorded using cameras,

$$I_m = |E_m|^2 = \left| \sum_{n=1}^N |t_{mn}| e^{i\phi_{mn}} A_n e^{i\varphi_n} \right|^2. \quad (2.4)$$

In order to focus light and obtain a single speckle through scattering media, i.e., to maximize the amplitude of E_m , all terms in Eq. (2.3) have to be in phase [10, 29], as shown in Fig. 2.1(b). So far numerous methods have been reported to obtain the required phase values of input optical fields, which will be briefly discussed in the following sections.

Fig 2.1. (a) A coherent light beam with unmodulated wavefront undergoes scattering when it passes through a disordered medium, and speckles are observed outside the medium. (b) A coherent light beam with modulated wavefront is focused to a point after passing through a disordered medium.

2.3.1. Optical wavefront shaping

In order to pre-compensate the phase distortions induced by multiple scattering, the phase values of light are modulated before it enters into disordered media. In the pioneering work by Vellokoop and Mosk [10], an iterative optimization approach is implemented to compute the incident optical phase patterns such that light can be focused to the desired position after scattering. The phase value of each incident optical mode φ_n is cycled from 0 to 2π , and the value that maximizes the recorded light intensity at the target position is recorded. This procedure is repeated for all optical modes, then the incident phase values are simultaneously set to their recorded values, effectively leading to constructive interference at the target position as all terms in Eq. (2.3) are in phase, forming a focal point as the background intensity is significantly suppressed [27]. Besides the stepwise sequential algorithm mentioned above, other iterative optimization algorithms such as continuous sequential algorithm [19], partitioning algorithm [30], and genetic algorithm [20] are also adopted in wavefront shaping, successfully focusing light through static turbid media.

Besides the iterative optimization methods, Popoff *et al.* firstly demonstrated the measurement of a transmission matrix using an interferometric method [11]. With the transmission matrix which is used to describe scattering properties inside a turbid medium, focused speckle can be obtained by simply setting $\varphi_n = -\phi_{mn}$, where φ_n and ϕ_{mn} are the phase value of n^{th} input optical mode and transmission matrix element t_{mn} [27], respectively, circumventing the needs of repeating iterative optimization algorithms every time the target focal point changes. Then Drémeau *et al.* reported a reference-less method to measure the transmission matrix, and they used the Bayesian phase retrieval algorithm to estimate the phase information of the transmission matrix

[22]. Most recently, Tzang *et al.* proposed a 1D-to-2D transform technique to measure the transmission matrix, significantly increasing wavefront shaping speed [31].

2.3.2. Guidestar-assisted wavefront shaping

All wavefront shaping techniques require feedback signals in order to modulate incident optical phase patterns. The feedback signals can be light intensity [10, 11], ultrasound [4-6], photoacoustic signals [3, 32], second-harmonic radiation [33, 34], two-photon [17, 35], magnetic field [36], and fluorescence [37-39]. Considering that light intensity has to be recorded by cameras outside turbid media, it can only be applied to situations of focusing light or imaging through a medium. By comparison, guidestars such as photoacoustic signals can help to conduct wavefront shaping inside a medium.

The wavefront shaping techniques guided by fluorescence, two-photon, and second-harmonic signals are promising but invasive since they require the injection of nanoparticles. Vellekoop *et al.* demonstrated light focusing at the target position labeled by a fluorescence nanoscale probe inside disordered photonic metamaterials, and the intensity of fluorescence light emitted by the probe is used as the feedback since it can reflect received light intensity at the target position [39]. Katz *et al.* reported a diffraction-limit focus through highly scattering samples, using two-photon fluorescence signals as feedback [17]. Hsieh *et al.* placed barium titanate (BaTiO₃) nanocrystals behind a turbid layer, and these BaTiO₃ nanocrystals were used to generate second-harmonic waves. The emitted second-harmonic signals are scattered then phase-conjugated [33]. Ruan *et al.* used magnetic particles as guidestars, and scattered light that interacts with magnetic fields are selectively phase-conjugated using a digital optical phase conjugation (DOPC) system [36].

In contrast, photoacoustic signals, which are generated through thermo-elastic expansion, serve as potential noninvasive guidestars as their amplitude is proportional

to the received light intensity [40]. Kong. *et al* firstly demonstrated the photoacoustic-guided wavefront shaping, successfully focusing light to a single point whose resolution is limited by the acoustic diffraction [32]. Then Lai. *et al* generated nonlinear photoacoustic signals based on the Grueneisen relaxation effect, and the nonlinear signals are utilized as feedback to break the acoustic-diffraction limitation, obtaining a single speckle grain on the scale of 5 - 7 μm [3]. While the above two methods both rely on iterative optimization algorithms, the transmission matrix measurement using photoacoustic signals as the feedback has also been reported [14].

Ultrasounds serve as important guidestars in time-reversal and phase conjugation. Xu. *et al* firstly proposed the time-reversed ultrasonically encoded (TRUE) optical focusing through scattering media [4]. The scattered light is encoded by ultrasound which is provided by a transducer, then only the ultrasonically encoded light is time reversed, and focused inside the transducer focal zone. Then Benjamin *et al.* put forwarded an approach named time reversal of variance-encoded light (TROVE) that encodes individual optical mode with its distinct variances. After time-reversal of computationally decoded modes, the acoustic resolution can be broken [6]. Besides, time-reversed ultrasound microbubble encoded light (TRUME), which takes advantage of changes in optical fields induced by microbubbles, is also capable of breaking the acoustic-diffraction limitation, achieving an optical focus with the size of $\sim 2 \mu\text{m}$ [41].

As most methods reviewed above take advantage of implanted or virtual guidestars to focus light inside or through scattering media, Ma *et al.* proposed a scheme that adopts intrinsic dynamics as guidestars, named as Time-Reversed Adapted-Perturbation (TRAP) [42]. With TRAP, light scattered from the origin of the perturbation can be adaptively time-reversed. Edward *et al.* put forward a technique termed Time Reversal by Analysis of Changing wavefronts from Kinetic targets (TRACK) that analyzes the

differences between time-variant scattering fields induced by a moving object [43].

2.3.3. Deep learning-assisted wavefront shaping and image recovery

Recent years have witnessed fast developments in deep learning, which is a data-driven approach and often known as deep neural networks (DNNs). Deep learning has been widely used to solve inverse problems like denoising [44-47], image reconstruction [48-52], and super-resolution imaging [53-56] owing to its superior ability in revealing complex relationships through transforming representations at one level to a higher and more abstract level [57]. The idea has also been exploited to focus light [21] and reconstruct images [58-60] through static scattering media. For example, Turpin *et al.* introduced neural networks for binary amplitude modulation. With the well-trained network, they reported the realization of light focusing as well as image recovery through a single diffuser [21]. Jin *et al.* employed a convolutional neural network (CNN) for wavefront aberrations. The low-order aberrations can be calculated by the pre-trained CNN, then aberrations are compensated [61]. Li *et al.* trained U-Net with speckles generated by various objects with four diffusers [58]. The pre-trained network can be generalized to “unseen” objects or diffusers. All these diffusers, however, are with the same macroscopic parameters. Sun *et al.* [62] trained five neural networks to model five different scattering conditions, then blurred images are first classified to one of the five situations, then 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 gets a rough classification and reconstruction.

2.3.4. Key parameters

The focal intensity enhancement ratio η is introduced to quantify the light focusing performance, defined as the ratio between the optimized intensity at the region of interest (ROI) and the average intensity before optimization [10]. Theoretically, $\eta \propto \left(\frac{\pi}{4} \times \frac{N}{M}\right)$ if SLM, which can tune the phases of incident optical modes from 0 to 2π , is adopted, where N is the number of elements that are independently controlled on the SLM, and M is the number of output modes within the ROI [63]. Besides SLM, digital micromirror devices (DMDs), which are only capable of binary amplitude modulation but are much faster than SLM, are also widely used in wavefront shaping. With DMD, $\eta \propto \left(\frac{1}{2\pi} \times \frac{N}{M}\right)$. In addition, peak-to-background-ratio (PBR), which is defined as the ratio between the intensity of optimized spot and background [64], also serves as a quantification of light focusing performance.

As mentioned above, multiple optical scattering can be regarded as deterministic within a time window characterized by speckle decorrelation time (SDT). The intensity autocorrelation function $g_{II}(\tau)$ is calculated by

$$g_{II}(\tau) = \frac{\langle I(t_0) \cdot I(t_0 + \tau) \rangle}{\langle I(t_0) \rangle \cdot \langle I(t_0 + \tau) \rangle} \quad (2.5)$$

where $I(t_0)$ and $I(t_0 + \tau)$ are image intensity captured at time t_0 and $t_0 + \tau$, respectively, and $\langle \dots \rangle$ indicates averaging over all the elements of the element-wise multiplication [65]. SDT is defined as the time that $g_{II}(\tau)$ decreases to $1/e$ of its initial value. SDT will influence the wavefront shaping reconstruction fidelity [66], and it is also vital when wavefront shaping techniques are applied to *in vivo* situations.

2.4. Applications of wavefront shaping techniques

The development of wavefront shaping techniques benefits numerous applications in various fields, especially in biological imaging which has been suffering from the obstacle caused by optical scattering for decades.

2.4.1. Photoacoustic imaging

Caravaca-Aguirre *et al.* firstly demonstrated three-dimensional photoacoustic imaging through scattering media using photoacoustic signals as guidestars to pre-compensate wavefronts. A polypropylene tube (90 μm inner and 120 μm outer diameter) filled with India ink is placed behind a diffuser, and the ink will generate photoacoustic signals, serving as the feedback signals. With iterative optimization using the genetic algorithm, photoacoustic signal strength is increased by 10 times, meanwhile, image resolution and contrast are also improved [67]. Then the same group reported super-resolution photoacoustic imaging with wavefront shaping. They took advantage of the Gaussian spatial sensitivity of a focused ultrasound transducer, which indicates that higher weighting factors will be given to photoacoustic signals generated by speckles that are closer to the center of the ultrasound focal zone [68].

2.4.2. Optical imaging

Wavefront shaping techniques open up a new path for existing modalities to overcome scattering, thus deep tissue imaging can be achieved by techniques such as Optical Coherence Tomography (OCT), two-photon microscopy, and fluorescence imaging.

Tang *et al.* proposed a method named as iterative multiphoton adaptive compensation technique (IMPACT), which utilizes the nonlinearity of multi-photon signals as feedback to compensate for the distortion caused by light scattering, and they

achieved a diffraction-limited focal point inside deep tissues with multiphoton microscopy [69]. However, the requirement of implanted guidestars restricts its application in *in vivo* imaging [70].

OCT is widely used as a non-invasive imaging technique, but its penetration depth is severely limited by optical scattering. Fiolka *et al.* presented a wavefront optimization approach that uses the backscattered light as a feedback signal to work with a configuration that is similar to optical coherence microscopy [71], achieving light focusing inside deep tissue with both tissue-mimicking phantoms and fixed brain slices. Park's group reported the enhancement of OCT imaging depth as well as signal-to-noise ratio [72] through the combination of wavefront shaping and spectral domain OCT. They also demonstrated *in vivo* deep tissue imaging of a live mouse tail with wavefront shaping OCT [73].

As one of the important biomedical research tools, the depth of fluorescence imaging is restricted to the ballistic regime due to light scattering. Wang *et al.* took advantage of digitally time-reversing ultrasound-tagged light, and achieved fluorescence imaging of tumor microtissues at a depth of 2.5 mm in biological tissues [8]. Similarly, Si *et al.* combined single-cycle pulsed ultrasound modulation and digital optical phase conjugation to break the limitation caused by scattering, and realized fluorescence imaging with resolution smaller than 40 μm [74]. Most recently, Ruan *et al.* reported ultrasound modulated fluorescence imaging through dynamic media, which takes advantage of the spatiotemporal dynamic properties of speckles [75].

2.4.3. Endoscopic imaging

Endoscopic imaging plays a significant role in modern gastrointestinal diagnosis as well as surgery [70]. However, the implementation of multi-mode optical fibers (MMF), which are often regarded as turbid media, hinders high-resolution imaging in deep

depths. Choi *et al.* proposed a method named lensless microendoscopy by a single fiber (LMSF) that utilizes the measured transmission matrix to reverse the scattering-induced wavefront distortion. They demonstrated partially flexible wide-field endoscopic 3D imaging with a single multimode fiber at a relatively high speed (1 frame/s for 3D imaging) with a high spatial resolution (1.8 μm) [76]. Imaging speed is accelerated by adopting GPU [77].

2.4.4. Optical manipulation

So far, most applications of wavefront shaping techniques focus on imaging, however, they also benefit optical manipulation including optogenetics. Cizmar *et al.* reported *in situ* wavefront aberrations correction based on orthogonal mode decomposition, and the optimal phase mask will be obtained through iterative optimization [78]. They firstly demonstrated the optimized trapping through turbid media. Kim *et al.* reported a technique named tomographic mould for optical trapping (TOMOTRAP) that can realize real-time stable control of arbitrarily shaped particles and biological cells, such as red blood cells, by generating the 3D light field distributions of a trapping beam based on the measurements of 3D refractive index distribution of samples [79].

As for optogenetics, Yoon *et al.* reported non-invasive optogenetic control through a 300 μm -thick mouse skull, which is a highly scattering medium, using wavefront shaping [80]. Ruan *et al.* presented optical focusing in 2-mm-thick living brain tissue using the time-reversed ultrasonically encoded (TRUE) technique [81]. Then they achieved optogenetic modulation of neural activity in 800- μm -thick acute mouse brain slices.

2.5. Memory effect

The optical memory effect describes an intrinsic isoplanatism of the scattering process that tilting the incident light by a small angle will not change the speckles after scattering; instead, speckle patterns remain the same but will tilt accordingly [82]. It is firstly reported by Feng *et al.* and is derived in a waveguide geometry [83]. Then Li *et al.* experimentally demonstrated the correlation in laser speckles, and gave a general expression for angular correlation functions with arbitrary shift [84]. Recently, Osnabrugge *et al.* demonstrated a generalized optical memory effect that both tilting and shifting correlations occur in scattering media [85]. Based on this property, the imaging field can be widened.

Bertolotti *et al.* reported the non-invasive imaging of a fluorescent object that is hidden behind an opaque scattering layer [1]. It is found that within the memory effect range, the autocorrelation of measured speckle intensity is related with the autocorrelation of the imaging object. Using the detected scattered speckle intensity and iterative phase retrieval algorithm, the images of a fluorescent object can be retrieved. Then Yang *et al.* demonstrated imaging of blood cells enclosed in chicken skin tissue using speckle scanning microscopy which relies on the correlation of speckle patterns, achieving the subcellular resolution [86]. Benefitting from the memory effect, Katz *et al.* used spatially incoherent light and demonstrated single-shot high-resolution imaging through scattering media varying from white paint to the dynamic biological tissues [16].

2.6. Summary

Ever since Vellokoop *et al.* firstly demonstrated the control of optical wavefronts through disordered media, over the past several years, wavefront shaping techniques have obtained significant improvements towards faster computation, higher resolution,

and deeper penetration. Nevertheless, so far, there still exists some obstacles that restrict the applications of wavefront shaping. For wavefronts pre-compensation techniques, the biggest restriction is the optimization speed, considering that iterative optimization is generally needed. Besides, the frame rate of commercially available liquid crystal on silicon-spatial light modulators (LCoS-SLMs) is usually 60Hz, which also poses a limitation on light focusing speed. Although DMD can be much faster as its frame rate can reach 23 kHz [63], it is only capable of binary amplitude modulation, leading to the compromising of light focusing performance. With iterative wavefront shaping techniques, it may take up to several minutes even hours to obtain a focused speckle. Improvement roadmaps include improvements in hardware such as faster phase modulators, more powerful GPU, and onboard data acquisition, and developments in software such as more efficient optimization algorithms [25]. In addition, for guidestar-assisted wavefront shaping, the modulation efficiency and introduction of external objects (e.g., fluorescence particles) also pose limitations on wavefront shaping implementations. Ben *et al.* have reported that for photoacoustic guided wavefront shaping, once the dimension of optical absorbers is much larger than the speckle size, wavefront shaping performance will be significantly degraded [87]. Besides, pulsed lasers are required to generate photoacoustic signals, whose coherence length, however, can be short and may not satisfy the strict coherence requirement in wavefront shaping [27].

As for wavefront shaping techniques relying on time reversal and phase conjugation, realizing light focusing within several milliseconds is achievable. As guidestars such as ultrasounds and fluorescence objects are required in these techniques, they suffer from low modulation efficiency. Analog optical phase conjugation-based systems may result in limited energy in time-reversed light [25]. In addition, the experimental setup can be

much more complicated than that used in pre-compensated wavefront shaping methods [8, 88].

Each aforementioned method has its own advantages and limitations. As reviewed above, deep learning has witnessed significant developments in recent years, and it has been used in various fields due to its strong ability in feature extraction and modeling. Besides, some neural networks have similar structures with reconstruction models used in inverse scattering problems, which will be elaborated in the following chapters. Given the success of conventional iterative reconstruction models, it can be expected that neural networks will be powerful in resolving inverse problems. Therefore, this thesis concentrates on applying deep learning to pre-compensation wavefront shaping, aiming at comprehensively solving wavefront shaping problems in nonstatic scenarios, which is of great significance in practical applications.

Chapter 3. Focusing Light Through Scattering Media by Reinforced Hybrid Algorithms

As reviewed in Chapter 2, so far numerous algorithms have been reported in wavefront shaping to pre-compensate scattering-induced phase distortions. Whilst promising as aforementioned representative algorithms are all able to focus light through scattering media, each has its own limitations. Employing a single algorithm towards a globally optimal solution may consume extremely long time and enormous computational source. To alleviate the burden, in this chapter, a new thought is exploited that one algorithm can be reinforced by another complementary algorithm since they effectively compensate each other's weaknesses, leading to a more efficient hybrid algorithm. Herein a systematical approach named as Genetic Neural Network (GeneNN) is introduced as a proof of concept. Preliminary light focusing has been achieved by a deep neural network (DNN), whose results are fed to a genetic algorithm (GA) as an initial condition. The genetic algorithm furthers the optimization, evolving to converge to the global optimum. Experimental results demonstrate that with the proposed GeneNN, optimization speed is almost doubled, and wavefront shaping performance can be improved up to 40% over conventional methods. The reinforced hybrid algorithm shows great potentials in facilitating various biomedical and optical imaging techniques.

3.1. Working Principle

3.1.1. Wavefront shaping using genetic algorithm

The forward scattering process has been described with Eqs. (2.3) and (2.4), which are also working principles for all pre-compensated wavefront shaping techniques. The process of employing GA for wavefront shaping roughly consists of the following five steps: initialization, ranking, breeding, mutation, and iteration, as shown in Fig. 3.1.

First of all, a certain number G of phase patterns are created, while each phase value is chosen from a uniform pseudorandom distribution [20]. Then these patterns are scored by a specially designed fitness function, which is the light intensity at a chosen point, as indicated by Eq. (2.4). The phase patterns are ranked in the light of fitness function evaluation results. The higher the score, the higher the ranking. The next step is breeding. Offspring is generated through $offspring = T \times ma + (1-T) \times pa$, where T is a random binary template, and ma and pa are parent patterns. These two parents are selected according to the rule that patterns with higher ranking are more likely to be chosen. After breeding, some segments in the offspring are mutated and randomly changed. The mutation rate R decreases with the increase of generations n to avoid over mutation, following $R = (R_0 - R_{end}) \times \exp(-n / \lambda) + R_{end}$, where R_0 , R_{end} , and λ are initial mutation rate, final mutation rate, and decay factor, respectively [20]. The offspring will also be evaluated by the fitness function. During each generation, a particular number (in this Chapter, $G/2$) of offspring will be reproduced to replace $G/2$ existing patterns with lower rankings [20]. Afterwards, all G phase patterns are ranked again based on their scores. The above breeding and mutation procedures will be iterated multiple times until the end condition is met. In general, iteration stops when a pre-defined number of generations are reproduced, or the fitness function evaluation result reaches a threshold.

GA can find a fit solution in a short time, and it is robust to noise. However, GA is prone to be stuck in a local minimum, posing the risk that probably better solutions are not able to be explored. Consequently, employing a good initialization is crucial to achieve the global optimum or near-global optima. In addition, GA results are affected by a lot of factors, thus with random initializations, optimization results can vary significantly each time GA is run owing to the uncertainty in randomly created initial solutions.

Fig 3.1. Working principle of employing genetic algorithm for wavefront shaping. First of all, a specific number G of initial phase patterns are created, then they are ranked according to their fitness function scoring results. Two patterns are selected as parents to generate offspring. Several segments of the offspring are mutated (indicated by white circles), then the offspring are evaluated by the fitness function. Each generation $G/2$ offspring are reproduced to replace existing $G/2$ patterns, then all phase patterns are ranked again based on their fitness function scores. The whole breeding and mutating process is iterated multiple times until the end condition has been met, typically, a pre-defined number of generations are reproduced, or the fitness function evaluation result reaches a threshold.

3.1.2. Wavefront shaping using deep neural networks

As a data-driven approach, deep learning uses a totally different strategy to compute

the phase pattern required for light focusing. Known as universal approximators, deep CNNs (DCNNs) have been proved to be powerful in resolving inverse problems [59, 89, 90]. Hence DCNNs are suited to be adopted to model the inverse scattering process via supervised learning, revealing the relationship between transmitted speckles and incident optical phase patterns. As shown in Fig. 3.2, DCNN inputs are the intensity distributions of transmitted speckles recorded by a camera while the outputs are their corresponding incident phase patterns modulated by a SLM. After training, DCNN will establish an accurate mapping from speckles to incident phase patterns, thus DCNN is capable to predict the required phase pattern to focus light through a specific scattering medium.

Fig 3.2. Working principle of applying deep neural networks for wavefront shaping. First of all, a DCNN is trained with a number of samples. DCNN inputs are speckles recorded by a camera and outputs are the corresponding incident phase patterns modulated by a SLM. After training, the DCNN can accurately map speckles to their corresponding phase patterns. Afterwards, a preset focused speckle is sent to the DCNN as the desired pattern, and the phase pattern predicted by DCNN is able to focus light through the medium when the pattern is loaded on the SLM.

Comparing with iterative wavefront shaping methods, the deep learning approach is simpler since the relationship between speckles and incident phase patterns is learnt

straightforwardly through training, circumventing the need for interferometers or iterations. Nevertheless, its performance is significantly affected by training samples. Only when the training sample size is sufficiently large, meanwhile, these samples are typical enough to represent the entire scattering processes, then DCNN will be able to predict the optimal phase pattern for light focusing after training. Unfortunately, it is impractical to include all possible phase patterns and speckles for training, and the typicality of samples is also difficult to measure, thus the supervised learning method genuinely suffers from the dilemma in reaching the global optimum.

Notwithstanding this, the result from DCNNs can be regarded as a good initialization for GA. As for GA, convergence to the global maximum is achieved on the condition that the initial value is around the global optimum [91]. Hence the possibility of the optimization trapped in local minima for GA will be largely reduced thanks to the initial patterns provided by DCNN. Considering that DCNN and GA serve as the complementary algorithm to each other, the hybrid of them, named as Genetic Neural Network (GeneNN), will demonstrate significant improvements against GA or deep learning alone in terms of convergence speed and focusing efficiency.

The proposed GeneNN consists of two parts. The first part is collecting samples to train a DCNN. After training, an initially focused speckle can be obtained with the phase pattern predicted by the DCNN. The second part is adopting GA to further the optimization of the focusing process. Two methods are put forward to create initial phase patterns, both utilizing the DCNN results. The first method, named as GeneNNv1, is taking the pattern predicted by DCNN as one of the initial patterns, while all the other patterns are created following a uniform pseudorandom distribution [20], as shown in Fig. 3.3(a). The pattern from DCNN will certainly have the highest ranking, so it has the highest opportunity to be chosen for breeding. The other method, named as

GeneNNv2, is that all the initial patterns are created based on the DCNN predicted pattern but by augmenting various additional phase patterns to this common base, while all the additional patterns are also produced with a uniform pseudorandom distribution, as shown in Fig. 3.3(b). In the following section, the experimental results of both methods will be presented in detail.

Fig 3.3. Two methods to apply DCNN results to GA in Genetic Neural Network (GeneNN). (a) GeneNNv1. The phase pattern predicted by the DCNN is one of the initial phase patterns in GA. This pattern should have the highest ranking, so it is of great possibility of being chosen to breed the next generation. (b) GeneNNv2. The SLM pattern from DCNN works as the common base, and initial phase patterns in GA are created by augmenting various additional patterns to it.

3.2. Experimental results and comparisons of algorithms

The experiment setup is illustrated in Fig. 3.4. Light emitted from a He-Ne CW laser (633nm, Melles Griot) passes through a ND filter (NDC-50C-4M, Thorlabs), then the light beam is collimated and expanded by 4.3 times with a telescope. A half-wave plate (WPH10M-633, Thorlabs) and a polarizer (LPVISC100-MP2, Thorlabs) are employed together to adjust the polarization state of incident light to be parallel along the long axis of the spatial light modulator (SLM, X13138-01, Hamamatsu). Light is modulated and

reflected by the SLM, hence the SLM patterns are used to represent the phase patterns of light. Light is focused onto the front surface of a diffuser (120 Grit, 83-419, Edmund) by an objective lens (50X/0.80, Nikon TU Plan Fluor). Then light is scattered inside the diffuser. Transmitted optical speckles are collected by another objective lens (20X/0.45, Nikon TU Plan Fluor) placed behind the diffuser. At last, the intensity distribution of the speckles is recorded by a camera (Zyla 4.2 sCMOS, Andor). The input signal levels to the SLM is set as 32, i.e., 32 different grey levels are employed to represent the phase values from 0 to 2π . The resolution of the SLM screen is 1280×1024 , while in experiments the SLM pattern size is 32×32 , i.e., the whole SLM screen is divided into 32×32 macropixels, with one macropixel containing 40×32 pixels. The size of the speckle patterns is 64×64 pixels.

Fig 3.4. Experiment setup. Light emitted from the laser passes through a ND filter (F), then light is expanded and collimated by a telescope (L1 and L2). A half-wave plate (HWP) and a polarizer (P) are used together to adjust the polarization state of light. Light is modulated and reflected by the SLM. The following two lenses (L3 and L4) reduce the beam size so that light can enter an objective lens (OBJ1). OBJ1 focuses light on the surface of the diffuser (D), and another objective lens (OBJ2) placed behind the diffuser collects scattered light and projects it onto the camera.

CNNs have been proved to be intrinsically suitable for solving inverse scattering problems [92]. In addition, DCNNs are known as good at processing images [53, 54, 57,

93]. These properties make DCNNs suited for solving wavefront shaping problems. The structure of the DCNN employed is shown in Fig. 3.5. The DCNN has three inputs and one output. They work collaboratively, contributing together to guarantee that the proposed DCNN can learn sufficient information of the scattering medium, hence an accurate mapping from speckles to their corresponding SLM patterns will be established. Input 2 is randomly generated SLM patterns, while the speckles obtained with them are noted as Input 1. Input 3 is the target speckle, and DCNN output is the required SLM pattern to generate the desired speckle. In general, the last few layers are task specific while the earlier parts are modality specific [94]. With Input 1, Input 2, the first three convolutional layers (Conv1, Conv2, and Cov3), and the first fully-connected layer (FC1), the DCNN is supposed to learn the information of scattering processes in the medium. The last fully-connected layer (FC2) is trained to establish the mapping from speckles to SLM patterns. Conv1, Conv2, and Cov3 work together to extract and abstract image features from Input 1, then these features are flattened to a 1D array. Input 2, which is a 2D SLM pattern, is also flattened to concatenate with the features from Input 1, followed by a fully-connected layer (FC1). Output of FC1 is concatenated with flattened image features from Input 3, then the final fully-connected layer (FC2) predicts the SLM patterns required in order to get the desired speckle indicated as Input 3. Convolutional layers (Conv1, Conv2, and Cov3) consists of 16, 32, and 48 filters, and the filter size of each layer is 7×7 , 5×5 , and 3×3 with stride set as 3×3 , 2×2 , and 1×1 , respectively. Conv1 and Conv4, Conv2 and Conv5, and Conv3 and Conv6 have the same structure. FC1 consists of 512 neurons, whose dropout rate is set as 0.5. FC2 has 1,024 neurons. The activation function of the last output layer is sigmoid, while the activation functions of all other layers are tanh. 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. The Tensorflow Keras library is used to construct the CNN model. The GPU type is NVIDIA GEFORCE GTX 980.

Fig 3.5. The structure of the proposed deep convolutional neural network (DCNN). Input 1 is speckle patterns, and their corresponding SLM patterns are noted as Input 2. Input 3 is the desired speckle, and DCNN output is its required SLM pattern corresponding to Input 3. With Input 1, Input 2, the first three convolutional layers (Conv1, Conv2, and Cov3), and the first fully-connected layer (FC1), the DCNN is supposed to learn the information of scattering processes in the medium. The last fully-connected layer (FC2) is trained to establish the mapping from speckles to SLM patterns.

Experiment results are reported here. First of all, samples were collected for training. 10,000 SLM patterns were randomly created, then they were sequentially sent to the SLM to modulate the incident light phases. Each time a SLM pattern was loaded, the camera recorded the corresponding speckle pattern. Afterwards, both SLM patterns and speckles were respectively normalized to be between 0 and 1. Epoch was set as 15. Training time was approximately three minutes. After training, a preset focused speckle pattern (a Gaussian beam spot generated in MATLAB) was sent to the DCNN as the

desired pattern, then one SLM pattern was predicted and sent to the SLM for light modulation, resulting in a transmitted speckle shown as Fig. 3.6(a) in practice. A focal point has already been observed. The intensity enhancement factor η is introduced to quantitatively evaluate light focusing performance, which is defined as the ratio between the intensity at the focal point after optimization and the average intensity before light focusing [10]. The enhancement factor η achieved by DCNN is 33. Next, the initial phase patterns for GA were produced according to the two proposed algorithms indicated above (GeneNNv1 and GeneNNv2). For GeneNNv2, additional phase values were selected from $[-\pi/2 \pi/2]$, following a uniform pseudorandom distribution as well. The summation patterns were scaled to $[0 2\pi]$. The optimization performance of both algorithms were tested when the population size G was 10, 20, 30, 40, and 50, while measurement times varied with different sizes. Each algorithm was run 10 times under each population size, then the results were averaged. For comparison, GA was also tested 10 times using a random initialization. Experimental results were shown from Fig. 3.6(c) - 3.6(g). Fig. 3.6(b) is a representative focused speckle achieved using GeneNNv2 and the population size G is set to 50. The focal point in Fig. 3.6(b) is much brighter and the background is also darker than Fig. 3.6(a), with η improved more than four times, reaching 148. Fig. 3.6(a) and (b) use the same colormap and scale. As illustrated in Fig. 3.6(c) - 3.6(g), η reached by the two hybrid algorithms (GeneNNv1 and GeneNNv2) are always higher than the one by random GA regardless of measurement times. As more and more generations are reproduced, the merits of the hybrid algorithm become more significant as its enhancement factor η rises faster. To achieve the same η , with any population size tested, the time spent by hybrid algorithms is only a half of that random GA needs. With different hybrid algorithms or population sizes, the increase percentage in η over randomly initialized GA varies, as listed in Table 3.1. The increase percentage

of η is defined as $[(\eta \text{ achieved by GeneNN} / \eta \text{ achieved by a randomly initialized GA}) - 1] \times 100\%$. Except the differences in initial conditions, all the other parameters are the same. The improvement in η ranges from 20% - 40%, depending on the algorithm and population size. However, in all cases, the hybrid algorithms contribute significantly to better focusing.

Fig 3.6. Experiment results of GeneNNv1, GeneNNv2, randomly initialized GA, and DCNN. (a) A focused speckle obtained using the SLM pattern predicted by the pre-trained DCNN. (b) Focusing result with GeneNNv2, while the population size G is 50. (a) and (b) use the same colormap and scale. (c)-(g) Intensity enhancement factor η achieved by the three algorithms: GeneNNv1, GeneNNv2, and randomly initialized GA when population size G is set as 10, 20, 30, 40, 50, respectively. (h) Different numbers of samples are used for DCNN training, resulting in different η after light focusing. The increase percentage in η over the result achieved when training sample size is 10,000 is a function of training sample size. (i) To increase η to 60, the time spent by three algorithms (GeneNNv1, GeneNNv2, and randomly initialized GA) are different at different population size. (j) To increase η to 50, the time spent by three algorithms (GeneNNv1, GeneNNv2, and DCNN) are different at different population size. (k) With the same time spent $t=500$ s, the η achieved by different algorithms (GeneNNv1, GeneNNv2, and randomly initialized GA) are different at different population size. (l) With the same time spent $t=500$ s, the η achieved by different algorithms (GeneNNv1, GeneNNv2, and DCNN) are different at different population size.

Table 3.1. Increase percentage in the intensity enhancement factor η comparing with randomly initialized GA.

	$G=10$	$G=20$	$G=30$	$G=40$	$G=50$
GeneNNv1	18%	27%	21%	44%	35%
GeneNNv2	20%	23%	21%	31%	33%

In GeneNN, DCNN was trained with 10,000 samples, then GA continued the reinforced optimization based on the DCNN prediction. In order to compare with optimization performance relying on deep learning individually, the size of training samples is enlarged, improving DCNN performance accordingly. Results are illustrated in Fig. 3.6(h). The increase percentage of η is calculated as the ratio between η achieved by the DCNN when the number of training samples is increased and η reached with 10,000 samples. It can be expected that the larger the sample size, the higher the η . Enlarging training sample size by about 60% is able to improve the focusing performance by almost 50%. Nevertheless, limited by the SLM frame rate, collecting 6,000 samples takes nearly 18 minutes. In contrast, from Fig. 3.6(c)-(g), with GeneNN, increasing DCNN performance by 50% requires only 50-200 measurements, depending on the algorithm and population size. At all events, the optimization time is always less than 10 minutes, which is almost half of the time spent in deep learning. Therefore, the hybrid GeneNN can be claimed as a more efficient approach in optimizing SLM patterns against individual deep learning.

The experimental results confirm the efficacy of the GeneNN in improving the optimization performance in various aspects. The hybrid of the DCNN and GA successfully leads to a reinforced algorithm considering that they compensate for each other's drawbacks. Instead of employing a random initialization, the initial phase patterns in GA are created with the help of the pre-trained DCNN. Phase patterns from DCNN have already realized light focusing, indicating that they are already close to the globally optimal SLM pattern. During breeding, offspring are generated by amending the SLM pattern from the DCNN, thus the optimization process is effectively guided, avoiding the time-consuming random trials at the initial stage of GA. In addition, since GA inherently favors patterns with higher fitness scores, a good initialization decreases

the risk that the gene pool is trapped around one local minimum, which may result in the failure in exploring globally optimal solutions. It can be observed from experimental results that faster and smoother convergence has been achieved by the GeneNN, indicating that a better convergence direction has been found. However, the focusing performance achieved by the proposed two hybrid algorithms (GeneNNv1 and GeneNNv2) are almost the same, which suggests that with the two methods, DCNN actually imposes almost equal influence on phase patterns generation, and contributes robustly to the performance improvement of subsequent GA algorithm.

3.3. Discussion

Focusing light and imaging inside or through disordered media is an important yet challenging problem. As a widely accepted solution, wavefront shaping successfully achieves light focusing by virtue of various algorithms, whilst each has its own pros and cons. Hence, a hybrid of complementary algorithms is proposed to benefit each other, towards developing a reinforced algorithm. As a proof of concept, the GeneNN has demonstrated effective performance enhancement in many aspects including convergence speed and light focusing performance. Applying DCNN results to assist the creation of initial patterns in GA, a much better focused speckle could be observed, and the optimization process is also faster than GA or CNN algorithm alone.

The performance of the GeneNN algorithm is affected by lots of elements, among which the influence from population size G and mutation rate R are particularly significant. First of all, the impact of population size G is discussed. With different G , the number of measurements required to increase η to 60, 70, and 80 also varies, as listed in Table 3.2. In general, the larger the population size G , the less the measurements needed to achieve the same η , regardless of hybrid algorithms. With GeneNNv1, in order

to improve η to 60, the necessary measurements are almost 1/5 of that needed when G is only 10. By virtue of the larger diversity in a higher population, the overall probability of generating better offspring becomes larger. Moreover, with a larger number of patterns, better light focusing can be achieved without significantly increasing computational time [95]. For instance, with GeneNNv2, when the population size G is 10, nearly 7 minutes (150 measurements) were spent to increase η to 60 (Fig. 3.6(c)), while less than 6 minutes (less than 50 measurements) were costed once the G was enlarged to 50 (Fig. 3.6(g)). In this circumstance, a larger population size not only contributes to better light focusing, but also saves computational time. Besides, in Fig. 3.6(f) and 3.6(g), with the increase of measurement times, η rises much more smoothly than in Fig. 3.6(c)-(e), which implies that as the population size G becomes larger, GeneNN will be more stable, and more robust to environmental disturbance. In addition, the improvement percentage realized by GeneNN over randomly initialized GA is also affected by the population size G . In Table 3.1, when G is increased from 10 to 50, the improvement percentage also rises up from nearly 20% to 40%, while the way of applying DCNN results to GA (GeneNNv1 or GeneNNv2) does not induce obvious difference on final optimization results. Despite this, a larger population size does not always be helpful. Chen *et al.* has discussed that in some situations, a large population may degrade algorithm performance [96]. The increase of the population size can improve the accuracy, but it will saturate when reaching a certain size, then the adoption of more phase patterns will only consume more computational resources without any improvement in accuracy [97]. Finding a suitable population size is nontrivial, requiring both experience and time.

Table 3.2. The number of required measurements to reach different η with varying population size G and hybrid algorithms.

	Population size G	GeneNNv1	GeneNNv2
$\eta=60$	10	173	161
	20	100	89
	30	81	71
	40	58	68
	50	36	42
$\eta=70$	10	214	177
	20	118	106
	30	106	87
	40	76	79
	50	44	50
$\eta=80$	10	282	269
	20	146	136
	30	123	106
	40	93	105
	50	49	59

Figs. 3.6(i)-(l) are shown to demonstrate the influence of the population size G on the optimization speed and performance of different algorithms (GeneNNv1, GeneNNv2, randomly initialized GA, and DCNN) more clearly. As indicated in Fig. 3.6(i) and Table 3.2, to enhance the η to the same value (set as 60 here), when $G=10$, the time spent by GeneNNv1 and GeneNNv2 are 484 seconds and 451 seconds, respectively. As G is increased to 50, the time spent by GeneNNv1 and GeneNNv2 are reduced to 234 seconds and 273 seconds, respectively. The optimization time is reduced by nearly 50%. Besides, as stated above, with the enlarging of G , the reduction percentage in optimization time achieved by the hybrid algorithm against randomly initialized GA also increases. The reduction percentage is defined as $[(1 - \text{optimization time spent by the hybrid algorithm} / \text{optimization time spent by randomly initialized GA}) \times 100\%]$. When G is increased from 10 to 50, the optimization time reduction percentage increases from 20% to almost 40%.

In addition, to enhance the η to the same value (set as 50 here), the time spent by

two proposed GeneNN algorithms and DCNN training are also investigated, and results are shown in Fig. 3.6(j). To improve the η achieved by DCNN, the training sample size needs to be enlarged. The time spent by DCNN is only determined by the number of samples (with a fixed sample collection speed), and it has nothing to do with the population size G . From Fig. 3.6(h), to increase η from 33 to 50, 6,000 samples are needed, and sample collection needs 1080 seconds. By comparison, to enhance η from 33 to 50, with GeneNNv1, the time spent varies from 370 seconds ($G=10$) to 182 seconds ($G=50$), which is only 17% ~ 34% of the time required by DCNN.

Besides the optimization speed, the optimization performance (quantified by the η) is also a function of the population size. With the optimization time fixed as 500 seconds, the η achieved by the two proposed hybrid algorithms, randomly initialized GA, and DCNN training are shown in Fig. 3.6 (k) and (l), respectively. As seen from Fig. 3.6(k), as G is enlarging from 10 to 50, η achieved by GeneNNv1 increases from 62 to 107, realizing a nearly 70% enhancement in η . As for DCNN training, nearly 3,000 samples can be collected within 500 seconds, and the η is enhanced to nearly 40.

In GeneNN, breeding leads the convergence to a specific minimum, while mutation tries to expand the searching space and avoid convergence. As a consequence, mutation rate R plays an important role in the optimization process. Finding a balance between exploration and exploitation is vital. An experimental comparison of the η achieved with various mutation rate R using two GeneNN algorithms is shown in Fig. 3.7. Initial mutation rate R_0 , final mutation rate R_{end} , and decay factor λ were adjusted respectively to discover their effect on the GeneNN optimization process. The population size G was set at 30. Comparing the results under different conditions, in general, a lower mutation rate R leads to faster convergence, regardless of the hybrid algorithm. Among the three parameters, initial mutation rate R_0 imposes larger influence on the GeneNN

optimization than the other two factors, considering that the mutation rate R decays following the equation $R = (R_0 - R_{end}) \times \exp(-n / \lambda) + R_{end}$, the diminution in R_0 contributes to faster decreasing in R . Nevertheless, employing a smaller mutation rate R poses a risk of not reaching the global optimum due to the lack of explorations. If the mutation rate R is too high, then optimization process will be degraded to a random trail. The best mutation rate R is always problem-specific, and an adaptive steering of mutation rate R is preferred [98]. In the reported work, a gradually decreasing mutation rate R is adopted. The initial larger mutation rate is able to induce high diversity in phase patterns to effectively avoid converging to local maxima. With the increase of iteration times, mutation rate keeps reducing adaptively in order to deliver a specific and best solution.

Fig 3.7. GeneNN optimization results with different mutation rate R using GeneNNv1 and GeneNNv2.

The experimental results presented in Section 3.2 demonstrate that adopting DCNN results as initial conditions for GA efficiently contributes to improving light focusing performance. Herein the influence of the quality of DCNN predictions on the performance of the two GeneNN algorithms is further explored, and results are shown in Fig. 3.8. DCNN was trained by five sets of samples whose sizes are varied, resulting in different η after light focusing. With each sample set, DCNN was trained five times. Each time after training, the light was modulated by the SLM pattern output from the

DCNN, and a focused speckle could be recorded. Then results were averaged. The enhancement factor η achieved by DCNN with different sample sets was 42, 37, 32, 26, and 19, respectively. As expected, the less the training samples, the lower the η .

Fig 3.8. The focusing performance of GeneNNv1 and GeneNNv2 with various initial η achieved by DCNN at different population size G . (a) The enhancement factor η achieved by GeneNNv1 when $G=10$ with different initial η offered by DCNN. The initial η realized by DCNN are 42 (grey line), 37 (red line), 32 (blue line), 26 (green line), and 19 (purple line), respectively. (b) The enhancement factor η achieved by GeneNNv2 when $G=10$ with different initial η . (c) The enhancement factor η achieved by GeneNNv1 when $G=20$ with different initial η . (d) The enhancement factor η achieved by GeneNNv2 when $G=20$ with different initial η . (e) The enhancement factor η achieved by GeneNNv1 when $G=30$ with different initial η . (f) The enhancement factor η achieved by GeneNNv2 when $G=30$ with different initial η .

Afterwards, using different DCNN results as the initial conditions, each GeneNN algorithm was run ten times with different population size G , then the mean value was calculated. The number of phase patterns was set as 10, 20, and 30. At the initial stage of GA, the phase pattern from DCNN is randomly modified to some extent due to

breeding and mutation, which causes a drop in η , as observed in Fig. 3.8. For better vision, the initial η offered by the DCNN is shown as the first measured value in Fig. 3.8, while the second value is with the first generation with GA. Steep drops exist in all figures in Fig. 3.8. But soon the performance will be recovered and exceeded. With the increase of the population size G , η recovery speed becomes higher. To fully restore the DCNN performance, approximately 100 measurements are required when G is 10, while less than 50 measurements are enough as G is enlarged to 30. Besides, under the same population size, the drop of η is more significant with GeneNNv1 than GeneNNv2 after the first measurement of GA, as shown in Table 3.3. The drop percentage of η is defined as $[1 - (\eta \text{ obtained with offspring of the first generation in GA} / \eta \text{ achieved by the pre-trained DCNN})] \times 100\%$. For GeneNNv1, the drop percentage of η is 80% - 90% compared with the original DCNN performance, whereas, using GeneNNv2, the drop percentage is 60% - 80%. This phenomenon implies that larger changes in phase patterns are induced by GeneNNv1. The result is reasonable considering that with GeneNNv2, all the initial phase patterns are created based on the SLM pattern from DCNN. No matter which two patterns are selected for breeding, the next generation will still be influenced and guided more or less by the DCNN results. In contrast, in GeneNNv1, the SLM pattern from the DCNN only serves as one initial phase pattern. In this situation, offspring receive less information from the DCNN, thus are more susceptible to random changes. In addition, with the same hybrid algorithm and measurement times, better preliminary focusing results from the DCNN contribute to higher GeneNN performance. As can be seen in Fig. 3.8, when η achieved by DCNN is 42 or 37, after reinforced optimization, η reached by GeneNN is approximately twice of that obtained when DCNN performance is only 19 or 26. In general, the higher the η achieved by the DCNN, the better the focusing performance of the GeneNN algorithms, regardless of the

population size. Higher η suggests that the phase pattern from DCNN is closer to the optimal one, and is less likely to be around local optima. During the reinforced optimization process, better SLM patterns are further generated as they are obtained by applying adjustments to the initial SLM pattern, thus improved focused speckles will be observed.

Table 3.3. The drop percentage of the enhancement η after the first measurement of GA against initial η offered by the DCNN with different population size G , hybrid algorithms, and initial η .

η achieved by the DCNN		42	37	32	26	19
$G=10$	GeneNNv1	77.8%	81.7%	85.5%	79.3%	87.9%
	GeneNNv2	75.5%	76.4%	76.9%	76.9%	61.7%
$G=20$	GeneNNv1	84.8%	89.8%	83.5%	91.3%	89.6%
	GeneNNv2	57.2%	80.6%	89.0%	70.9%	84.2%
$G=30$	GeneNNv1	88.6%	83.4%	83.1%	92.6%	89.4%
	GeneNNv2	61.4%	80.7%	74.3%	67.2%	71.1%

3.4. Summary

In summary, a new strategy is introduced to develop a hybrid algorithm for adaptive wavefront shaping to reinforce a single algorithm by compensating its limitations. As an institutive approach, the hybrid of deep neural networks and genetic algorithm is proposed, named as GeneNN, considering that they can effectively compensate for each other's drawbacks. The phase pattern output by a pre-trained DCNN has already realized a preliminary focused speckle, then optimal focusing performance is reinforced by successive optimization with a GA. Two different hybrid algorithms are put forward to apply DCNN results to GA. Despite that both of them demonstrate similar improvements, GeneNNv2 leads to less drop of the η in the initial stage when GA is applied after the DCNN. It has been proved that GeneNN achieves reliably higher η and faster convergence rate than individual GA or DCNN can do. The pre-focusing performance reached by the DCNN and the population size G can significantly influence

the reinforced optimization performance of the GeneNN. In general, the higher the η reached by the DCNN, the better the performance of the GeneNN. With the increase of population size G , the optimization process will be more stable, and the performance improvement percentage realized by the GeneNN over single algorithm will also be higher. This pioneering work demonstrates that the hybrid of supervised learning and further reinforced algorithms can effectively enhance individual algorithm, which has great potentials in boosting global optimization efficiency in various aspects.

Chapter 4. Deep Learning-Assisted Wavefront Shaping through Disordered Media with Perturbations

Optical focusing at depths in tissue is the Holy Grail of biomedical optics that may bring revolutionary advancement to the field. Wavefront shaping is a widely accepted approach to solve this problem, however, as reviewed in Chapter 2, implementations thus far have their own advantages and limitations. As indicated before, deep learning is intrinsically powerful to solve inverse scattering problems without complicated computation. In this chapter, first of all, a deep convolutional neural network named as ReFocusing-Optical-Transformation-Net (RFOTNet), which is implemented as a Multi-Input-Single-Output network, is proposed to focus light through scattering media based on the mathematic analysis. More importantly, it tackles the challenge of light focusing recovery through scattering media under perturbations. Considering that optical speckles of the medium before and after moderate perturbations are correlated, an optical focus can be rapidly recovered based on fine-tuning of pre-trained neural networks, significantly reducing the time and computational cost in refocusing. The feasibility has been validated experimentally.

4.1. Theoretical analysis of implementing deep convolutional neural networks for light focusing and refocusing through scattering media with perturbations

4.1.1. Inverse scattering problems and deep convolutional neural networks

As shown in Fig. 4.1, the scenario is that a monochromatic optical wave field propagates from the source to a scattering layer, and the transmitted scattered light is collected by a camera. After scattering, the optical field at the receiving plane r_c (e.g. the camera plane in Fig. 4.1) is given using the Fresnel-Kirchhoff diffraction formula [99], where the effects of absorption are neglected,

$$E(\mathbf{r}_c) = \iint d^2\mathbf{r}_b g(\mathbf{r}_c - \mathbf{r}_b) \frac{i}{\lambda} E(\mathbf{r}_b) = \iint d^2\mathbf{r}_b g(\mathbf{r}_c - \mathbf{r}_b) \frac{i}{\lambda} \left(\frac{i}{\lambda} \iint d^2\mathbf{r}_a G(\mathbf{r}_b, \mathbf{r}_a) E(\mathbf{r}_a) \right), \quad (4.1)$$

where $\mathbf{r}_a$ is the source plane (e.g., the SLM plane in Fig. 4.1), $\mathbf{r}_b$ is the back surface of the scattering layer. $E(\mathbf{r}_a)$, $E(\mathbf{r}_b)$, and $E(\mathbf{r}_c)$ are optical fields at $\mathbf{r}_a$, $\mathbf{r}_b$, and $\mathbf{r}_c$, respectively [99]. $G(\mathbf{r}_b, \mathbf{r}_a)$ is the unknown Green function for light traveling from $\mathbf{r}_a$ to $\mathbf{r}_b$, while $g(\mathbf{r}_c - \mathbf{r}_b)$ is the free space Green function for light traveling from $\mathbf{r}_b$ to $\mathbf{r}_c$ [99]. Regular cameras only record the light intensity distribution of the speckle patterns on $\mathbf{r}_c$, thus

$$I(\mathbf{r}_c) = |E(\mathbf{r}_c)|^2. \quad (4.2)$$

Based on the above forward scattering model, to precisely compute the required incident complex optical field $E(\mathbf{r}_a)$ with which light is focused at $\mathbf{r}_c$, the inverse scattering model has to be obtained based on the recorded transmitted light intensity distribution $I(\mathbf{r}_c)$. From Eqs. (4.1) and (4.2), it can be seen that the inverse scattering problem is nonlinear and ill-posed. These properties prohibit the adoption of direct inversion methods; iterative optimization with regularization is necessary to encounter this problem [100].

In order to fit the general regularized formulation and solve the specific light scattering problem, here Eqs. (4.1) and (4.2) are combined into a discretized form

$$I_c = \left| \sum_a^N t_{ca} E_a \right|^2, \quad (4.3)$$

where t_{ca} is a complex transmission coefficient describing light propagation from the $\mathbf{r}_a$ on the source plane to the receiving plane $\mathbf{r}_c$. Therefore, the general Tikhonov regularized formulation is given as

$$\hat{f}_\alpha = \arg \min_f \left\| I_c - |Hf|^2 \right\|^2 + \alpha \|f\|^2, \quad (4.4)$$

where H is the forward scattering model relating the measured speckle patterns I_c and light source f , $I_c = Hf$, and α is the regulation parameter. In the wavefront shaping, f will be the input optical field E^{in} that consists of N different optical modes, $E^{in} = [E_1^{in} \ E_2^{in} \ \cdots \ E_N^{in}]$.

Fig 4.1. A brief illustration of light scattering. The phase patterns of light are modulated by a spatial light modulator (SLM), then light is focused onto a scattering layer (S) by an objective lens (OB1). Scattered light transmitted the scattering layer is collected by another objective lens (OB2) and recorded by a camera.

A lot of iterative methods, such as distorted Born iterative method [101], subspace optimization method (SOM) [102], and iterative shrinkage and thresholding algorithm (ISTA) [103], have been reported to solve Eq. (4.4). Among them, Yuan *et al.* prove the existence and efficacy of the inverse scattering function in light scattering problem [104]; Xu *et al.* confirm that CNNs can be used to solve the inverse scattering problem thanks to the kernel separability [89], and Jin *et al.* construct a model revealing the intrinsic relationship between the CNN and the iterative optimization methods for shift-invariant systems [92]. Note that from Eq. (4.1), due to multiple scattering, the whole system is no longer shift-invariant. As CNN has the capability to learn the degree of shift variance and partially compensate it [59], CNN has the potential to resolve inverse scattering problems.

In regard to Eq. (4.4), most conventional iterative algorithms rely on the building block model [92]:

$$p^{m+1} = A_\theta \left[\frac{1}{L} W * H * I_c + \left(I - \frac{1}{L} W * H * HW \right) p^m \right], \quad (4.5)$$

where the desired estimation $\hat{E}^{in} = Wp$, in which W is a convolutional transformation and p is the transformation coefficients. L is the Lipschitz constant, $L \leq \text{eig}(W * H * HW)$. From Eq. (4.5), the iterative optimization process can be regarded as a sequence of linear filtering by kernel $I - (1/L)W * H * HW$ and bias $(1/L)W * H * I_c$, followed by a point-wise nonlinear operation A_θ by value θ . This is actually quite similar to a typical structure of CNN [92]. As the abovementioned iterative methods can effectively (albeit time consuming) solve the inverse scattering problem, it is natural to hypothesize that CNNs can be implemented to tackle the problem.

4.1.2. Speckle correlation theory

When the scattering medium is subject to perturbations, both forward scattering model H and inverse scattering model H^{-1} change accordingly. 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 changes are correlated [83], which can be shown below

$$\begin{aligned} C_{nmn'm'} &= C_{nmn'm'}^{(1)} + C_{nmn'm'}^{(2)} + C_{nmn'm'}^{(3)} \\ &= D_1 \langle T_{nm} \rangle \langle T_{n'm'} \rangle \delta_{\Delta \mathbf{k}_n, \Delta \mathbf{k}_m} F_1(\Delta \mathbf{k}_n L) \\ &\quad + D_2 g^{-1} \langle T_{nm} \rangle \langle T_{n'm'} \rangle \delta_{\Delta \mathbf{k}_n, \Delta \mathbf{k}_m} [F_2(\Delta \mathbf{k}_n L) + F_2(\Delta \mathbf{k}_m L)] \\ &\quad + D_3 g^{-2} \langle T_{nm} \rangle \langle T_{n'm'} \rangle \end{aligned} \quad (4.6)$$

Here, T_{nm} is the transmission coefficient between the incident mode n and transmitted mode m (similar for $T_{n'm'}$), D_1, D_2, D_3 are constants, $\Delta \mathbf{k}_n \equiv |\mathbf{k}_n - \mathbf{k}_{n'}|$ (similar for $\Delta \mathbf{k}_m$), where $\mathbf{k}_n$ and $\mathbf{k}_m$ are the incident and transmitted wave vectors, respectively. $F_1(x)$

and $F_2(x)$ are form-factor functions. $C_{nmn'm'}^{(1)}$, $C_{nmn'm'}^{(2)}$ and $C_{nmn'm'}^{(3)}$ contribute to the short-range correlation, long-range correlation, and infinite-range correlation, respectively [105]. For most scattering media, the magnitude of $C_{nmn'm'}^{(1)}$, $C_{nmn'm'}^{(2)}$ and $C_{nmn'm'}^{(3)}$ decrease in sequence. After perturbation, the medium is changed to some extent, and the correlation before and after the change is governed by $C_{nmn'm'}$. Based on the correlation, the inverse scattering function H_2^{-1} after the change can be deduced from H^{-1} as long as the CNN has learnt the difference between these two states, calling for the necessity of adopting the fine-tuning of CNN using samples collected from the changed state. Thanks again to the correlation, the number of fine-tuning samples can be much smaller than the samples used for training the CNN from random initialization, which contributes meaningfully to decreasing the time spent in re-optimization. From Eq. (4.6), it is obvious that the larger the change, the smaller the correlation coefficient $C_{nmn'm'}$, and more samples will be needed for fine-tuning. After fine-tuning, the new CNN is able to adapt to the new relationship between the speckles and incident light wavefronts for the perturbed medium.

4.2. Working Principle

As known, U-Net has been widely used to solve image reconstruction problems through scattering media [92, 100, 106]. Image features are extracted and abstracted in contracting paths, which significantly enlarges the effective receptive field of the network and improves the output prediction [100]. Then expansion paths upsample feature maps back to the original image resolution [107]. Nevertheless, image reconstruction concerns more about the quality of the restored images, while it does not pose high demands on time. As for wavefront shaping, time is critical due to the

limitation of speckle decorrelation time, thus a simpler while efficient network specifically targeting on learning the information of scattering processes and modeling inverse functions is preferred in order to speed up SLM patterns computation.

The structure of the proposed RFOTNet is shown in Fig. 4.2(a). The RFOTNet has three inputs and one output. They work collaboratively as a team to learn the information of the medium, establishing the transformation from speckles to their corresponding SLM patterns. Input 1 is the speckle pattern recorded by a camera, while the corresponding phase pattern of the light entering the scattering medium is noted as Input 2. Since phase patterns of light are usually adjusted by a spatial light modulator (SLM), they are represented by SLM patterns. Input 3 is the targeted speckle pattern, while RFOTNet output is the SLM pattern needed in order to get Input 3 through the scattering medium. Each time the scattering medium is changed, the RFOTNet has to be adjusted to model the new medium. In transfer learning, generally, only the last few layers, rather than the whole neural networks, are fine-tuned [94, 108-110] as the last layers are task specific while the earlier ones are modality specific [111]. Information learnt 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, in our experiment, when the fine-tuning of RFOTNet is required, only the last fully-connected layer (FC2) in the RFOTNet is fine-tuned while all the other layers are frozen. By doing so, both time and computational resources can be saved without significant sacrifice of accuracy. The three convolutional layers, Conv1, Conv2, and Cov3, extract image features from Input 1, then these features are flattened to a 1D array to concatenate with Input 2, which has also been flattened. The combination serves as the input to the first fully-connected layer (FC1), followed by a dropout layer. The outputs of FC1 concatenate with the image features extracted from Input 3. The final fully-connected

layer (FC2) predicts the SLM pattern needed for Input 3. Conv1, Conv2, and Conv3 consist of 16, 32, and 48 filters, respectively, and the filter size of each layer is 7×7 , 5×5 , and 3×3 with stride setting as 3×3 , 2×2 , and 1×1 , respectively. Conv1 and Conv4, Conv2 and Conv5, and Conv3 and Conv6 have the same structure, respectively. The number of neurons in FC1 is 512 with a dropout rate set to 0.5. The number of neurons in FC2 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. The activation function of all layers are tanh, except for the last output layer whose activation function is sigmoid. The Tensorflow Keras library is used to construct the model, and the GPU used in computation is NVIDIA GEFORCE GTX 980.

It is worth noting here that the lynchpin in solving nonstatic problems is the proposed framework, rather than any other specific implementations. Despite this, comparing with a conventional CNN which simply learns the mapping from speckles to SLM patterns, the RFOTNet demonstrates superior performance in terms of convergence speed, light focusing, and refocusing results (Comparisons are shown in Section 4.3).

Fig 4.2. (a) Structure of the proposed RFOTNet. RFOTNet has 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 targeted speckle pattern, while RFOTNet output is the SLM pattern needed in order to get Input 3 through the scattering medium. (b) Illustration of the working principle of the proposed method. First of all, samples are collected to train RFOTNet. After training, the RFOTNet can model the inverse scattering function, and output a SLM pattern that can focus light through the scattering medium. When the medium is altered due to perturbations, the focal point can be degraded or even lost. Benefiting from the correlation of speckles before and after the change, only a small number of extra samples are sufficient to fine-tune the pre-trained RFOTNet in step 3. After the tuning, the RFOTNet can adapt to the concurrent medium state and recover the optical focus.

Fig. 4.2(b) illustrates the working principle of the proposed method. In the first step, samples are collected for neural network training. After training, RFOTNet will precisely model the inverse scattering function and establish the relationship between SLM patterns and the corresponding speckles patterns. Then a focused (speckle) pattern is sent to the RFOTNet through Input 3 as the target, and RFOTNet predicts the SLM pattern required in order to generate the focused target in the current system. Light modulated by this SLM pattern propagates through the same scattering medium, and is supposed to focus to a single point, which should be the same as the target. Affected by unavoidable perturbations from the environment, the scattering medium is changed to some extent, causing the focus to be degraded or even lost. To recover this focus, samples collected in the changed medium are used to fine-tune the RFOTNet trained in

Step 1. As the optical fields of the medium before and after the change are correlated, the number of samples for fine-tuning is much smaller than that used for retraining a network. After the directed adjustment, the RFOTNet will be able to adaptively model the perturbed medium, and produce an SLM pattern that can recover the optical focusing. The scattering medium can be perturbed continually, and each time it is altered, the RFOTNet is tuned to adapt to the concurrent state, and helps to refocus light. These four steps function together, assuring optimal light focusing through the nonstatic medium.

4.3. Simulation and experimental results

4.3.1. Simulation results

To verify the feasibility of the proposed method, simulation was conducted. A transmission matrix generated from a circularly symmetric Gaussian distribution [20] is used to represent light scattering paths inside a disordered medium. The size of the SLM patterns was 32×32 , while the size of speckle patterns was 64×64 . In Step 1, 10,000 samples were generated based on this transmission matrix for RFOTNet training. After training, a SLM pattern can be predicted by the RFOTNet. With such a SLM pattern, light was successfully focused through the scattering medium as shown in Fig. 4.3(a). Note that Figs. 4.3a-c use the same colormap. Next, different levels of perturbations are modeled by adding matrices with different variances to the original transmission matrix. The additional matrices also follow a circularly symmetric Gaussian distribution. After perturbation, the optical focus was degraded (Fig. 4.3b). To quantitatively evaluate the focusing performance, the intensity enhancement factor η is defined as the ratio between the optimized intensity at the chosen position and the average intensity before optimization [10]. Correlation coefficient of speckles is introduced to identify the extent of medium change; two speckle patterns were recorded before and after the medium

change, while the SLM pattern remained identical. As shown in Fig. 4.3(d), with increased variance, which suggests larger perturbations, both the correlation coefficient and η drop accordingly. Relative η is calculated by the ratio of η after medium change over the original η . New samples were generated from the altered transmission matrix to fine-tune the pre-trained RFOTNet. The simulated focusing recovery effect is shown in Fig. 4.3(c).

Fig 4.3. Simulation results. (a) Focused speckle pattern obtained in Step 2. (b) Speckle pattern after the medium change. (a) and (b) are recorded with the identical modulation pattern on the SLM. (c) The focused speckle pattern can be recovered after fine-tuning the pre-trained RFOTNet. (a), (b), and (c) use the same colormap. (d) Different levels of variances/perturbations result in different levels of scattering medium change. (e) The relationship between the degree of medium change and the number of required fine-tuning samples to recover the focusing.

Twenty different matrices were generated to model different levels of perturbations, and for each matrix, simulation was conducted five times. Results were averaged and shown in Fig. 4.3(e), which illustrates the number of fine-tuning samples required in order to recover the original focusing under different levels of medium change. When the correlation coefficient is around 0.6, the number of fine-tuning samples is only a quarter of that used for training RFOTNet in Step 1, and the original focusing performance can be recovered. When the correlation coefficient drops to 0.3 or even

lower, the number of fine-tuning samples is larger than the half required in Step 1. However, the number of fine-tuning samples increases sharply when the correlation coefficient drops below 0.2, which means little information of the original status is preserved. These simulation results are consistent with predictions from the correlation theory [83]. With the increase of perturbations, the correlation coefficient drops following a single-sided Gaussian distribution [84, 112], which is reflected in Fig. 4.3(d). As indicated by the speckle correlation theory, as environmental perturbations become stronger, the magnitude of $C_{nmn'm'}$ drops accordingly. In this case, the distinction due to the change is enlarged, and hence the fine-tuning of the pre-trained neural networks needs more samples and time to adapt itself to model the new situation. Nevertheless, so long as the correlation coefficient remains to be larger than 0.2, the number of samples for fine-tuning is convincingly smaller than the original number (10,000 samples).

In order to compare the performance of the proposed RFOTNet and conventional single-input-single-output CNN in light focusing and refocusing through scattering media with perturbations, simulation was conducted, and results are shown in Fig. 4.5. The structure of RFOTNet is illustrated in Fig. 4.2(a), while the structure of the conventional single-input-single-output CNN is shown in Fig. 4.4. The CNN consists of three convolutional layers and two fully connected layers. The input to the CNN is speckle patterns while the output is their corresponding SLM patterns. All convolutional (Conv 1, Conv 2, and Conv 3) and fully-connected layers (FC1 and FC2) have the same structure and parameters as the convolutional (Conv 1, Conv 2, and Conv 3) and fully-connected layers (FC1 and FC2) in RFOTNet, respectively. Training and fine-tuning settings are also the same.

Fig 4.4. The structure of a conventional single-input-single-output CNN. Input is speckle patterns and output is corresponding SLM patterns. The structure and parameters of all the convolutional layers and fully-connected layers in the CNN are the same as those in RFOTNet.

First of all, 10,000 samples were generated using a transmission matrix to train both networks. After training, the same focused speckle was sent to the pre-trained RFOTNet and CNN, and the focusing results with the SLM pattern predicted by them are shown in Figs. 4.5 (c) and (e), respectively. Note that Figs. 4.5c-f use the same color bar. As shown in Fig. 4.5 (a), RFOTNet demonstrated much higher converging speed than the conventional CNN during training, and the focusing result (Fig. 4.5c) is also much better. Then an environmental perturbation was modeled by adding another matrix to the original transmission matrix. After perturbation, the optical focusing was degraded. Then 3,000 fine-tuning samples were generated based on the new transmission matrix. During the fine-tuning process, RFOTNet still converged faster (Fig. 4.5b). Focusing recovery results by RFOTNet and CNN are shown in Fig. 4.5 (d) and (f), respectively. RFOTNet not only demonstrates superior ability in establishing the transformation from speckles to SLM patterns, it is also more powerful in adapting to the new medium state. In RFOTNet, Input 1 and Input 2 provide rich information about the scattering medium to facilitate the establishment of a more precise mapping from Input 3 to Output. Moreover, during fine-tuning, RFOTNet is better at discovering the differences before and after the perturbation, and adjusting the weights to model the new status.

Fig 4.5. Simulation results of RFOTNet and a single-input-single-output CNN in light focusing and refocusing. (a) Training performance of RFOTNet and CNN. Straight lines indicate the results of RFOTNet, while dashed lines show the results of CNN. Black and red lines show the training loss values (mean-squared-error) and validation loss values (mean-squared-error), respectively. (b) The performance of RFOTNet and CNN during fine-tuning. (c) Focused speckle obtained using RFOTNet after training. (d) Focusing recovery results with fine-tuned RFOTNet after medium change. (e) Focused speckle obtained using CNN after training. (f) Focusing recovery results with fine-tuned CNN after medium change. Note that (c)-(f) use the same color bar.

4.3.2. Experimental results

The experimental setup is illustrated in Fig. 4.6. Light emitting from a He-Ne CW laser (633nm, 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 the SLM (X13138-01, Hamamatsu). The light wavefront is modulated by the SLM, after which the light passes through two successive lenses and is focused onto the surface of a diffuser (120 Grit, Edmund ground glass diffuser, 83-419) by an objective lens (TU Plan Fluor 50X/0.80, Nikon). The light undergoes

multiple scatterings inside the diffuser, then the scattered light is collected by another objective lens (TU Plan Fluor 20X/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 size of speckle patterns recorded by the camera is 64×64 pixels. In experiments, 32 grey levels are used in SLM to represent phase values from 0 to 2π .

Fig 4.6. Schematic of the experimental setup. Light is expanded by two lenses (L1 and L2), and then half-waveplate (HW) and polarizer (P) adjust the polarization state of light incident onto the spatial light modulator (SLM). Light is modulated and reflected by the SLM, then passes through two lenses (L3 and L4) and is focused onto a diffuser (D) surface by an objective lens (OB1). Scattered light is collected by another objective lens (OB2) and recorded by a camera.

In experiments, in Step 1, 10,000 pairs of samples were collected for training, and the epoch was set to 10. The reason why 10,000 samples were used here is that, based on the training results with different numbers of samples, training RFOTNet with 10,000 samples could balance the trade-off between focusing performance and training cost including time and memory (Results will be presented in Section 4.4). Before training, the intensity of all collected speckles and SLM patterns are respectively normalized to between 0 and 1. Training time was approximately 2 minutes. After training, a targeted

speckle (shown as Fig. 4.7m) with a single speckle grain was sent to the RFOTNet, predicting an output, which was then loaded onto the SLM to modify the incoming optical wavefront. The final speckle pattern recorded using the predicted SLM pattern was shown in Fig. 4.7(a). As seen, light was focused to a single point with diameter ~ 30 μm , and the enhancement factor η was 64. After training, the RFOTNet has learned the information of the current medium state, thus is able to focus light through the diffuser. It is worth noting that although here light is only focused to a single position, the trained RFOTNet is capable to focus light to an arbitrary position on the image plane.

Then perturbations were applied to the scattering medium by mechanically moving the diffuser along one direction. After the movement, the speckle pattern also changed, as shown in Fig. 4.7(b). Note that Fig. 4.7a-l use the same colormap and scale. While the SLM pattern remained identical, the focusing performance was degraded as indicated by the appearance of more speckle grains in the background. The correlation coefficient between (a) and (b) is 0.5221, with which the η dropped from 64 to 27 (Fig. 4.7n, Group 1), only 42% of the original η . To recover the focal point, 3,000 pairs of SLM patterns and speckles were collected to fine-tune the pre-trained RFOTNet. The reason why 3,000 samples were chosen is that, based on simulation results, when the correlation coefficient is around 0.5, $\sim 3,000$ samples is required to recover the original focusing performance. Iterating 60 epochs took about 3 minutes. The target pattern was the same as the one used in Step 1. The SLM pattern predicted by the fine-tuned RFOTNet resulted in a new focus shown in Fig. 4.7(c), whose η is around 50, indicating that η has been improved by more than 85% compared to η of Fig. 4.7(b), and the background is significantly suppressed.

Fig 4.7. Experimental results of four different groups to achieve focusing recovery through the diffuser. (a)-(c) Experimental results of Group 1. (a) is the original focal pattern obtained using the RFOTNet trained in Step 1, (b) is the pattern after the perturbation, and (c) is the recovered focal pattern after fine-tuning the RFOTNet. (d)-(f), (g)-(i), and (j)-(l) are the speckle patterns corresponding to before change, after change, and after RFOTNet fine-tuning for Groups 2 to 4, respectively. (a) - (l) use the same colormap and scale. (m) The target speckle. (n) Intensity enhancement factor η of the focus of speckle patterns. Black squares indicate the η after Step 2, red circles indicate the η after the medium change, and blue triangles indicate the η of recovered focus after Step 4.

After Group 1 demonstration, experimental conditions were manually varied, and the whole experiment was repeated three more times from Steps 1 to 4. The number of samples in training and fine-tuning, the RFOTNet structure, and the training settings were all identical to the first experiment (Group 1). Results are illustrated in Fig. 4.7 as Groups 2-4, respectively. As given in Fig. 4.7(n), the focused speckle patterns had η of 49, 36, and 23, respectively, after Step 1. After the medium (diffuser) was moved

laterally by various distances, the focusing performance was obviously degraded (Figs. 4.7e, h, and k). More and more bright speckle grains appeared in the background, and eventually, the original focal point was submerged and hard to be distinguished (Fig. 4.7k). In these three groups, the correlation coefficients with the original patterns were 0.4822, 0.4252, and 0.6976, respectively, and the η dropped to 45%, 47%, and 43% of its original value, degrading to 22, 17, and 10, respectively. Then for each group, 3,000 samples were used to fine-tune each individual pre-trained RFOTNet. Focusing recovery results are shown in Figs. 4.7(f), (i), and (l), with η improved by 82%, 88%, and 120% over η before fine-tuning, reaching 40, 32, and 22, respectively. There is no doubt that after fine-tuning the RFOTNet, the clear focus is recovered from medium perturbations.

It is worth noting that the correlation coefficients before and after medium change in this study were all larger than 0.2, for which one may doubt whether it was within the traditional memory effect range [82]. However, retaining the original phase modulation pattern on the SLM can no longer keep light focused after the medium change in experiments, which is consistent with the finding reported by Benjamin *et al.* [112]. As known, if it is within the memory effect range, when the SLM pattern is shifted slightly, say, along a lateral direction, the resultant focal point shifts laterally accordingly, although the focal intensity decreases with the displacement distance, following a bell-shaped curve [112]. In the experiment, the maximal medium movement is smaller than $2.5\ \mu\text{m}$, which is much less than the individual element size of the camera ($6.5\ \mu\text{m}$) and the SLM ($12.5\ \mu\text{m}$). So the movement of the focal point is not obvious. But when the scattering medium is disturbed while the SLM pattern remains unchanged, the intensities of the focal point are significantly reduced, resulting in much more random speckle patterns that may even submerge the focal point. It is thus confirmed that using

traditional memory effect methods cannot keep the clear focal point after the medium change.

It also should be highlighted that if conventional iterative wavefront shaping [10] or transmission matrix methods [11] are applied to solve this problem, as discussed earlier, the whole optimization process has to be repeated from the beginning. Imagine the SLM pattern is 32×32 as in the experiment, 4,096 measurements are required to calculate the transmission matrix. Moreover, interference between the modulated light and reference light is required, which significantly increases the system complexity and reduces the efficiency of the SLM considering that part of the SLM pixels work as the reference. As for the iterative optimization methods, such as CSA, the phase value of each macropixel on SLM has to be cycled from 0 to 2π in multiple steps. As in the experiment, 2π is divided into 32 grey levels, which requires altogether 32,768 measurements to calculate the phase mask towards an optimum focusing. In comparison, using the proposed method, only 3,000 (*one order of magnitude less*) measurements are used to recover the focal point from disturbance. In addition, CSA optimizes each macropixel independently, the detected intensity improvement at the output plane is small, which may lead to errors in phase selection especially when the signal-to-noise ratio is low [113]. In contrast, the proposed approach optimizes all the SLM pixels together, leading to significant increases in signal-to-noise ratio and computational efficiency.

4.4. Discussion

Optical focusing plays a central role in many biomedical applications, which, however, is challenging at depths in tissues due to the strong scattering of light. In recent years, progress in wavefront shaping has been achieved to focus light tightly and

efficiently through or within thick scattering media. Most of these works (including CNN-based studies) [3, 10, 14, 21], however, have been limited to mechanically static samples as each optimization process is specified to only one state of the medium. When the medium is perturbed or starts to move, the quality of light focusing will degrade in terms of focal intensity and focal-to-background ratio. If the change is beyond the memory effect regime, with existing methods, a new time- and/or resource-demanding optimization process is required to recover the focusing. To tackle such a challenge, in this study, for the first time to the best of our knowledge, the fine-tuning of the pre-trained RFOTNet is proposed, and the degraded or even extinctive optical focus after perturbation can be recovered much more rapidly than the conventional methods.

In order to manifest the performance of the proposed method against conventional memory effect more clearly, experiments were conducted, and results are shown below. Fig. 4.8(a) is a focused pattern achieved using RFOTNet. Then the diffuser was moved left by $1\mu\text{m}$, resulting in the obvious degradation in focusing, as shown in Fig. 4.8(b). The correlation between (a) and (b) is 0.5303, while the η in (b) drops to only 32% of (a). The correlation between the original focused speckle (Fig. 4.8a) and speckles after shifting the diffuser is a function of translation distance, as shown in Fig. 4.8(d). The focal intensity also drops as the translation distance increases, following a bell shape [112], as illustrated in Fig. 4.8(e). Relative intensity is calculated against the focal intensity in (a). Despite that $1\mu\text{m}$ does not exceed the memory effect range, when the SLM pattern was shifted left accordingly, it resulted in a speckle pattern shown in Fig. 4.8(c), failing to recover the focal point. Therefore, a slight disturbance in the medium can lead to severe decrease in speckle correlation as well as focal intensity, while the traditional memory effect is no longer able to achieve a clear focal point. With the proposed method, as shown above, a clear focused speckle pattern can be reliably

recovered through fine-tuning the pre-trained RFOTNet.

Fig 4.8. Experimental results of focusing recovery with memory effect. (a) Original focused speckle obtained using the RFOTNet. (b) Speckle pattern after moving the diffuser left by 1μm while the SLM pattern is the same as (a). (c) Speckle pattern after the SLM shifted left accordingly. (d) Correlation between the original focus (a) and speckles after medium change (with the same SLM pattern) is a function of translation distance. (e) Focal peak intensity as a function of diffuser translation distance. Figs. a-c use the same colormap and scalar bar.

The feasibility of the proposed method has been clearly demonstrated, as shown in Fig. 4.7. After the medium change, the intensity enhancement factor η is degraded to different extents. Fig. 4.9 further illustrates how the drop of η is affected by the degree of medium movement. The original focus is the same, so the influence from the focusing performance has been eliminated; the correlation coefficient serves as the only convincing factor that describes the change of speckle patterns and measures the difficulty of focus recovery. As seen, when the correlation coefficient is larger than 0.72, the η drops to 0.7-0.9 of the original value. In this situation, medium change is mild, and the refocusing is relatively easier. When the correlation coefficient goes lower than 0.5, suggesting a larger medium change, the η drops to less than half of its original value. In this circumstance, recovering the focusing can be difficult. Generally speaking, larger

is the medium change, smaller is the correlation coefficient before and after the change, more reduction is the η , and more challenging is the focusing recovery.

Fig 4.9. The relationship between the intensity enhancement factor η remaining ratio and the levels of scattering medium change. The original focal point was the same, and the diffuser was moved by various distances to introduce different levels of perturbations. All results are scaled as the original η to be 1. Red dashed line illustrates the change of median values while blue dotted line shows the mean values.

Next, let's examine how the focusing recovery performance is related with the number of fine-tuning samples and the degree of medium change. Experiments were conducted five times, with averaged results shown in Fig. 4.10. Note that the original focal point was the same, and the η recovery percentage is measured by the ratio between the η after the RFOTNet fine-tuning and the original η before the medium change. 3,000, 5,000, and 6,000 samples were used to fine-tune the pre-trained RFOTNet, and the corresponding results are shown in grey, red, and blue colors, respectively. As seen, when the correlation coefficient is 0.7-0.8 (mild medium perturbation), 3,000 fine-tuning samples are enough to fully restore the original η . When the correlation coefficient is reduced to 0.4-0.5 and 0.3, respectively, 5,000 and 6,000 samples are required accordingly to recover the original performance. If the same number of tuning

samples is provided, the higher is the correlation coefficient, the smaller is the medium change, and the higher is the refocusing quality (as measured by η). Although there are some unavoidable interferences and/or noise in experiments, the experimental results agree well with simulation. Compared with existing methods such as iterative optimization and transmission matrix, the proposed method needs fewer measurements or simpler setup to realize focusing recovery. It is worth noting that under some circumstances, the η recovery percentage can be larger than 1, indicating that the focused speckle achieved by the fine-tuned RFOTNet may even be superior to the original pre-trained RFOTNet. This phenomenon manifests one attractive merit of the proposed method that the fine-tuning of a neural network is not only capable to adapt it to different medium status, it also can compensate the deficiency of the pre-trained network, improving it towards optimal performance.

Fig 4.10. The focusing recovery performance, as measured by the relative η , as a function of the number of samples in fine-tuning and the degree of medium change.

As mentioned above, the reason why 10,000 training samples is adopted is to balance the trade-off between computational time and light focusing performance. Herein, a detailed discussion is conducted. Different numbers of samples are offered for

RFOTNet training, and results are shown in Fig. 4.11. Except the number of samples, all the other parameters remain the same all the time. Epoch is fixed as 15. As the number of training samples enlarges from 6,000 to 14,000, both sample collection and training time increase linearly. By contrast, light focusing performance, which is represented by the η , is improved almost 20% when the number of samples is increased from 9,000 to 10,000, then η remains around 17 as training samples varying from 10,000 to 14,000. Therefore, according to the experimental results, choosing the number of training samples as 10,000 can balance the trade-off between computational cost and light focusing performance.

Fig 4.11. The intensity enhancement factor η achieved and training time cost as the number of training samples varies.

Lastly, since the performance of neural networks highly relies on the training samples, using more samples will definitely lead to better focusing and refocusing, at the expense of spending more time and memory in sample collection and training. So far, such trade-off between performance and time cannot be eliminated. Currently, the speed in the study is mainly limited by the slow frame rate of the liquid crystal-based

SLM, whose nominal frame rate is up to 60 Hz, but due to the existence of rising and falling response time, in experiments, the SLM operated at only ~ 6 Hz. Digital micromirror device (DMD) can be much faster (the settling time is only $18\ \mu\text{s}$ [114]), but it can only provide binary magnitude modulation for optical wavefronts. Thus, the focusing performance is inherently poorer; the η is only 1/5 of that with SLM when the same number of modulating megapixels are actively assigned [63]. In the meanwhile, to speed up the computation, a more powerful GPU and the adoption of FPGA are highly desired. In summary, the proposed method has demonstrated great potential in implementing wavefront shaping-based optical focusing through dynamic media, which could be of great significance to the field.

4.5. Summary

In conclusion, in this chapter, a deep learning framework is proposed to tackle the challenge of light focusing and refocusing through disordered media with perturbations. First of all, the proposed ROFTNet, which is implemented by a multi-input-single-output convolutional neural network, is trained to focus light through scattering media. Next, taking advantage of the speckle correlation theory and the fine-tuning properties of neural networks, once perturbations occur and the focused speckle is degraded, which is unavoidable in practice, only a small number of newly available samples from the changed environment are required to fine-tune the original trained model, and the focusing performance can be fast recovered. The feasibility of the proposed framework has been proven by both simulation and experimental results. Experimental results indicate that while 10,000 samples are used in training, only 3,000 fine-tuning samples are required when the correlation coefficient between media statuses before and after perturbation is larger than 0.4. It is also experimentally proved that the proposed

framework can still realize light refocusing in the circumstances where traditional memory effect fails.

The number of fine-tuning samples required to recover the original focusing is related to the levels of medium change. Besides, simulation comparisons between light focusing performance of the proposed ROFTNet and a conventional single-input-single-output CNN demonstrate that with the same training samples and hyperparameter settings, RFOTNet shows much higher converging speed than a conventional CNN during training and fine-tuning, and the light focusing and refocusing results are also much better.

In addition, the relationship between light focusing performance and the number of training samples is explored, trying to balance the trade-off between η and training cost. Moreover, the relationship between the drop of η and the degree of medium change, and how the focusing recovery performance is related with the number of fine-tuning samples at different degrees of medium change are both experimentally investigated, which can serve as guidelines for further research works. The proposed deep learning-empowered wavefront shaping framework has great potentials in facilitating optimal optical focusing and imaging in deep and dynamic tissues.

Chapter 5. Deep Learning-Empowered Wavefront Shaping through Changing Media

Chapter 4 aims at solving the problems of light focusing and refocusing through scattering media suffering from perturbations, which is inevitable in practice. In this chapter, another deep learning-empowered framework is put forward to resolve wavefront shaping problems in media that are changing continually, not just limited to perturbations at times. Specifically, the proposed framework is implemented by a network named as Timely-Focusing-Optical-Transformation-Net (TFOTNet), a Multi-Input-Single-Output deep Convolutional Long Short-Term Memory (ConvLSTM) network, and the framework effectively tackles the grand challenge of light focusing and refocusing through time-variant media without complicated computation. Same as before, benefitting from the speckle correlation theory, the introduction of fine-tuning of the pre-trained network efficiently deals with focusing degradation over time or due to inevitable environmental disturbance, permitting rapid recovery of an optical focus. Experimental results show that the iterations of training and fine-tuning processes with the proposed framework assure light focusing through time-variant media.

5.1. Theoretical analysis of implementing deep learning framework for light focusing and refocusing through nonstatic scattering media

5.1.1. Inverse scattering problems in time-variant environment

In Section 4.1, the inverse scattering problems in the static environment are analyzed, and the mathematic models of deep neural networks are also reviewed. However, in the circumstances of time-variant media, amendments to the previous models are required as the model is not only space-dependent but also time-dependent.

Fig 5.1. A brief illustration of light scattering through time-variant media. The phase patterns of light are modulated by a spatial light modulator (SLM), then light is focused onto a nonstatic scattering layer (S) by an objective lens (OB1). Scattered light transmitted the scattering layer is collected by another objective lens (OB2) and recorded by a camera. The time-variant scattering medium is created by a rotating diffuser.

In this scenario, at time t , a monochromatic optical wave field propagates from the source to a randomly changing scattering layer, and the transmitted scattered light is collected by a camera. Thus

$$I_c(t) = |E_c^{out}(t)|^2 = \left| \sum_a^N t_{ca}(t) E_a^{in}(t) \right|^2, \quad (5.1)$$

where $\mathbf{r}_a$ is the source plane (e.g., the SLM plane in Fig. 5.1), $E_a^{in}(t)$ and $E_c^{out}(t)$ are optical fields at $\mathbf{r}_a$ and $\mathbf{r}_c$ at time t , respectively [99]. $t_{ca}(t)$ is a complex transmission coefficient describing light propagation from $\mathbf{r}_a$ on the source plane to $\mathbf{r}_c$ on the receiving plane at time t . The effects of absorption are neglected. Thus, as analyzed in Section 4.1, iterative optimization with regularization is necessary [100] to resolve the inverse scattering problem at time t . The objective is to find a desired reconstruction $E^{in}(t) = W(t)p(t)$ that minimizes the cost function formulated as below:

$$\arg \min_{E^{in}(t)} \left\| I_c(t) - |H(t)E^{in}(t)|^2 \right\|^2 + \alpha(t)T(p(t)), \quad (5.2)$$

where $H(t)$ is the forward scattering model at time t , which is a matrix, relating transmitted light intensity $I_c(t)$ and the reconstructed electrical field $E^{in}(t)$, and $\alpha(t)T(p(t))$ is the regulation. $\alpha(t)$ is the regulation coefficient, $W(t)$ is a convolutional transformation, and $p(t)$ is the transformation coefficients at time t . In wavefront shaping,

$E^{in}(t)$ consists of N different input optical modes, $E^{in}(t) = [E_1^{in}(t) \ E_2^{in}(t) \cdots E_N^{in}(t)]$.

For dynamic media, media statuses at time t and $(t-1)$ are correlated, indicating that $H(t)$ is not only determined by the current situation, it is also influenced by its previous values:

$$H(t) = g\left(\beta_1^{t-1} H(t-1) + \beta_2^t x(t)\right), \quad (5.3)$$

where β_1^{t-1} and β_2^t are time-dependent parameters, while $g(\bullet)$ is a nonlinear function, $H(t-1)$ is the scattering model at time $(t-1)$, and $x(t)$ represents the information from current scattering media. Hence, in dynamic situations, Eq. (5.2) can still be solved using iterative algorithms based on the building block model, but with temporal information included in it, and $p(t)$ at the $(m+1)^{th}$ iteration is given below:

$$p^{m+1}(t) = \left\{ A_\theta \left[\frac{1}{L} W(t) * H(t) * I_c(t) + \left(I - \frac{1}{L} W(t) * H(t) * H(t) W(t) \right) p^m(t) \right] \right\} \\ = \left\{ A_\theta \left[\begin{array}{l} \frac{1}{L} W(t) * g\left(\beta_1^{t-1} H(t-1) + \beta_2^t x(t)\right) * I_c(t) \\ + \left(I - \frac{1}{L} W(t) * g\left(\beta_1^{t-1} H(t-1) + \beta_2^t x(t)\right) \right) p^m(t) \\ * g\left(\beta_1^{t-1} H(t-1) + \beta_2^t x(t)\right) W(t) \end{array} \right] \right\}, \quad (5.4)$$

where $L(t)$ is the Lipschitz constant at time t , $L(t) \leq \text{eig}(W(t) * H(t) * H(t) W(t))$. From Eq. (5.4), 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 media statuses is transmitted over time, and imposes influence on the inverse scattering model at time t . Eqs. (5.2)-(5.4) suggest that TFOTNet is suitable to solve 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 [115, 116]:

$$MISE = E \left\{ \int_0^T \iint_{S_a} [y_t(r_a) - \hat{y}_t(r_a)]^2 d^2 r_a dt \right\}, \quad (5.5)$$

where $y_t(r_a)$ and $\hat{y}_t(r_a)$ respectively denote 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 predicted and true phase values of incident optical modes $E_a^{in}(t)$, respectively.

As for the speckle correlation theory in dynamic media whose properties are time-variant, both spatial and temporal speckle correlation exist, thus the speckle correlation should be shown as below [117]:

$$\begin{aligned} C(t-t', r-r') &= \frac{\langle I(t, r) I(t', r') \rangle - \langle I(t, r) \rangle \langle I(t', r') \rangle}{\langle I(t, r) \rangle \langle I(t', r') \rangle} \\ &= C_1(t-t', r-r') + C_2(t-t', r-r') + C_3(t-t', r-r') \end{aligned} \quad (5.6)$$

According to the results reported by Feng *et al.* [118], 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 [105]. 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' [117]. The proposed framework encodes the correlation between medium statuses, propagating the information over time, as a consequence, an accurate inverse model can be constructed.

5.1.2. ConvLSTM layer

A ConvLSTM layer consists of a forget gate layer, input gate layer, and output gate layer. The forget gate passes the useful contextual spatial information gathered from

previous statuses, meanwhile, redundant features are “forgotten” [119]. The input gate allows the real-time interaction with the environment, while the output gate selects the information which is fit to output at the current state, speeding up convergence [120]. They work collaboratively, encompassing both long-term and short-term memories. For details about ConvLSTM network, please refer to [119, 121]. In short, parameters of a ConvLSTM layer are updated with the following equations [119]:

$$\begin{aligned}
i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\
f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\
C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c), \\
o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \\
H_t &= o_t \circ \tanh(C_t)
\end{aligned} \tag{5.7}$$

where $X_1, \dots, X_t$ are inputs, $C_1, \dots, C_t$ are cell outputs, and $H_1, \dots, H_t$ are hidden states from the beginning to time t . f_t , i_t , and o_t are the outputs of forget gate layer, input gate layer, and an output gate layer, respectively, at time t [119]. “*” denotes convolutional operator, while “ $\circ$ ” indicates the Hadamard product. The sharing of parameters within ConvLSTM network not only reduces the number of parameters so that the generalization of a ConvLSTM network can be improved, it also contributes to the increased accuracy of the model considering that both network outputs and previous features are used in optimization [122]. By contrast, CNNs update parameters independently at each iteration, resulting in severe redundancy due to the correlation among data at different time steps. The proposed TFOTNet takes advantage of the virtue of ConvLSTM cells, but it takes one step further from the conventional single-input-single-output networks, which will be clarified in the following section.

5.2. Working principle

The structure of the proposed TFOTNet is shown in Fig. 5.2. TFOTNet shares a similar structure with the RFOTNet presented in Chapter 4, but significant modifications are made in layers as temporal information is encoded in TFOTNet. TOFTNet has three inputs and one output. Inputs 1 and 2 are paired, while Input 3 and the output are paired. Referring to Fig. 5.2, light is firstly reflected by the spatial light modulator (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 is 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 Inputs 1 to Inputs 2, and it acts as a regularization term. Incorporating this regularization input into the TOFTNet, the targeted relationship from Inputs 3 to the output is obtained, and it is used to real-time resolve the inverse scattering problems based on the regularized cost function (Eq. 5.2). Input 3 is the speckle pattern desired to be seen by the camera after light passing through the scattering medium in experiments 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 [123]. Offering prior information to regularize the inverse problem can mitigate the burden in training, which plays a significant role in successfully resolving inverse problems [124, 125]. Besides setting analytic priors manually, it has also been reported that prior terms can be directly learnt during the training of neural networks, which is tailored to the statistics of the training images, indicating a stronger regularization [90, 126]. Chang *et al.* adopted the adversarial method to jointly train two networks that one offers prior information while the other one conducts inverse

projection [126]. Inspired by these, the proposed TFOTNet consists of two parts: prior knowledge about scattering provided by Input 1 and 2, and inverse mapping from Input 3 to output. Through training, the network learns to extract suitable priors from Input 1 and 2, and they are passed to facilitate the resolving of the inverse problem represented by Input 3 and output, alleviating the training burden and improving the modelling accuracy when compared with methods that directly learn the inverse mapping without any other knowledge.

Same as that in Chapter 4, when fine-tuning is needed, only the last layer of TFOTNet is adjusted while the weights of all the other layers are frozen. ConvLSTM1 and ConvLSTM2 consist of 16 and 32 filters, respectively, and the filter size of each layer is 7×7 and 5×5 with stride setting as 3×3 and 2×2 , respectively. ConvLSTM1 and ConvLSTM 3, ConvLSTM 2 and ConvLSTM 4, share the same structure and weights, respectively. The number of neurons in 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. 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 are set as hard sigmoid. The Tensorflow Keras library is used to construct the model. The computational unit is an Acer Predator G9-792, 16 Gb RAM, and a GTX 980M GPU. It is worth noting here that the lynchpin in solving nonstatic problems is the proposed framework, rather than any other specific implementations. Despite this, the introduced TFOTNet allows for more efficient modeling over a conventional single-input-single-output CNN (Comparisons will be given in the following section).

Fig 5.2. Illustration of the proposed deep learning-empowered 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 passing through the scattering medium in experiments 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 a SLM pattern that can obtain the target through the current medium. Inevitable environmental disturbance (disturbance) or the nonstationary 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. After tuning, the TFOTNet can adapt to the concurrent medium state and recover the optical focusing performance.

The working principle of the proposed framework is illustrated in Fig. 5.2. In the first step, samples are collected with the continually changing medium for TFOTNet training. Medium changing at a constant speed can be regarded as a stationary stochastic process. After training, the well-trained TFOTNet will be able to establish an accurate inverse scattering model of the altering medium. Then a preset focused (speckle) pattern is sent to the TFOTNet through Input 3 as the target, and the TFOTNet outputs the SLM pattern required in order to generate the focused target through the current system. To

cope with focus fading over time or due to environmental perturbations, ad hoc samples from the real-time medium are offered to fine-tune the TFOTNet obtained in the previous time, while the total number of fine-tuning samples should be much less than that used for retraining a network from random initialization. The fine-tuned TFOTNet will be able to produce an SLM pattern to recover focusing performance. Noted that during the whole process of fine-tuning, the scattering medium is allowed to be changing as it is. The iterations of training and fine-tuning processes ensure light focusing through nonstatic media.

5.3. Experimental results

The experimental setup is shown in Fig. 5.3, and most components are the same as those in Fig. 4.6. The major difference is that the diffuser is not fixed or translated by a large distance at times; instead, it is rotating mechanically at a constant speed to create a continually changing scattering medium. The size of SLM patterns is 32×32 , while the size of speckle patterns recorded by the camera is 64×64 pixels.

In Step 1, 10,000 pairs of samples are collected for training, during which period the diffuser is rotating mechanically at a constant speed, creating a successively changing scattering medium. In the experiments, seven scenarios (Group 1-7) are discussed with the diffuser rotating at different speeds. Media altering speed is characterized by speckles decorrelations time (SDT), which is defined as the time duration that intensity autocorrelation function decreases to $1/e$ of its initial value [65], as shown in Table 5.1. Meanwhile, the degree of media change before and after the total 10,000 sample collection is evaluated using the correlation of speckles recorded when sample collection begins and ends while the SLM pattern remains the same, listed in Table 5.1 as well. From Group 1 to 7, the decrease of SDT indicates the increase of medium

changing speed. Before training, the intensity of all collected speckles and SLM patterns are respectively normalized to between 0 and 1. The same samples are fed to both TFOTNet and CNN for training. Training with TFOTNet consumes approximately 75 seconds, while for CNN, nearly 35 seconds is cost. At first, Group 1-4 are sent for training. After training, a target (shown as Fig. 5.4d) with a single speckle grain is sent to both TFOTNet and CNN, predicting an output, which is then loaded onto the SLM to modify the incoming optical wavefront. The final speckle patterns recorded using the predicted SLM patterns through the current scattering system are shown in Fig. 5.4(a). Focused speckle patterns with TFOTNet are shown in the upper row, while CNN results are listed in the bottom row. All figures in Figs. 5.4(a) – (c) use the same colormap and scale. As seen, under these four conditions, the well-trained TFOTNet successfully focuses light to a single point with diameter $\sim 30 \mu\text{m}$. Light focusing performance is evaluated by peak-to-background-ratio (PBR), which is defined as the ratio between optimized spot and background [64]. From Group 1-4, the PBR is decreasing accordingly, reaching 38.5, 24.3, 23.4, and 13.7, respectively. Comparably, CNN is able to focus light in Group 1 and 2, with PBR achieving 27.5 and 20.8, but fails in Group 3 and 4. Such results suggest that faster changing scenarios pose higher difficulty in establishing statistical models with neural networks, in these cases, CNN, which is local in both time and space [120], fails to realize light focusing. TFOTNet has been proven to be able to more powerful at encoding both spatial and temporal information with learnable weights, demonstrating its superior ability in dealing with dynamic situations over a conventional CNN. It is worth noting that although in the reported experiments, light is focused to a single position, the trained TFOTNet is capable to focus light to an arbitrary position on the image plane.

Fig 5.3. Schematic of the experimental setup. Light is expanded by two lenses (L1 and L2), then a half-waveplate (HWP) and a polarizer (P) adjust the polarization state of light incident onto the spatial light modulator (SLM). Light is modulated and reflected by the SLM, then passes through two lenses (L3 and L4) and is focused onto the surface of a mechanically rotating diffuser (D) by an objective lens (OB1). Scattered light is collected by another objective lens (OB2) and recorded by a camera.

Table 5.1. Summary of experiment scenarios.

	Speckle decorrelation time/min (training phase)	Correlation of media before and after 10,000 training samples collection	Speckle decorrelation time/min (fine-tuning phase)	Correlation of media before and after 3,000 fine-tuning samples collection
Group 1	266	0.9383	-	-
Group 2	108	0.7884	-	-
Group 3	56	0.6196	-	-
Group 4	30	0.4161	-	-
Group 5	25	0.3470	16	0.5798
Group 6	19	0.2616	30	0.7670
Group 7	13	0.1349	4	0.1684

After Group 1-4 demonstration, samples from Group 5 are sent to train TFOTNet and CNN, and results are shown in Fig. 5.4(b). Neither network could build a precise model to achieve light focusing as the diffuser rotates much faster such that the correlation between medium statuses before and after 10,000 training sample collection is lower than 0.37, which is regarded as the loss of correlation [65]. The same phenomenon is observed in Groups 6 and 7. In this circumstance, the fine-tuning of the pre-trained network is introduced to compensate the deficiency of the pre-trained network, and improve focusing performance. Herein, 3,000 fine-tuning samples from the changing medium are used, less than 1/3 of the training data. During the fine-tuning

process, the diffuser is still rotating but at a faster speed than that in the training phase, as seen in Table 5.1. With TFOTNet, fine-tuning 15 epochs takes about 2 minutes; while for CNN, iterating 15 epochs costs nearly 30 seconds. The target pattern is the same (Fig. 5.4d). The SLM pattern predicted by the fine-tuned TFOTNet successfully results in a refocus, shown in the bottom row in Fig. 5.4(b), with the PBR reaching 29.7. In contrast, with CNN, even after fine-tuning, the time-varying relationship between speckles and SLM patterns has not been learnt, due to the lack of ability in encompassing long-range correlation, much less for situations that are worse than Group 5, i.e., much significant media change. Through fine-tuning, the ingestion of more data permits the improvement of the original insufficiently trained networks on-the-fly, keeping pace with the medium changing trends. The training phase can be treated as a good initialization for the successive fine-tuning process, while the fine-tuning effectively leads to the track of focusing through the time-variant media.

During all the above discussed cases, training and fine-tuning phases are conducted with similar diffuser rotating speeds. However, in practice, environmental disturbances can impose persistent influence on the scattering medium, leading to speeding up or slowing down of the medium changing speed. Given this, the scenario that media change at a decreased speed is discussed first, and results are shown in Fig. 5.4(b) as Group 6. First of all, 10,000 training samples indicated as Group 6 in Table 5.1 are sent to train a TFOTNet from random initialization. Without question, the pre-trained TFOTNet is not able to realize light focusing through the fast changing medium. Then, during the fine-tuning phase, the diffuser rotating speed is reduced, and the pre-trained TFOTNet is fine-tuned with another 3,000 samples. As seen, with the fine-tuned TFOTNet, a clear focal point is observed, and PBR has reached 45.8. A discussion about the number of fine-tuning samples and light focusing performance will be given in the following

content.

Fig 5.4. Experimental results. (a) Light focusing results through continually changing media (Group 1-4) with SLM patterns precited by a pre-trained TFOTNet (upper row) and a pre-trained CNN (bottom row). Medium changing speed in Group 1-4 is listed in Table 5.1. Light focusing performance is measured by PBR. (b) Light focusing results with trained (upper row) and fine-tuned (bottom row) TFOTNet and CNN under three circumstances (Groups 5-7). (c) Left image: the original focus obtained through a continually changing medium. Middle image: The degraded speckle pattern after the diffuser is suddenly rotated by a large angle while the SLM pattern is the same as the left one. Right image: The recovered focal pattern after fine-tuning the TFOTNet. During the fine-tuning phase, the scattering medium keeps changing at the same speed before the disturbance happened. All figures in (a) – (c) use the same colormap and scale. (d) The targeted focused speckle pattern.

Next, another situation that media changing is speeded up during the fine-tuning process is explored, and results are shown in Fig. 5.4(b) as Group 7. TFOTNet is trained with 10,000 samples denoted as Group 7 in Table 5.1, but the pre-trained model fails to focus light through the rapidly changing medium. Then the diffuser rotating speed is increased such that SDT is as low as 4 minutes. The light focusing result through the faster changing medium with TFOTNet which has been fine-tuned with 3,000 samples is shown in the bottom row. Surprisingly, although some speckles are observed in the

background, a focused speckle has appeared in the targeted position, with PBR reaching 12.6.

Besides the persistent environmental influence, transient perturbations may result in a dramatic alteration in the medium, thus a sudden drop in light focusing performance. Herein, the scenario that a continually changing medium is affected by some unexpected, transient, dramatic influence, such as vibrations, is explored, and results are shown in Fig. 5.4(c). A focal point achieved through a changing medium (left image in Fig. 5.4c) suffers from a sudden, dramatic disturbance induced by rotating the diffuser by a large angle, resulting in the consequent vanish of the focused speckle (middle image in Fig. 5.4c), and the PBR drops from 39.8 to 2.78. The correlation coefficient between speckles before and after the sudden change is as low as 0.15, exceeding the memory effect range [82]. After the perturbation, the diffuser still rotates at the same speed as before. 3,000 samples are fed to fine-tune the previously obtained TFOTNet which has achieved the focused speckle (left image) before the perturbation. The iteration epoch is 15. As shown in the right image in Fig. 5.4(c), after fine-tuning, light focusing has been successfully recovered through the nonstatic medium, with the PBR rising to 25.2. The results demonstrate that the proposed framework is not only able to model successively changing media, it can also deal with sudden, dramatic corruptions in light focusing, and rapidly recover it.

As stated above, the introduction of neural networks' fine-tuning engine enables the proposed framework to cope with focusing deterioration induced by environmental disturbances. Meanwhile, performance can also fade away with time as media are changing continually, as shown in Figs. 5.5 a-c. Fig. 5.5(a) is the original focus, while Fig. 5.5(b) and 5.5(c) are recorded at different time with the same SLM pattern. As expected, the longer the time lapses, the severer degradation in focusing performance

(lower PBR), the smaller the correlation with the original focus. Therefore, determining when to initiate the fine-tuning process and how many samples to be used in fine-tuning are crucial, governed by the trade-off between computational cost and light focusing performance, and they are influenced by various parameters, especially the medium changing speed. Generally, more fine-tuning samples leads to a higher PBR, at the expense of costing longer time, which will induce a larger medium change that may increase the difficulty in focusing recovery. Herein, six scenarios with diverse changing speeds are explored, denoted as Group A-F, and the fine-tuning process is initialized at different time as well, characterized by the correlation with the original focus, as listed in Table 5.2. Experimental results are shown in Fig. 5.4(d). After fine-tuning, the PBR in all cases are increased, suggesting better focused speckles. For Group A and B, their fine-tuning phase is initialized at the same condition (same PBR and same correlation with the original focus), but samples offered in Group B is doubled compared to fine-tuning samples used for Group A, considering that the SDT of Group B is nearly 1/5 of SDT of Group A, which suggests a much faster changing medium. After fine-tuning, Group A achieves 75% enhancement in PBR over the value before fine-tuning, while Group B only realizes 43% PBR improvement. Comparing Group D and F, with the same number of fine-tuning samples and similar fine-tuning initialization conditions (correlation coefficients with their original focus are similar), the recovered PBR achieved with Group F is 40 while Group D only increases the PBR to 26 after fine-tuning, which is only 65% of the value achieved by Group F. The SDT of Group D is only 60% of the SDT of Group F, and faster changing medium definitely leads to lower fine-tuned PBR. In addition, with Group E and F whose media are changing at similar speed, the recovered PBR in Group E (29.7) is only 75% of the fine-tuned PBR in Group F (40). In Group E, fine-tuning begins when PBR has dropped to 2.2 while Group F

initializes fine-tuning when PBR is 3.6, and the severer focus degradation leads to poorer fine-tuning results. For Group C where the medium is changing fastest, and fine-tuning is started with the most serious focus deterioration (the lowest speckle correlation), the recovered PBR will certainly be the lowest.

Table 5.2. Summary of six fine-tuning scenarios.

	Medium changing speed (characterized by SDT/ min)	Correlation with original focus when fine-tuning is initialized	Number of fine-tuning samples used in total
Group A	50	0.7494	1,000
Group B	9	0.7494	2,000
Group C	5	0.2192	3,000
Group D	8	0.4849	3,000
Group E	14	0.3470	3,000
Group F	15	0.4538	3,000

Fig 5.5. Experimental results of light focusing recovery using TFOTNet through a nonstatic medium under various conditions. (a) A focused speckle obtained through a continually changing medium. (b) and (c) The degraded speckle pattern recorded after different time while the SLM pattern is the same as (a). Speckle correlation with (a) are 0.7547 and 0.5521, respectively. (d) Light focusing recovery performance in the six scenarios with different medium changing speeds, fine-tuning initialization conditions, and fine-tuning samples.

5.4. Discussion

Optical focusing is significant yet challenging in many applications due to the strong scattering of light in heterogeneous media. Most wavefront shaping works (including

CNN-based studies) [3, 10, 14, 21] have been limited to static samples as the each optimization process is specified to only one state of the medium. To tackle such a challenge, for the first time to the best of our knowledge, a deep learning-empowered framework is proposed, which is implemented by TFOTNet, to realize light focusing and fast focusing recovery through nonstatic media. With the iterations of training and fine-tuning of TFOTNet, light can be safely focused through media that are continually changing, in spite of the inevitable environmental disturbance which can be significant enough to induce alterations that have exceeded the memory effect range. Each training can be treated as a good initialization for the successive fine-tuning process, leading to the optimization of focusing through the nonstatic media, and taking a large step towards functional in realistic applications.

In order to clearly manifest the performance of the proposed framework against conventional iterative wavefront shaping methods, experimental comparisons with widely accepted genetical algorithm (GA) [20] and continuous sequential algorithm (CSA) [19] are conducted, and the results are demonstrated in Fig. 5.6 and Fig. 5.7, respectively. During each experiment, the diffuser is rotating at a constant speed, while the speed varies in different experiments, leading to diverse degrees of media change within the same time interval. First of all, GA is tested with three different rotating speeds. The initial mutation rate R_0 , final mutation rate R_{end} , and decay factor λ are set as 0.01, 0.0025, and 650, respectively. The mutation rate R decreases with the increase of generations n to avoid over mutation, following $R = (R_0 - R_{end}) \times \exp(-n / \lambda) + R_{end}$ [20]. The SLM pattern size is set as 32×32 , the number of phase patterns used in each generation is 50, and totally 200 measurements are conducted. The optimization process takes approximately 27 minutes, which is almost the same as time cost in collecting 10,000 samples for neural network training. Light focusing results with GA under three

circumstances are shown in Figs. 5.6a-c. As before, speckle decorrelation time (SDT) is adopted to characterize medium changing speed, and in these three experiments, SDT is 53 minutes, 22 minutes, and 12 minutes, respectively. Moreover, the correlation of media statuses before and after the whole optimization process is 0.66, 0.31, and 0.11, respectively. As seen, the obtained focal point more or less deviates from the desired focused speckle which is supposed to be at the center of the image (Fig. 5.4d, indicated by the black dashed circle in Fig. 5.6). In general, the faster the medium changes, the longer distance the focal point shifts from the target position. In these three situations, the PBR achieved are 31.99, 13.26, and 4.85, respectively. Interestingly, if the diffuser is rotated back to the position that the whole optimization process starts while the SLM pattern remains the same, the focused speckle moves back to the center of the image, as shown in Figs. 5.6d-f, leading to the consequent increase of PBR to 41.75, 58.67, and 90.98, respectively. Figs. 5.6a-f use the same colormap and scale. As a stepwise optimization approach, GA is sensitive to initial seeds, thus GA may intrinsically favor the initial condition, which has been reflected by the experimental results. GA may be good at working with stationary media. However, it shows worse performance in changing media than the proposed framework, especially when the medium is altering fast.

CSA is also tested with the diffuser rotating at three different speeds, and results are shown in Fig. 5.7. The SLM pattern size is set as 32×20 , and only 8 grey levels ($n\pi/4$, $n=1, 2, \dots, 8$) are used with the SLM. It takes almost 20 minutes to run CSA one time. The SDT in these three experiments are 120 minutes, 27 minutes, and 9 minutes, leading to the correlation of media before and after the whole optimization process reach 0.89, 0.49, and 0.10, respectively. As shown in Figs. 5.7a-c, obvious shift of the focal point from the target position (indicated by the black dashed circle) is observed in all three

situations, similar to the results obtained with GA. The PBR achieved is 25.62, 14.94, and 2.87, respectively. In the second experiment, the diffuser is also rotated back to the position where the experiment begins while the SLM pattern remains the same, and the focus moves back to the center with PBR being enhanced to 48.42, as shown in Fig. 5.7(d). Figs. 5.7a-d use the same colormap and scale.

Fig 5.6. Experimental results of applying GA to focus light through media that are changing continually at various speeds. (a) – (c) Light focusing results with GA. From (a) to (c), the diffuser rotating speed is increased accordingly. (a) - (c) are recorded at the position where the whole optimization process ends. The black dotted circle indicates the position of target focused speckle. (d) – (f) Light focusing results when the diffuser is rotated back to the position where the experiment starts. (a) and (d), (b) and (e), (c) and (f), are recorded with the same SLM pattern, respectively. Figs. a-f use the same colormap and scale. (g) The PBR achieved by GA under the three conditions. PBR of focused speckles obtained at positions that experiment ends (gray square) and starts (red dot) are evaluated.

Fig 5.7. Experimental results of applying CSA to focus light through media changing continually at various speeds. (a) – (c) Light focusing results under three conditions. From (a) to (c), the rotating speed is increased accordingly. (a) - (c) are recorded at the position where the whole optimization process ends. The black dotted circle indicates the target position of the focused speckle. (d) Focused speckle pattern recorded at the position where the experiment starts. (b) and (d) are recorded with the same SLM pattern. Figs. a-d use the same colormap and scale. (e) The PBR achieved by CSA under the three conditions. The PBR at positions that the experiment ends (gray square) and starts (red dot) are evaluated.

Comparing all the experimental results achieved by TFOTNet, GA, and CSA, it can be deduced that when the scattering medium is changing moderately, TFOTNet, GA, and CSA demonstrate similar performance in light focusing. Nevertheless, with the increase of medium changing speed, the superiority of the proposed framework over GA and CSA is becoming more prominent, in terms of the light focusing accuracy and PBR. Moreover, the adoption of fine-tuning engine enables rapid focusing recovery from degradations caused environmental perturbations which can be drastic enough to exceed the memory effect regime.

With the experimental results reported so far, the proposed TFOTNet will collapse when the correlation between medium statuses when sample collection begins and ends, or before and after the disturbance, is below 0.1, even though 3,000 samples are offered for fine-tuning. It can be expected that expediting the sample collection process will alleviate the restriction since more samples can be obtained within a certain time interval,

hence, in theory, the proposed network can cope with faster changing media, or more significant environmental disturbances than the results reported in the article. Currently, the speed in the study is mainly limited by the slow frame rate of the liquid crystal-based SLM. As discussed above in Chapter 4, the employment of DMD can increase sample collection speed, at the expense of the reduction of light focusing performance.

Last but not the least, as indicated in the working principle, a focused speckle is generated in Matlab, and then sent to the well-trained neural network as a target. Hence, it can be expected that the properties of targets will affect network's predictions. Herein, the influence of target size on light focusing performance is experimentally investigated. Targets with three different diameters are generated in Matlab, shown in Figs. 5.8(a) - (c). From Fig. 5.8(a) to Fig. 5.8(c), the diameter of the focal point keeps increasing. Figs. 5.8(a) - (c) use the same colormap. Fig. 5.8(a), (b), and (c) is sent to the same pre-trained TFOTNet, respectively, and the experimental results are shown in Figs. 5.8(d), (e), and (f). Figs. 5.8(d) - (f) use the same colormap and scale. As seen, the diameter of the focal point in Fig. 5.8(d), (e), and (f) are $\sim 19\ \mu\text{m}$, $\sim 51\ \mu\text{m}$, and $\sim 64\ \mu\text{m}$, respectively, demonstrating the same increasing tendency in the focused speckle size as their targets. Moreover, the focusing performance in Fig. 5.8(e) and (f) are worse than that in Fig. 5.8(d) as in Fig. 5.8(e) and (f), speckles are both distorted to some extent. The PBR of Figs. 5.8(d), (e), and (f) show a decreasing tendency, reaching 15.1, 13.7, and 9.6, respectively. Therefore, in order to obtain a better focus and higher PBR, targets with smaller sizes are preferred.

It is worth noting here that wavefront shaping techniques are developed to overcome the resolution limit posed by multiple scattering, rather than diffraction. Therefore, only a diffraction-limited focal point can be obtained through the scattering medium with wavefront shaping techniques. Assuming that the light is monochromatic, and speckles

are fully developed. With a free space propagation geometry, the typical transverse and longitudinal dimension of a speckle grain are $\phi_s \sim \lambda \frac{z}{D}$ and $l_s \sim 7\lambda \left(\frac{z}{D}\right)^2$, respectively, where λ is the optical wavelength, z is the distance from the scattering object to the imaging plane, and D is the lateral dimension of the illuminated surface of the scatterer [27]. The dimensions of speckles are the same as those of the diffraction-limited focal spot of a lens with aperture D and focal distance z [27]. In order to overcome the diffraction limit, it is necessary to include other techniques to work with wavefront shaping methods, such as metasurfaces.

Fig 5.8. Experimental results with targets of different sizes. (a) – (c) Target focused speckles generated in Matlab. From (a) to (c), the diameter of the focused speckle keeps increasing. (a) - (c) use the same colormap. (d) – (f) Experimental results. Focused speckles obtained with pre-trained TFOTNet using (a) (b), and (c) as the target, respectively. (d) - (f) use the same colormap and scale.

5.5. Summary

In summary, this chapter concentrates on solving the problems of light focusing and refocusing through the scattering medium that is continually changing at a constant speed, which can be regarded as a stationary stochastic process. The proposed deep learning framework implemented by TFOTNet, which is a multi-input-single-output

ConvLSTM network, is able to recover light focusing performance from degradation induced by continual changes of media or strong environmental disturbances.

First of all, it is theoretically proved that TFOTNet is intrinsically suitable to tackle dynamic inverse scattering problems. Then experimental results show that the proposed TFOTNet-based deep learning framework can not only tackle the focusing and refocusing problems through media that are continually changing at a constant speed, it is also able to recover the optical focus from environmental perturbations which can be strong enough to exceed the memory effect range. As two representative conventional algorithms in wavefront shaping, both GA and CSA are investigated to focus light through continually changing media. However, neither of them successes in achieving a focal point at the target position; instead, the obtained focal point is shifted, and the faster the media change, the longer the deviation distance from the target position, suggesting that they are not suitable in dealing with nonstatic media. By comparison, the proposed framework can realize faster and accurate light focusing and refocusing in this situation, demonstrating great potential in practical applications.

Chapter 6. Deep Learning-Empowered Adaptive Recursive Framework for Dynamic Wavefront Shaping through Nonstationary Scattering Media

In Chapters 4 and 5, the wavefront shaping problems through media suffering from environmental perturbations and continually changing at constant speeds are solved. This chapter aims to tackle the grand challenge of light focusing and refocusing through randomly changing media without complicated computation. The randomly changing media can be treated as nonstationary stochastic processes, and a nonstationary process is regarded as consisting of multiple piece-wise stationary variations, while the statistical properties of each stationary stochastic process are analysed to guide fine-tuning. The introduction of adaptive adjustments of hyperparameters of the proposed Timely-Focusing-Optical-Transformation-Net(TFOTNet) effectively circumvents the drawbacks of traditional LSTM that it tends to remember only stationary variations [127]. 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 modelling spatiotemporal non-stationarity becomes possible. Another essential of the proposed framework is the recursive fine-tuning. It makes the best use of the correlation between media statuses before and after the change, which is indicated by the speckle correlation theory [16, 83, 105]. Therefore, only a small number of newly available samples is required to fine-tune the previous network, permitting fast recovery of the focusing performance. Note that the number of fine-tuning samples is even smaller than that reported in the previous two chapters. Although recursive fine-tuning has already allowed timely focusing recovery, an adaptive recursive estimation takes one step further, efficiently balancing the trade-off between fine-tuning time and refocusing performance, allowing controllable light delivery

through the time-variant scattering media. Note that during the whole phases, the medium is generally nonstationary; it keeps changing. 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 consumed in fine-tuning over traditional methods.

6.1. Working principle

The general working flow of the proposed adaptive deep learning-empowered framework and the structure of TFOTNet are the same as that shown in Fig. 5.2. Nevertheless, herein, the adaptive recursive algorithm is employed in fine-tuning to handle spatiotemporal non-stationarity, and it is elaborated in Fig. 6.1. In the whole chapter, 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 one of background [64]. Medium changing speed is characterized by speckle decorrelation time (SDT), which is defined as the time duration that intensity autocorrelation function decreases to $1/e$ of its initial value [65]. Smaller SDT corresponds to faster media altering speed. With the adaptive recursive algorithm, at time t , the SDT of the current medium is detected, and the PBR target, $Target_t$, is determined based on SDT. The methods to calculate the PBR target will be elucidated in Section 6.2. PBR target indicates the pre-defined PBR that targeted 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 the adaptive PBR target, a faster SDT accommodates to a relatively lower PBR target, which needs fewer fine-tuning samples, shortening the fine-tuning time. The instant PBR is compared with $Target_t$, and fine-tuning will not be initialized until the PBR is lower than $Target_t$. Pairs of SLM patterns and the corresponding speckle patterns are collected during the changing process of the scattering medium for fine-tuning, and the required number of fine-tuning samples and hyperparameters of a network are all chosen based on SDT. The influence of hyperparameters on fine-tuning time is discussed in Section 6.2. 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 adaptive recursive fine-tuning process, the optical focus can be recovered from deterioration in time, allowing for the maintenance of a focal point with acceptable performance through time-variant scattering media. 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 this chapter, such as PBR target and the number 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.

It is worth noting that the proposed TFOTNet is a general network that can be applied to deal with speckles and SLM patterns of arbitrary size. Herein, a specific implementation for the proposed typical setup is introduced as a proof-of-concept. The output size of TFOTNet is determined by the size of 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.

Fig 6.1. Flow chart of the proposed adaptive recursive algorithm for light focusing and refocusing through nonstationary media.

6.2. Simulation and experimental results

6.2.1. Simulation results

To simulate the scattering process, a transmission matrix $TX(t)$ is used to describe a disordered medium at time t , following circularly symmetric Gaussian distribution [20]. For a medium that is not static, medium status at time $t+\Delta t$ is represented by $TX(t)+\Delta TX(\Delta t)$, and $\Delta TX(\Delta t)$ also follows circularly symmetric Gaussian distribution. The size of the SLM patterns is set as 32×32 , while the size of speckle patterns is $64 \times$

64. First, in Step 1, 10,000 samples are created for TFOTNet initialization and training. Sample collection time is estimated using the maximal frame rate of a commercial LCOS-SLM, which is generally 60 Hz, and the SDT of the medium is as long as 10 minutes, resulting in the correlation between medium statuses when training sample collection initiates and ends reaching 0.8. After training, a focused target speckle (as shown in Fig. 5.4d) 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.

The training phase can be treated as a good initialization for the recursive fine-tuning process, while the introduction of adaptive adjustments will lead to the efficient tracking the focusing through time-variant media. Herein, 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 the 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 media 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 one second. 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. 6.2. From Fig. 6.2(a) to (d), it can be seen that when medium altering is mild and

SDT is larger than 10.8 s, varying one hyperparameter does not lead to significant differences in 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 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 epoch, 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.

Fig 6.2. The influence of hyperparameters on fine-tuning time cost to reach a pre-defined PBR when a 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 that the scattering medium is changing at different speeds (indicated by lines of different colors, and quantified by speckle decorrelation time).

Among all the hyperparameters evaluated above, the number of fine-tuning samples imposes the most significant influence on the fine-tuning time cost. A comparison about the required number of fine-tuning samples with and without adaptive adjustments of hyperparameters in the situations of different SDTs is shown in Fig. 6.3(a). As seen,

without adaptive modifications of hyperparameters, the required sample amount is 2~3 times of that needed by the adaptive algorithm to reach a pre-defined PBR target, resulting in much longer fine-tuning time cost. The results 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 totally 1,000 samples are used for fine-tuning to avoid overfitting. Increasing 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 changes to the medium status during this time period, which will degrade the fine-tuned PBR. With the goal of balancing the trade-off between overall PBR and fine-tuning time cost, the relationship between PBR after fine-tuning and fine-tuning samples amount using the adaptive algorithm when the medium changes at different speeds is explored, and results are shown in Figs. 6.3(b) and (c). In all cases, fine-tuning is conducted after a fixed time interval when the initial focused speckle is obtained, and it is chosen as one second as an example, which is scalable. With slow medium changes (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 s to 1.2 s, several hundred samples are needed to surpass the PBR threshold. With further expediting of the medium change (SDT is 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 to a relatively lower PBR target, less time is needed. With instructive results shown in Fig. 6.3, 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, $PBR\ target = [\max(PBR) + \min(PBR)]/2$, serving as a criterion to evaluate the focusing recovery performance. Six SDT intervals are used, and their corresponding PBR target is given in Table 6.1, serving as a criterion to evaluate focusing recovery performance. Although different setups result in different initial focused speckles, the ratio between 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.

Fig 6.3. (a) The required number of fine-tuning samples with and without adaptive adjustments of hyperparameters as the medium changes at different speeds (characterized by SDT). (b) - (c) The relationship between PBR after fine-tuning and the fine-tuning samples amount using the adaptive algorithm as the medium changes at different speeds.

Table 6.1. SDT dependent PBR target.

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

Proof-of-concept continuous nonstationary processes are simulated to clearly manifest the effect of the adaptive recursive fine-tuning engine in dealing with spatiotemporal non-stationarity. 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. Herein, ten nonstationary processes are randomly generated. Three fine-tuning algorithms, i.e., adaptive recursive algorithm, nonadaptive recursive algorithm, and traditional algorithm, demonstrate significant differences in the PBR recovering performance, which will be discussed separately in the following. First of all, a focused speckle with PBR achieving 41.5 has already been obtained using the original trained model. 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 time period. Simulated focusing recovery results with three algorithms using TFOTNet in ten nonstationary processes are shown in Figs. 6.4(a) - (j), where the same legends are used. For each stationarily altering sub-process, its PBR target (the ideal case) is indicated by yellow dash line in the figures. With the adaptive recursive algorithm, hyperparameters are selected on the basis of statistical properties of each stationary variation; with the nonadaptive recursive algorithm, the hyperparameters remain the default values (shown in Fig. 6.2) 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 engine of sensing intermediate processes, the hyperparameters remain as the default values, and total number of samples is the same as those in the recursive methods. Given space limitation,

the specific hyperparameters and the number of fine-tuning samples are listed in Appendix I. It can be seen from Figs. 6.4 (a) - (j) that, qualitatively, the PBR achieved by the adaptive recursive approach (grey line) is always the highest, while the traditional fine-tuning (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 analysing the statistical properties (including mean value and autocovariance) of diverse stationary variations, hyperparameters can be adjusted adaptively, compensating for the limitations of conventional ConvLSTM networks that they lack the capability to encode high-order nonstationary spatiotemporal variations [127].

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

$$SUM_M = \sum_{j=1}^M SDT(j) \times Duration(j), \quad (6.1)$$

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

$$Global \text{ focusing performance } (GFP) = \sum_{i=1}^N PBR(i) / N. \quad (6.2)$$

With the three algorithms, PBR at the time when each fine-tuning starts and ends are recorded, then all 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 ten nonstationary processes is illustrated in Fig. 6.4(k). Fig. 6.4(k) and (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 algorithm over the traditional one are 82% ~ 264% and 39% ~ 132%, respectively, which is calculated by

$$GFP \text{ enhancement percentage} = \left(\frac{GFP \text{ with adaptive or nonadaptive recursive algorithm}}{GFP \text{ with traditional algorithm}} - 1 \right) \times 100\% .$$

These results manifest the merits of the adoption of recursive approach that they can make the best use of both temporal and spatial correlation of media for fine-tuning, allowing much better focusing recovery. Meanwhile, the adaptive algorithm realizes 12% ~ 57% GFP enhancement over the nonadaptive recursive algorithm, confirming that through the adaptive adjustments of hyperparameters, spatiotemporal non-stationarity can be better learnt by neural networks. As for the traditional algorithm, the medium status at the time that fine-tuning is conducted has little correlation with the status at the time that 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:

$$Global \ tracking \ error(GTE) = \sqrt{\frac{\sum_{i=1}^N [PBR(i) - PBR \ target \ (i)]^2}{N}} . \quad (6.3)$$

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 ten nonstationary processes is illustrated in Fig. 6.4(l). As seen, the adaptive recursive algorithm shows the lowest

values (1.8 to 7.1), while the traditional algorithm shows the largest values (17 to 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

$$GTE \text{ drop percentage} = \left(1 - \frac{GTE \text{ with adaptive or nonadaptive recursive algorithm}}{GTE \text{ with traditional algorithm}} \right) \times 100\% .$$

As seen, the adaptive and nonadaptive recursive method reduces 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 of environmental disturbance, PBR may drop sharply (as seen in Figs. 6.4d-j), indicating that more time is required to recover PBR. Therefore, the increase of GTE in these circumstances is inevitable, as shown in Fig. 6.4(l). In situations without perturbations (as shown in Figs. 6.4a-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 note that the conclusions deduced above are effective for all implementations of the proposed deep learning-empowered adaptive recursive framework, not just limited to the setup and random nonstationary processes demonstrated in this chapter, and for all realizations, similar performance improvements can be obtained.

In Fig. 6.4(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 to make 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 modelling deficiency.

Fig 6.4. Fine-tuning results with ten random nonstationary processes using three different algorithms. One nonstationary process can be regarded as consisting of multiple piecewise stationary stochastic sub-processes, while the SDT and time duration of each sub-process are randomly chosen. (a) - (j) Fine-tuning results with the adaptive recursive algorithm (grey line), nonadaptive recursive algorithm (red line), and traditional fine-tuning algorithm (blue line) in the ten nonstationary process. Each process is characterized by SUM_M . Figs. (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. Figs. (k) and (l) use the same legend.

It is worth noting that the fine-tuning time can be significantly reduced and SDT can be much smaller than the 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-SLMs, which is 60Hz. As for computation, the Tensorflow Keras library is used,

and the computing unit is an Acer Predator G9-792, 16 Gb RAM, and a GTX 980M GPU. However, for instance, if a DMD is applied to conduct wavefront modulation whose frame rate has reached 23 kHz [63], together with onboard data acquisition, the fine-tuning time can be expedited by nearly 400 times. Furthermore, if a more powerful GPU or workstation such as Nvidia Tesla series is employed, the computation speed will be improved by at least 3 times. Thus, the fine-tuning process can be speeded up by nearly 1,000 times, indicating that the proposed framework becomes potential to achieve wavefront shaping in dynamic situations, such as *in vivo* tissues that decorrelate as fast as several milliseconds [128].

6.2.2. Experimental results

After the verification with simulations, experiments are conducted. The experimental setup is the same as Fig. 5.3. The resolution of the SLM screen is 1280×1024 , and it is divided into 32×32 macropixels to display the SLM patterns, i.e., one macropixel contains 40×32 pixels. The dimension of speckle patterns recorded by the camera is 64×64 pixels. The diffuser is rotating at different speeds to create various SDT. Due to the limitation of precision from the rotating stage (Motorized Precision Rotation Stage PRM1/MZ8, Thorlabs), the diffuser is rotated once as per 100 samples are collected, and the equivalent rotating speed varies from 2.5 mdeg/s $\sim$ 10 mdeg/s. Note that the nominal frame rate of the SLM (X13138-01, Hamamatsu) 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 system, the frame rate achieved in operation is only ~ 6 Hz. This has restricted the medium change speed demonstrated in the current phase of experiments. 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 experiments, the ratio between the PBR target and the initial PBR in a stationary stochastic process whose SDT is 200 seconds should be the same as that in the stationary variation whose SDT is 20 seconds 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 is 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, and it has been proved by simulation that hybridizing these two systems will demonstrate better performance in handling non-stationarity 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. Given space limitation, the specific hyperparameters and the number of fine-tuning samples used in experiments are listed in Appendix II. Experimental results are shown in Fig. 6.5 and Fig. 6.6. Results without environmental perturbations are shown in Figs. 6.6(a-c), while results with disturbance are given in Figs. 6.6(d-f). The same legend is used in these figures. Figs. 6.5(a-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. 6.6(a-c), under circumstances without sudden perturbations, the PBR target can always be reached after fine-tuning using the adaptive recursive algorithm (grey line), while the traditional algorithm (red line) never reaches the target. In these circumstances, the adaptive recursive algorithm always shows much better GFP (17 to 25) than traditional algorithm results (8 to 17). As seen from Fig. 6.5(c), with the adaptive recursive algorithm, 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. 6.5(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 statuses increases overall with SUM_M . The low statistical correlation between these two statuses induces degradation in the modelling accuracy, increasing the difficulty of focusing recovery. Moreover, sending the entire 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. 6.5(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 exacerbating 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. 6.5(d).

Fig 6.5. 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 the adaptive recursive algorithm and traditional algorithm. Figs. (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.

In situations where environmental disturbances occur, the experimental results are shown in Figs. 6.6(d-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 Fig. 6.5(b) and Fig. 6.4(l), the adaptive recursive algorithm still demonstrates much lower GTE in experiments (2 to 15) than that achieved by the traditional algorithm (13 to 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 improvements 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. 6.6(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 Fig. 6.4(i).

Actually, all the enhancement and reduction results obtained in the six experiments agree well with the simulations. As seen from Figs. 6.5(a) and (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 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.

Fig 6.6. (a) - (c) Results of the three experiments without environmental disturbance. The SDT of each stationary sub-process is shown in the figure, and the PBR target is indicated by yellow dashed lines. (d) - (f) Results of the three experiments with environmental disturbance. Figs. (a) - (f) use the same legend. (g) Speckle images recorded during a nonstationary process (Fig. f) using the adaptive recursive and traditional algorithm. All speckle images use the same colormap and scale, and are interpolated to 253×253 for a better view. The color bar indicates the detected light intensity in arbitrary unit. The middle image in the bottom row (traditional algorithm) is an interpolated result using recorded speckles.

The speckle images recorded in the nonstationary process indicated by Fig. 6.6(f) with the adaptive recursive and traditional algorithm are shown in Fig. 6.6(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 $\sim 30 \mu\text{m}$. Speckle patterns before and after each fine-tuning are demonstrated. As seen in Fig. 6.6(g), the adaptive recursive algorithm can recover the focal point in time, then the

focus can be remained over time. By contrast, with the traditional algorithm, due to the lack of ability in timely tracking, it cannot recover the focal point even though fine-tuning is conducted. It is worth noting that although in the reported experiments, light is focused to a single position, the trained TFOTNet is capable to focus 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, the middle image in the bottom row in Fig. 6.6(g) is an interpolated result using recorded speckles.

6.2.3. Comparison of light focusing and refocusing performance with TFOTNet, conventional ConvLSTM, and CNN

The ability of a conventional single-input-single-output ConvLSTM network [119, 121] and a single-input-single-output CNN in light focusing and refocusing through nonstationary scattering media are investigated using both simulation and experiments, and results are shown in Fig. 6.7. The structure of the conventional single-input-single-output ConvLSTM network and the single-input-single-output CNN are shown in Fig. 6.8(a) and (b), 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 structure with their corresponding ones in TFOTNet, including kernel size, the number of filters, activation function, *etc.* As for CNN, it consists of two convolutional layers, one fully-connected layer, and the last fully-connected layer serves as the output layer. Except timestep which is not included in CNN, all the other parameters are the same as the ConvLSTM network.

Fig 6.7. Comparisons about fine-tuning ability in a nonstationary process using three different networks. (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) are shown. All images use the same colormap and scale, and the color bars indicate the light intensity in arbitrary unit. (b) Experimental results. Light focusing and refocusing results using three networks with 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 unit. (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 unit. (d) Details of the nonstationary process and PBR with the three networks in simulation (Fig. a). (e) Details of the nonstationary process and PBR with the three networks in experiments (Fig. b). (f) Details of the nonstationary process and PBR with the three networks in experiment (Fig. c). These three figures (f-h) use the same legend.

Fig 6.8. (a) Structure of a conventional single-input-single-output ConvLSTM network. (b) Structure of a conventional single-input-single-output CNN.

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. 6.7(a). The PBR 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 increased by 50% compared with that needed by TFOTNet. Even though, 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 that as both temporal and spatial weights have to be learnt during training, a large number of samples are required. However, TFOTNet significantly enhances the modelling 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 Fig. 6.7(a) and (d), with the same nonstationary process and fine-tuning algorithm, TFOTNet always exhibits the best performance in light focusing and refocusing. Nevertheless, with ConvLSTM network or CNN, over time, the background has become 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 initialize and train TFOTNet, ConvLSTM, and CNN, and training results are shown in Fig. 6.7(b). With TFOTNet, 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 Fig. 6.7(b) and (e). With the 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 Fig. 6.7(c) and (f), among the three networks, only the proposed TFOTNet is able to obtain a focus after training even though the same training samples and parameters are used. In the situations indicated by Figs. 6.7(a) - (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 noise [129, 130]. As seen in Fig. 6.7(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 Fig. 6.7(a) and (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 noise than conventional single-input-single-output networks.

As mentioned above, the SLM used in experiments 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 recursive framework as much faster modulators can be employed. With a current commercially available DMD whose frame rate reached 23 kHz [63], both sample collection speed and the SDT of the altering medium can be improved by nearly 4,000 times, therefore, for the experimental results

shown in Fig. 6.6, 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 [128]. 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 to milliseconds variation media *in vivo* will be further work, and will be reported elsewhere.

6.2.4. 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 , and on average, the medium status changes per Δt . For the adaptive recursive algorithm, averagely, 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

$\left(KN^2 \times \frac{M_{total}}{M}\right) / F$ and $\left(4N^2 \times \frac{M_{total}}{M}\right) / F$, respectively, where N is the dimension of the

SLM patterns, K is the pixel grey level, and F is the frame rate of the SLM. $\frac{M_{total}}{M}$

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 fine-tuning is $\left(\frac{M}{F} + pt_p\right) \times \frac{M_{total}}{M}$. 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 p and time cost per epoch t_p . Considering that during 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, the number of fine-tuning samples, *etc.* It should note that taking a powerful GPU will reduce t_p , thus reducing $\lambda(N)$. As an example, with the proposed TFOTNet and reported computation unit (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 number of fine-tuning samples used here is 1,000, which is the largest sample amount used in the reported experiments, and timestep and batch size are set as 2 and 64, respectively. It can be expected that t_p will be further reduced if the number of fine-tuning samples is smaller. $\lambda(N) = t_p / N^2$ is calculated to vary from 9.2×10^{-6} to 5.9×10^{-3} . With the reported results, $N=32$, 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, are given below, based on the experimental setup reported in this chapter, 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 [19], 191 [131], or other values, and here K is set as 32 to be consistent with the reported experimental settings. Hence, more than 9 minutes is needed by CSA to complete an iterative optimization process. To measure a new transmission matrix that represents the changed medium status, nearly 70 seconds is required. As indicated by experimental results in Fig. 6.6, with the adaptive recursive algorithm, time spent in each fine-tuning varies from 6 s to 170 s with the frame rate of 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 seconds. 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 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 [113].

One promising application of the proposed framework is encryption. Recently, learning-based optical encryption has been reported [132] 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 much more parameters are required to precisely model the process. More importantly, the introduction of 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 the proposed 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 time 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 persistency is low between iterations, a new training or optimization cycle may be required. In this scenario, neither traditional iterative methods nor the proposed adaptive recursive framework can eliminate the trade-off.

6.3. Summary

In summary, the proposed deep learning-empowered adaptive recursive 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. With the proposed adaptive recursive fine-tuning of TFOTNet, optical focusing can be recovered in time from degradation, which is much more rapidly (with fundamental potential to achieve real-time) than 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 global focusing performance can be enhanced by 43% ~ 444% against the traditional algorithm, and the global

tracking error is reduced by 30% ~ 93%. Moreover, as the nonstationary process prolongs, both the global focusing performance enhancement percentage and the global tracking error reduction percentage over the traditional algorithm increase. It can be expected that similar performance improvements 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 is 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.

Chapter 7. Conclusion and Future Work

7.1. Conclusion

In summary, in this thesis, a comprehensive investigation on deep learning-empowered wavefront shaping is conducted, from theoretical analysis to implementations. A reinforced algorithm that hybridizes two or more complementary algorithms is proposed to increase wavefront shaping efficiency. Moreover, three deep learning frameworks are put forward to solve light focusing and refocusing problems through a scattering medium with perturbations, a medium that is continually changing at a constant speed, and a randomly altering medium, respectively.

Specifically, state-of-art wavefront shaping techniques are reviewed in Chapter 2, and their merits, as well as drawbacks, are analysed. Based on these, in Chapter 3, the idea of hybridizing complementary algorithms is proposed since they effectively compensate each other's weaknesses, leading to a more efficient hybrid algorithm. A systematical approach that combines the neural network and the genetic algorithm is put forward as a proof-of-concept, and improvements are experimentally verified in various aspects including computational efficiency and light focusing performance. As most pre-compensated wavefront shaping methods reported thus far only deal with static scattering media, in Chapter 4, a deep learning framework which is implemented by a specifically designed DCNN is proposed to solve light focusing and refocusing problems through the medium with perturbations. Next, Chapter 5 aims at realizing wavefront shaping through a medium that is continually changing at a constant speed with another deep learning framework which is implemented by a multi-input-single-output ConvLSTM network. At last, the grand challenge of light focusing and refocusing through a randomly altering medium are tackled with the introduction of

adaptive adjustments of neural network's hyperparameters and recursive fine-tuning engine.

Each method proposed in the thesis has its own merits and application scenarios. The hybrid algorithm proposed in Chapter 3 aims at increasing the efficiency of wavefront shaping algorithms, and it has been demonstrated that the proposed GeneNN can improve the focusing performance and optimization speed through a static scattering medium. The idea of hybridizing complementary algorithms can be extended to nonstatic scenarios. For example, the hybrid of the RFOTNet/TFOTNet and the dynamic mutation algorithm [133] will improve wavefront shaping efficiency through/within nonstatic media. The fine-tuning of RFOTNet demonstrated in Chapter 4 is proposed to cope with environmental perturbations, and it is a relatively simpler model compared with models used in randomly changing media. In situations where a medium is often subject to environmental perturbations, such as multimode fibers, the method proposed in Chapter 4 will be a potential solution. The deep learning framework implemented by TFOTNet in Chapter 5 is proposed to achieve light focusing and refocusing through a scattering medium that is changing continually at a constant speed. Naturally, the computational cost of TFOTNet is larger than RFOTNet since TFOTNet is made up of ConvLSTM cells which encode both spatial and temporal information. This framework is designed to deal with uniform changes. As for the adaptive recursive fine-tuning framework proposed in Chapter 6, investigations about the relationships between the hyperparameters and the fine-tuning cost as well as focusing performance have to be conducted, then the proposed framework can be employed to achieve adaptive light refocusing through a nonstationary scattering medium. Although the proposed framework is powerful in dealing with time-variant systems, it is much more complicated. In summary, the proposed methods in the thesis can be chosen based on

the specific implementations, with the goal to balance the trade-off between the computational cost and optimization performance.

In summary, this thesis puts forward an approach to improve wavefront shaping algorithm efficiency. Furthermore, for the kind of first time to the best of our knowledge, deep learning-empowered frameworks are proposed to tackle the challenge of optical focusing and refocusing through time-variant scattering media with wavefront shaping, demonstrating great potential in facilitating practical applications which require dynamic wavefront control.

7.2. Future work

Despite that promising results have been achieved in this thesis, the deep learning framework can be further improved in various aspects, such as computational speed and modelling accuracy. Moreover, the applications require more explorations, including extending the goal from light focusing to imaging, and applying the frameworks to other fields such as metasurfaces control. The perspective of the future work may concern the following aspects:

1. Inspired by the skyrocketing developments of reinforcement learning in recent years [134-139], it will be introduced to further improve the performance of the proposed framework. With reinforcement learning approaches such as Monte Carlo tree search [140-143] or deep Q-network [144, 145], it can be expected that light focusing and refocusing results can be further enhanced since reinforcement learning algorithms compensate the drawbacks of supervised learning that the optimization performance is influenced by the quantity and quality of training samples.
2. With the inspiring results of light focusing and refocusing achieved by the proposed adaptive deep learning framework, in the future, the work will be extended from

obtaining a focused speckle to forming clear images through dynamic media, which is of great significance in practical applications, especially in biomedical imaging.

3. Metasurfaces have received rapid developments as well as wide applications over the decades owing to their attractive properties such as anomalous reflection and refraction [146-148], polarization conversion [146, 149], and cloaking [150-152]. Metasurfaces have also been applied to wavefront shaping [153-156]. A metasurface serves as a good replacement for the commercial LCoS-SLM, which will lead to more accurate control of optical wavefronts.

Author's Publications

Journal papers:

1. **Luo, Y.**, Yan, S., Li, H., Lai, P., & Zheng, Y. (2020). Focusing light through scattering media by reinforced hybrid algorithms. **APL Photonics**, **5(1)**, 016109.
2. **Luo, Y.**, Yan, S., Li, H., Lai, P., & Zheng, Y. (2021). Towards smart optical focusing: Deep learning-empowered dynamic wavefront shaping through nonstationary scattering media. **Photonics Research**. Accepted.
3. **Luo, Y.**, Yan, S., Li, H., Lai, P., & Zheng, Y. (2019). Towards smart optical focusing: Deep learning-empowered wavefront shaping in nonstationary scattering media. **arXiv preprint arXiv:1909.00210**.
4. Li, H., Yu, Z., **Luo, Y. (Co-first author)**, Cheng, S., Wang, L., Zheng, Y., & Lai, P. Learning-based super interpolation and extrapolation for sub-Nyquist sampled optical speckles. **Submitted to Laser & Photonics Reviews**.
5. Li, H., Woo, C. M., Zhong, T., Yu, Z., **Luo, Y.**, Zheng, Y., & Lai, P. (2021). Adaptive optical focusing through perturbed scattering media with a dynamic mutation algorithm. **Photonics Research**, **9(2)**, 202-212.
6. Zhang, R., **Luo, Y.**, Jin, H., Gao, F., & Zheng, Y. (2020). Time-domain photoacoustic waveform analysis for glucose measurement. **Analyst**, **145(24)**, 7964-7972.
7. Zhang, R., Liu, S., Jin H., **Luo Y.**, Zheng, Z., Gao F., & Zheng, Y. (2019). Noninvasive Electromagnetic Wave Sensing of Glucose. **Sensors** **19(5)**, 1151.
8. Zhang, R., Gao, F., Feng, X., Liu, S., Kishor, R., **Luo, Y.**, & Zheng, Y. (2017). Noninvasive photoacoustic measurement of glucose by data fusion. **Analyst**, **142(16)**, 2892-2896.

9. Liu, S., Feng, X., Jin, H., Zhang, R., **Luo, Y.**, Zheng, Z., & Zheng, Y. (2019). Handheld photoacoustic imager for theranostics in 3D. **IEEE transactions on medical imaging**, **38(9)**, 2037-2046.
10. Liu, S., Feng, X., Gao, F., Jin, H., Zhang, R., **Luo, Y.**, & Zheng, Y. (2018). GPU-accelerated two dimensional synthetic aperture focusing for photoacoustic microscopy. **APL Photonics**, **3(2)**, 026101.
11. Liu, S., Zhang, R., **Luo, Y.**, & Zheng, Y. (2017). Magnetoacoustic microscopic imaging of conductive objects and nanoparticles distribution. **Journal of Applied Physics**, **122(12)**, 124502.

Conference Papers:

12. **Luo, Y.**, Li, H., Zhang, R., Lai, P., & Zheng, Y. (2019, February). Deep learning assisted optical wavefront shaping in disordered medium. **In Adaptive Optics and Wavefront Control for Biological Systems V (Vol. 10886, p. 1088612). International Society for Optics and Photonics.**
13. Zhang, R., Qu, Z., Jin, H., Liu, S., **Luo, Y.**, & Zheng, Y. (2018, May). Noninvasive glucose measurement by microwave biosensor with accuracy enhancement. **In 2018 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 1-4). IEEE.**

Bibliography

1. Bertolotti, J., et al., *Non-invasive imaging through opaque scattering layers*. Nature, 2012. **491**(7423): p. 232.
2. Shi, L. and R.R. Alfano, *Deep imaging in tissue and biomedical materials: using linear and nonlinear optical methods*. 2017: CRC Press.
3. Lai, P., et al., *Photoacoustically guided wavefront shaping for enhanced optical focusing in scattering media*. Nat Photonics, 2015. **9**(2): p. 126-132.
4. Xu, X., H. Liu, and L.V. Wang, *Time-reversed ultrasonically encoded optical focusing into scattering media*. Nat Photonics, 2011. **5**(3): p. 154.
5. Lai, P., et al., *Time-reversed ultrasonically encoded optical focusing in biological tissue*. J Biomed Opt, 2012. **17**(3): p. 030506.
6. Judkewitz, B., et al., *Speckle-scale focusing in the diffusive regime with time-reversal of variance-encoded light (TROVE)*. Nat Photonics, 2013. **7**(4): p. 300-305.
7. Lai, P., et al., *Reflection-mode time-reversed ultrasonically encoded optical focusing into turbid media*. Journal of Biomedical Optics, 2011. **16**(8): p. 080505.
8. Wang, Y.M., et al., *Deep-tissue focal fluorescence imaging with digitally time-reversed ultrasound-encoded light*. Nature communications, 2012. **3**: p. 928.
9. Yaqoob, Z., et al., *Optical phase conjugation for turbidity suppression in biological samples*. Nature photonics, 2008. **2**(2): p. 110.
10. Vellekoop, I.M. and A.P. Mosk, *Focusing coherent light through opaque strongly scattering media*. Opt Lett, 2007. **32**(16): p. 2309-11.
11. Popoff, S.M., et al., *Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media*. Phys Rev Lett, 2010. **104**(10): p. 100601.
12. Katz, O., E. Small, and Y. Silberberg, *Looking around corners and through thin turbid layers in real time with scattered incoherent light*. Nature photonics, 2012. **6**(8): p. 549.
13. Yu, H., et al., *Measuring large optical transmission matrices of disordered media*. Physical review letters, 2013. **111**(15): p. 153902.
14. Chaigne, T., et al., *Controlling light in scattering media non-invasively using the photoacoustic transmission matrix*. Nature Photonics, 2014. **8**(1): p. 58.
15. Takasaki, K.T. and J.W. Fleischer, *Phase-space measurement for depth-resolved memory-effect imaging*. Optics express, 2014. **22**(25): p. 31426-31433.
16. Katz, O., et al., *Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations*. Nature photonics, 2014. **8**(10): p. 784.
17. Katz, O., et al., *Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers*. Optica, 2014. **1**(3): p. 170-174.
18. Edrei, E. and G. Scarcelli, *Memory-effect based deconvolution microscopy for super-resolution imaging through scattering media*. Scientific reports, 2016. **6**: p. 33558.
19. Thompson, J., B. Hokr, and V. Yakovlev, *Optimization of focusing through scattering media using the continuous sequential algorithm*. Journal of modern optics, 2016. **63**(1): p. 80-84.
20. Conkey, D.B., et al., *Genetic algorithm optimization for focusing through turbid media in noisy environments*. Optics express, 2012. **20**(5): p. 4840-4849.
21. Turpin, A., I. Vishniakou, and J. d Seelig, *Light scattering control in transmission and reflection with neural networks*. Optics express, 2018. **26**(23): p. 30911-30929.
22. Drémeau, A., et al., *Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques*. Optics express, 2015. **23**(9): p. 11898-11911.
23. Menzel, W.P. and N. Satellite, *Remote sensing applications with meteorological satellites*. The Solar Spectrum, 2005. **3**(10).
24. Kortüm, G., *Single and Multiple Scattering*, in *Reflectance Spectroscopy*. 1969,

- Springer. p. 72-102.
25. Park, J.-H., et al., *Perspective: Wavefront shaping techniques for controlling multiple light scattering in biological tissues: Toward in vivo applications*. APL photonics, 2018. **3**(10): p. 100901.
26. Ntziachristos, V., *Going deeper than microscopy: the optical imaging frontier in biology*. Nature methods, 2010. **7**(8): p. 603.
27. Bossy, E. and S. Gigan, *Photoacoustics with coherent light*. Photoacoustics, 2016. **4**(1): p. 22-35.
28. Goodman, J.W., *Speckle phenomena in optics: theory and applications*. 2007: Roberts and Company Publishers.
29. Horstmeyer, R., H. Ruan, and C. Yang, *Guidestar-assisted wavefront-shaping methods for focusing light into biological tissue*. Nature photonics, 2015. **9**(9): p. 563-571.
30. Vellekoop, I.M. and A. Mosk, *Phase control algorithms for focusing light through turbid media*. Optics communications, 2008. **281**(11): p. 3071-3080.
31. Tzang, O., et al., *Wavefront shaping in complex media with a 350 KHz modulator via a 1D-to-2D transform*. Nature Photonics, 2019. **13**(11): p. 788-793.
32. Kong, F., et al., *Photoacoustic-guided convergence of light through optically diffusive media*. Optics letters, 2011. **36**(11): p. 2053-2055.
33. Hsieh, C.-L., et al., *Digital phase conjugation of second harmonic radiation emitted by nanoparticles in turbid media*. Optics express, 2010. **18**(12): p. 12283-12290.
34. Qiao, Y., et al., *Second-harmonic focusing by a nonlinear turbid medium via feedback-based wavefront shaping*. Optics Letters, 2017. **42**(10): p. 1895-1898.
35. Weiss, U. and O. Katz, *Two-photon lensless micro-endoscopy with in-situ wavefront correction*. Optics Express, 2018. **26**(22): p. 28808-28817.
36. Ruan, H., et al., *Focusing light inside scattering media with magnetic-particle-guided wavefront shaping*. Optica, 2017. **4**(11): p. 1337-1343.
37. Vellekoop, I.M., M. Cui, and C. Yang, *Digital optical phase conjugation of fluorescence in turbid tissue*. Applied physics letters, 2012. **101**(8): p. 081108.
38. Vellekoop, I.M. and C.M. Aegerter, *Scattered light fluorescence microscopy: imaging through turbid layers*. Optics letters, 2010. **35**(8): p. 1245-1247.
39. Vellekoop, I.M., et al., *Demixing light paths inside disordered metamaterials*. Optics express, 2008. **16**(1): p. 67-80.
40. Zhou, Y., J. Yao, and L.V. Wang, *Tutorial on photoacoustic tomography*. Journal of biomedical optics, 2016. **21**(6): p. 061007.
41. Ruan, H., M. Jang, and C. Yang, *Optical focusing inside scattering media with time-reversed ultrasound microbubble encoded light*. Nature communications, 2015. **6**(1): p. 1-8.
42. Ma, C., et al., *Time-reversed adapted-perturbation (TRAP) optical focusing onto dynamic objects inside scattering media*. Nature photonics, 2014. **8**(12): p. 931.
43. Zhou, E.H., et al., *Focusing on moving targets through scattering samples*. Optica, 2014. **1**(4): p. 227-232.
44. Xie, J., L. Xu, and E. Chen. *Image denoising and inpainting with deep neural networks*. in *Advances in neural information processing systems*. 2012.
45. Vincent, P., et al., *Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion*. Journal of machine learning research, 2010. **11**(12).
46. Zhang, K., et al., *Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising*. IEEE Transactions on Image Processing, 2017. **26**(7): p. 3142-3155.
47. Liu, D., et al., *When image denoising meets high-level vision tasks: A deep learning approach*. arXiv preprint arXiv:1706.04284, 2017.
48. Barbastathis, G., A. Ozcan, and G. Situ, *On the use of deep learning for computational imaging*. Optica, 2019. **6**(8): p. 921-943.

49. Sinha, A., et al., *Lensless computational imaging through deep learning*. Optica, 2017. **4**(9): p. 1117-1125.
50. Waller, L. and L. Tian, *Computational imaging: Machine learning for 3D microscopy*. Nature, 2015. **523**(7561): p. 416.
51. Goy, A., et al., *Low photon count phase retrieval using deep learning*. Physical review letters, 2018. **121**(24): p. 243902.
52. Deng, M., S. Li, and G. Barbastathis, *Learning to synthesize: splitting and recombining low and high spatial frequencies for image recovery*. arXiv preprint arXiv:1811.07945, 2018.
53. McCann, M.T., K.H. Jin, and M. Unser, *Convolutional neural networks for inverse problems in imaging: A review*. IEEE Signal Processing Magazine, 2017. **34**(6): p. 85-95.
54. Dong, C., et al., *Image super-resolution using deep convolutional networks*. IEEE transactions on pattern analysis and machine intelligence, 2015. **38**(2): p. 295-307.
55. Lim, B., et al. *Enhanced deep residual networks for single image super-resolution*. in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017.
56. Dong, C., et al. *Learning a deep convolutional network for image super-resolution*. in *European conference on computer vision*. 2014. Springer.
57. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning*. nature, 2015. **521**(7553): p. 436.
58. Li, Y., Y. Xue, and L. Tian, *Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media*. Optica, 2018. **5**(10): p. 1181-1190.
59. Li, S., et al., *Imaging through glass diffusers using densely connected convolutional networks*. Optica, 2018. **5**(7): p. 803-813.
60. Rahmani, B., et al., *Multimode optical fiber transmission with a deep learning network*. Light: Science & Applications, 2018. **7**(1): p. 69.
61. Jin, Y., et al., *Machine learning guided rapid focusing with sensor-less aberration corrections*. Optics express, 2018. **26**(23): p. 30162-30171.
62. Sun, Y., et al., *Image reconstruction through dynamic scattering media based on deep learning*. Optics express, 2019. **27**(11): p. 16032-16046.
63. Yu, Z., H. Li, and P. Lai, *Wavefront shaping and its application to enhance photoacoustic imaging*. Applied Sciences, 2017. **7**(12): p. 1320.
64. Strudley, T., et al., *An ultrafast reconfigurable nanophotonic switch using wavefront shaping of light in a nonlinear nanomaterial*. Light: Science & Applications, 2014. **3**(9): p. e207-e207.
65. Qureshi, M.M., et al., *In vivo study of optical speckle decorrelation time across depths in the mouse brain*. Biomedical optics express, 2017. **8**(11): p. 4855-4864.
66. Jang, M., et al., *Relation between speckle decorrelation and optical phase conjugation (OPC)-based turbidity suppression through dynamic scattering media: a study on in vivo mouse skin*. Biomedical optics express, 2015. **6**(1): p. 72-85.
67. Caravaca-Aguirre, A.M., et al. *Three-dimensional photoacoustic imaging through scattering media*. in *Biomedical Optics*. 2014. Optical Society of America.
68. Conkey, D.B., et al., *Super-resolution photoacoustic imaging through a scattering wall*. Nature communications, 2015. **6**: p. 7902.
69. Tang, J., R.N. Germain, and M. Cui, *Superpenetration optical microscopy by iterative multiphoton adaptive compensation technique*. Proceedings of the National Academy of Sciences, 2012. **109**(22): p. 8434-8439.
70. Yu, H., et al., *Recent advances in wavefront shaping techniques for biomedical applications*. Current Applied Physics, 2015. **15**(5): p. 632-641.
71. Fiolka, R., K. Si, and M. Cui, *Complex wavefront corrections for deep tissue focusing using low coherence backscattered light*. Optics Express, 2012. **20**(15): p. 16532-16543.

72. Jang, J., et al., *Complex wavefront shaping for optimal depth-selective focusing in optical coherence tomography*. Optics express, 2013. **21**(3): p. 2890-2902.
73. Yu, H., et al., *In vivo deep tissue imaging using wavefront shaping optical coherence tomography*. Journal of biomedical optics, 2016. **21**(10): p. 101406.
74. Si, K., R. Fiolka, and M. Cui, *Fluorescence imaging beyond the ballistic regime by ultrasound-pulse-guided digital phase conjugation*. Nature photonics, 2012. **6**(10): p. 657.
75. Ruan, H., et al., *Fluorescence imaging through dynamic scattering media with speckle-encoded ultrasound-modulated light correlation*. Nature Photonics, 2020: p. 1-6.
76. Choi, Y., et al., *Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber*. Physical review letters, 2012. **109**(20): p. 203901.
77. Plöschner, M., et al., *GPU accelerated toolbox for real-time beam-shaping in multimode fibres*. Optics express, 2014. **22**(3): p. 2933-2947.
78. Čižmár, T., M. Mazilu, and K. Dholakia, *In situ wavefront correction and its application to micromanipulation*. Nature Photonics, 2010. **4**(6): p. 388-394.
79. Kim, K. and Y. Park, *Tomographic active optical trapping of arbitrarily shaped objects by exploiting 3D refractive index maps*. Nature communications, 2017. **8**(1): p. 1-8.
80. Yoon, J., et al., *Optogenetic control of cell signaling pathway through scattering skull using wavefront shaping*. Scientific reports, 2015. **5**: p. 13289.
81. Ruan, H., et al., *Deep tissue optical focusing and optogenetic modulation with time-reversed ultrasonically encoded light*. Science advances, 2017. **3**(12): p. eaao5520.
82. Schott, S., et al., *Characterization of the angular memory effect of scattered light in biological tissues*. Optics express, 2015. **23**(10): p. 13505-13516.
83. Feng, S., et al., *Correlations and fluctuations of coherent wave transmission through disordered media*. Physical review letters, 1988. **61**(7): p. 834.
84. Li, J. and A. Genack, *Correlation in laser speckle*. Physical Review E, 1994. **49**(5): p. 4530.
85. Osnabrugge, G., et al., *Generalized optical memory effect*. Optica, 2017. **4**(8): p. 886-892.
86. Yang, X., Y. Pu, and D. Psaltis, *Imaging blood cells through scattering biological tissue using speckle scanning microscopy*. Optics express, 2014. **22**(3): p. 3405-3413.
87. Deán-Ben, X.L., et al., *Influence of the absorber dimensions on wavefront shaping based on volumetric optoacoustic feedback*. Optics letters, 2015. **40**(22): p. 5395-5398.
88. Yang, J., et al., *Focusing light through scattering media by polarization modulation based generalized digital optical phase conjugation*. Applied physics letters, 2017. **111**(20): p. 201108.
89. Xu, L., et al. *Deep convolutional neural network for image deconvolution*. in *Advances in Neural Information Processing Systems*. 2014.
90. Lucas, A., et al., *Using deep neural networks for inverse problems in imaging: beyond analytical methods*. IEEE Signal Processing Magazine, 2018. **35**(1): p. 20-36.
91. Bodenhofer, U., *Genetic algorithms: theory and applications*. 2003, Lecture notes, Fuzzy Logic Laboratorium Linz-Hagenberg, Winter.
92. Jin, K.H., et al., *Deep convolutional neural network for inverse problems in imaging*. IEEE Transactions on Image Processing, 2017. **26**(9): p. 4509-4522.
93. Kappeler, A., et al., *Video super-resolution with convolutional neural networks*. IEEE Transactions on Computational Imaging, 2016. **2**(2): p. 109-122.
94. Yosinski, J., et al. *How transferable are features in deep neural networks?* in *Advances in neural information processing systems*. 2014.
95. Phyu, S.P.T.P. and G. Srijuntongsiri. *Effect of the number of parents on the*

- performance of multi-parent genetic algorithm.* in *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*. 2016. IEEE.
96. Chen, T., et al., *A large population size can be unhelpful in evolutionary algorithms*. Theoretical Computer Science, 2012. **436**: p. 54-70.
 97. Roeva, O., S. Fidanova, and M. Paprzycki, *Population size influence on the genetic and ant algorithms performance in case of cultivation process modeling*, in *Recent Advances in Computational Optimization*. 2015, Springer. p. 107-120.
 98. Thierens, D. *Adaptive mutation rate control schemes in genetic algorithms*. in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*. 2002. IEEE.
 99. Vellekoop, I.M., *Controlling the propagation of light in disordered scattering media*. arXiv preprint arXiv:0807.1087, 2008.
 100. Wei, Z. and X. Chen, *Deep-learning schemes for full-wave nonlinear inverse scattering problems*. IEEE Transactions on Geoscience and Remote Sensing, 2018. **57**(4): p. 1849-1860.
 101. Chew, W.C. and Y.-M. Wang, *Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method*. IEEE transactions on medical imaging, 1990. **9**(2): p. 218-225.
 102. Chen, X., *Subspace-based optimization method for solving inverse-scattering problems*. IEEE Transactions on Geoscience and Remote Sensing, 2009. **48**(1): p. 42-49.
 103. Kamilov, U.S. and H. Mansour, *Learning optimal nonlinearities for iterative thresholding algorithms*. IEEE Signal Processing Letters, 2016. **23**(5): p. 747-751.
 104. Yuan, Z. and H. Wang, *Multiple Scattering Media Imaging via End-to-End Neural Network*. arXiv preprint arXiv:1806.09968, 2018.
 105. Sebbah, P., *Waves and imaging through complex media*. 2001: Springer Science & Business Media.
 106. Han, Y. and J.C. Ye, *Framing U-Net via deep convolutional framelets: Application to sparse-view CT*. IEEE transactions on medical imaging, 2018. **37**(6): p. 1418-1429.
 107. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
 108. Shin, H.-C., et al., *Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning*. IEEE transactions on medical imaging, 2016. **35**(5): p. 1285-1298.
 109. Christodoulidis, S., et al., *Multisource transfer learning with convolutional neural networks for lung pattern analysis*. IEEE journal of biomedical and health informatics, 2016. **21**(1): p. 76-84.
 110. Girshick, R., et al., *Region-based convolutional networks for accurate object detection and segmentation*. IEEE transactions on pattern analysis and machine intelligence, 2015. **38**(1): p. 142-158.
 111. Castrejon, L., et al. *Learning aligned cross-modal representations from weakly aligned data*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
 112. Benjamin Judkewitz, R.H., Ivo M. Vellekoop, Ioannis N. Papadopoulos and Changhuei Yang, *Translation correlations in anisotropically scattering media*. Nat Phys, 2015. **11**: p. 684-689.
 113. Fayyaz, Z., N. Mohammadian, and M.R. Avanaki. *Comparative assessment of five algorithms to control an SLM for focusing coherent light through scattering media*. in *Photons Plus Ultrasound: Imaging and Sensing 2018*. 2018. International Society for Optics and Photonics.
 114. Dudley, D., W.M. Duncan, and J. Slaughter. *Emerging digital micromirror device*

- (DMD) applications. in *MOEMS display and imaging systems*. 2003. International Society for Optics and Photonics.
115. Rueckert, E., et al. *Learning inverse dynamics models in $O(n)$ time with lstm networks*. in *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. 2017. IEEE.
 116. Marron, J.S. and M.P. Wand, *Exact mean integrated squared error*. The Annals of Statistics, 1992: p. 712-736.
 117. Breitzkreiz, M. and P.W. Brouwer, *Semiclassical theory of speckle correlations*. Physical Review E, 2013. **88**(6): p. 062905.
 118. Feng, S., et al., *Correlations and fluctuations of coherent wave transmission through disordered media*. Phys Rev Lett, 1988. **61**(7): p. 834-837.
 119. Xingjian, S., et al. *Convolutional LSTM network: A machine learning approach for precipitation nowcasting*. in *Advances in neural information processing systems*. 2015.
 120. Hochreiter, S. and J. Schmidhuber, *Long short-term memory*. Neural computation, 1997. **9**(8): p. 1735-1780.
 121. Xavier, A., *An introduction to ConvLSTM*. 2019.
 122. Qin, C., et al., *Convolutional recurrent neural networks for dynamic MR image reconstruction*. IEEE transactions on medical imaging, 2018. **38**(1): p. 280-290.
 123. Belfore, L., A. Arkadan, and B. Lenhardt, *ANN inverse mapping technique applied to electromagnetic design*. IEEE transactions on magnetics, 2001. **37**(5): p. 3584-3587.
 124. Li, L., et al. *Learning a discriminative prior for blind image deblurring*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
 125. Adler, J. and O. Öktem, *Solving ill-posed inverse problems using iterative deep neural networks*. Inverse Problems, 2017. **33**(12): p. 124007.
 126. Rick Chang, J., et al. *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.
 127. Wang, Y., et al. *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.
 128. Nissilä, I., et al., *Advances in Electromagnetic Fields in Living Systems, volume 4, chapter Diffuse optical imaging*. 2005, Springer.
 129. Kaufman, L., et al., *Measuring signal-to-noise ratios in MR imaging*. Radiology, 1989. **173**(1): p. 265-267.
 130. Welsh, B.M., *Speckle imaging signal-to-noise ratio performance as a function of frame integration time*. JOSA A, 1995. **12**(6): p. 1364-1374.
 131. Tay, J.W., et al., *Ultrasonically encoded wavefront shaping for focusing into random media*. Scientific reports, 2014. **4**: p. 3918.
 132. Zhou, L., Y. Xiao, and W. Chen, *Learning complex scattering media for optical encryption*. Optics Letters, 2020. **45**(18): p. 5279-5282.
 133. Li, H., et al., *Adaptive optical focusing through perturbed scattering media with a dynamic mutation algorithm*. Photonics Research, 2021. **9**(2): p. 202-212.
 134. Mnih, V., et al., *Human-level control through deep reinforcement learning*. nature, 2015. **518**(7540): p. 529-533.
 135. Silver, D., et al., *Mastering the game of Go with deep neural networks and tree search*. nature, 2016. **529**(7587): p. 484-489.
 136. Silver, D., et al., *Mastering the game of go without human knowledge*. nature, 2017. **550**(7676): p. 354-359.
 137. Littman, M.L., *Reinforcement learning improves behaviour from evaluative feedback*. Nature, 2015. **521**(7553): p. 445-451.
 138. Bueno, J., et al., *Reinforcement learning in a large-scale photonic recurrent neural*

- network. *Optica*, 2018. **5**(6): p. 756-760.
139. Mahmud, M., et al., *Applications of deep learning and reinforcement learning to biological data*. *IEEE transactions on neural networks and learning systems*, 2018. **29**(6): p. 2063-2079.
 140. Browne, C.B., et al., *A survey of monte carlo tree search methods*. *IEEE Transactions on Computational Intelligence and AI in games*, 2012. **4**(1): p. 1-43.
 141. Gelly, S. and D. Silver, *Monte-Carlo tree search and rapid action value estimation in computer Go*. *Artificial Intelligence*, 2011. **175**(11): p. 1856-1875.
 142. Guo, X., et al. *Deep learning for real-time Atari game play using offline Monte-Carlo tree search planning*. in *Advances in neural information processing systems*. 2014.
 143. Chaslot, G., et al. *Monte-Carlo Tree Search: A New Framework for Game AI*. in *AIIDE*. 2008.
 144. Hester, T., et al. *Deep q-learning from demonstrations*. in *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.
 145. Gu, S., et al. *Continuous deep q-learning with model-based acceleration*. in *International Conference on Machine Learning*. 2016.
 146. Chen, H.-T., A.J. Taylor, and N. Yu, *A review of metasurfaces: physics and applications*. *Reports on progress in physics*, 2016. **79**(7): p. 076401.
 147. Pendry, J.B., *Negative refraction makes a perfect lens*. *Physical review letters*, 2000. **85**(18): p. 3966.
 148. Smith, D.R., J.B. Pendry, and M.C. Wiltshire, *Metamaterials and negative refractive index*. *Science*, 2004. **305**(5685): p. 788-792.
 149. Kamali, S.M., et al., *A review of dielectric optical metasurfaces for wavefront control*. *Nanophotonics*, 2018. **7**(6): p. 1041-1068.
 150. Esfahlani, H., et al., *Acoustic carpet cloak based on an ultrathin metasurface*. *Physical Review B*, 2016. **94**(1): p. 014302.
 151. Yang, Y., et al., *Full-polarization 3D metasurface cloak with preserved amplitude and phase*. *Advanced Materials*, 2016. **28**(32): p. 6866-6871.
 152. Yang, Y., et al., *A metasurface carpet cloak for electromagnetic, acoustic and water waves*. *Scientific reports*, 2016. **6**: p. 20219.
 153. Chen, Y., X. Yang, and J. Gao, *Spin-controlled wavefront shaping with plasmonic chiral geometric metasurfaces*. *Light: Science & Applications*, 2018. **7**(1): p. 1-10.
 154. Jang, M., et al., *Wavefront shaping with disorder-engineered metasurfaces*. *Nature photonics*, 2018. **12**(2): p. 84-90.
 155. Veksler, D., et al., *Multiple wavefront shaping by metasurface based on mixed random antenna groups*. *ACS Photonics*, 2015. **2**(5): p. 661-667.
 156. Wang, Z., et al., *On-chip wavefront shaping with dielectric metasurface*. *Nature communications*, 2019. **10**(1): p. 1-7.

Appendix I

Table A.1 lists the hyperparameters and the number of samples used in each fine-tuning with the adaptive recursive algorithm. Here SUM_M is employed to characterize a nonstationary process. Each time fine-tuning is conducted, the same samples are offered to the adaptive and nonadaptive recursive algorithm, while the hyperparameters used with the nonadaptive algorithm remain as the default values, which are listed in Fig. 6.2. As for the traditional algorithm, samples collected for each fine-tuning during the whole nonstationary process are sent altogether to fine-tune the original trained model, while the hyperparameters also remain as the default values. The initial training results in the ten nonstationary processes are all the same.

Table A.1. Hyperparameters and the number of samples used in each fine-tuning with the adaptive recursive algorithm

	Number of Samples	Timestep	Batch size	Initial learning rate
$SUM_M=1408.95$				
1 st Fine-tuning	200	2	32	0.001
2 nd Fine-tuning	10	2	32	0.001
3 rd Fine-tuning	100	3	32	0.0025
4 th Fine-tuning	200	3	64	0.005
5 th Fine-tuning	20	3	64	0.001
6 th Fine-tuning	100	3	64	0.005
7 th Fine-tuning	1500	4	64	0.005
8 th Fine-tuning	200	3	64	0.005
$SUM_M=723.71$				
1 st Fine-tuning	300	2	32	0.001
2 nd Fine-tuning	10	2	16	0.0005
3 rd Fine-tuning	30	2	16	0.0005
4 th Fine-tuning	200	2	16	0.001
5 th Fine-tuning	400	2	32	0.001
6 th Fine-tuning	300	2	16	0.001
$SUM_M=487.4$				
1 st Fine-tuning	10	3	8	0.001
2 nd Fine-tuning	10	4	64	0.005
3 rd Fine-tuning	20	3	64	0.0025
4 th Fine-tuning	200	4	64	0.005
5 th Fine-tuning	400	3	64	0.0025
6 th Fine-tuning	100	4	32	0.0025

$SUM_M=447.96$				
1 st Fine-tuning	20	3	64	0.0025
2 nd Fine-tuning	10	4	64	0.001
3 rd Fine-tuning	200	3	64	0.0025
4 th Fine-tuning	10	4	32	0.0025
5 th Fine-tuning	30	3	32	0.0025
6 th Fine-tuning	400	3	32	0.0025
7 th Fine-tuning	100	4	32	0.005
$SUM_M=431.16$				
1 st Fine-tuning	10	3	8	0.001
2 nd Fine-tuning	100	3	32	0.0025
3 rd Fine-tuning	60	4	32	0.0025
4 th Fine-tuning	300	3	64	0.0025
5 th Fine-tuning	300	3	64	0.005
6 th Fine-tuning	200	3	32	0.0025
$SUM_M=369.14$				
1 st Fine-tuning	20	3	32	0.001
2 nd Fine-tuning	300	3	64	0.005
3 rd Fine-tuning	60	4	64	0.005
4 th Fine-tuning	100	3	32	0.005
5 th Fine-tuning	200	4	64	0.005
6 th Fine-tuning	200	4	64	0.0025
$SUM_M=334.35$				
1 st Fine-tuning	10	3	8	0.001
2 nd Fine-tuning	200	3	16	0.001
3 rd Fine-tuning	200	3	32	0.001
4 th Fine-tuning	100	4	32	0.0025
5 th Fine-tuning	100	3	32	0.0025
6 th Fine-tuning	200	4	32	0.005
$SUM_M=96.16$				
1 st Fine-tuning	10	3	32	0.0025
2 nd Fine-tuning	5	3	32	0.0025
3 rd Fine-tuning	20	3	64	0.0025
4 th Fine-tuning	10	4	64	0.0025
5 th Fine-tuning	20	2	32	0.001
6 th Fine-tuning	10	2	16	0.001
7 th Fine-tuning	5	2	16	0.0005
$SUM_M=61.96$				
1 st Fine-tuning	10	3	8	0.001
2 nd Fine-tuning	10	4	64	0.005
3 rd Fine-tuning	10	2	64	0.001
4 th Fine-tuning	10	4	64	0.005
5 th Fine-tuning	10	3	128	0.0025
$SUM_M=32.34$				
1 st Fine-tuning	10	4	64	0.005
2 nd Fine-tuning	20	2	32	0.001
3 rd Fine-tuning	10	3	64	0.0025
4 th Fine-tuning	10	3	32	0.0025
5 th Fine-tuning	10	4	32	0.001

Appendix II

Table A.2 and Table A.3 list the hyperparameters and the number of samples used in training and each fine-tuning with the adaptive recursive algorithm, respectively, in six experiments. Here SUM_M is employed to characterize a nonstationary process. With the adaptive recursive algorithm, the number of fine-tuning samples and hyperparameters are adaptively selected, allowing timely fine-tuning. As for the traditional algorithm, samples collected for each fine-tuning during the whole nonstationary process are sent altogether to fine-tune the original trained model, while the hyperparameters remain as the values used in training (Table A.2). The total fine-tuning samples used with the adaptive recursive algorithm and the traditional algorithm are the same.

Table A.2. Hyperparameters and the number of samples used in training

Number of Samples	Timestep	Batch size	Initial learning rate
10,000	4	16	0.0025

Table A.3. Hyperparameters and the number of samples used in each fine-tuning with the adaptive recursive algorithm

	Number of Samples	Timestep	Batch size	Initial learning rate
$SUM_M = 80.41 \times 10^3$				
1 st Fine-tuning	1000	2	16	0.001
2 nd Fine-tuning	50	2	16	0.001
$SUM_M = 42.94 \times 10^3$				
1 st Fine-tuning	1000	2	16	0.0025
2 nd Fine-tuning	30	4	32	0.001
3 rd Fine-tuning	100	3	16	0.001
$SUM_M = 37.98 \times 10^3$				
1 st Fine-tuning	50	2	16	0.001
2 nd Fine-tuning	1000	2	16	0.0025
$SUM_M = 4.16 \times 10^3$				
1 st Fine-tuning	30	4	32	0.005
2 nd Fine-tuning	50	4	16	0.001
3 rd Fine-tuning	30	3	16	0.001
4 th Fine-tuning	50	3	16	0.001
$SUM_M = 2.64 \times 10^3$				
1 st Fine-tuning	30	4	16	0.001

2 nd Fine-tuning	50	4	16	0.001
3 rd Fine-tuning	30	4	16	0.0025
$SUM_M=2.06\times 10^3$				
1 st Fine-tuning	30	4	16	0.001
2 nd Fine-tuning	50	4	16	0.001
3 rd Fine-tuning	50	4	16	0.001