



THE HONG KONG
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

**COMPUTATIONALLY ASSISTED DEEP-TISSUE
OPTICAL IMAGING AND PATTERNED LIGHT
DELIVERY**

SHENGFU CHENG

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University
Department of Biomedical Engineering

Computationally Assisted Deep-Tissue Optical Imaging and Patterned Light Delivery

Shengfu Cheng

*A thesis submitted in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy*

June 2024

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

_____ Shengfu Cheng (Name of student)

Abstract

Optical imaging has its unique values in biomedicine, as light is non-radiative and highly sensitive when interacting with biological tissues, offering structural and functional information with high spatial resolution. The biological tissues are, however, optically turbid with heterogeneous refractive index, causing multiple light scattering that prevents high-resolution optical focusing and imaging at depths in tissue. Over the past decades, many optical technologies have been developed that can overcome, reduce, or bypass strong light scattering for deep-tissue imaging. For example, wavefront shaping has been developed to overcome strong optical scattering and focus light through complex media, making it possible for high-resolution imaging within or through thick tissue. Photoacoustic imaging transforms the strongly scattered light into weakly scattered sound for detection, thus allowing a larger penetration depth with tunable spatial resolution. Fiber-optic endoscopic imaging can bypass optical scattering for in-situ microscopic observation with minimal invasion, especially when wavefront shaping is implemented with a multimode fiber (MMF) or fiber bundle. With photoacoustic imaging and the light control of MMF as the two primary technologies here, this thesis presents a comprehensive study aimed at pushing the boundaries of deep-tissue optical imaging and promoting light delivery related applications in complex media.

The research work of the thesis can be divided into three major parts. The first part (Chapter 2) is to explore deep-tissue optical-resolution photoacoustic microscopy (OR-PAM), which is based on image transformation on blurry yet “seeing-deeper” images of an acoustic-resolution microscopy (AR-PAM) through deep learning. The trained network successfully transformed AR-PAM images into high-resolution counterparts, with the lateral resolution improved from $54.0\ \mu\text{m}$ to $5.1\ \mu\text{m}$. Further, it significantly improved a mouse deep-brain image and retained high imaging resolution at tissue depths beyond one optical transport mean free path, which marks a big leap towards deep-tissue OR-PAM. The second part (Chapters 3 and 4) delves into the control of light delivery through MMF for deep-tissue applications based on wavefront optimization or retrieval of transmission matrix (TM). To be specific, by adopting a natural gradient ascent-based wavefront shaping strategy, coupled with a novel fitness function based on cosine similarity, we achieved high-quality pattern projection through a 15-meter-long unstable MMF. Apart from focusing, we also propose an alternating projection-based phase optimization method for suppressing scattered light in an arbitrary target region with a fast frame rate. In addition, we also study TM reconstruction from output intensity measurement by

developing a modified nonconvex optimization method, which demonstrates optimum efficiency with reduced running time or sampling ratio. The ability to retrieve a large-scale TM of MMF in a reference-less setup opens new avenues for deep-tissue optical imaging and manipulation, etc. In the last part (Chapter 5), we investigate MMF-based fluorescence imaging and especially explore the online calibration of MMF. We have confirmed the performance of point-scanning-imaging on the samples of microspheres and mouse brain slice via MMF calibration in the transmission mode. Besides that, a fluorescence -based intensity TM was proposed for reflection-mode calibration of MMF, which is feasible in principle but needs further experimental validation. These studies serve a common goal of advancing the field of deep-tissue optical imaging and focusing. Although the part of PA imaging seems to be less relevant to MMF-based focusing and imaging, the wavefront control methods we have developed may also be applied to enhance the focusing quality and AR-PAM imaging in deep tissues physically.

In summary, the thesis presents several computational and optimization methods that address key challenges faced by deep-tissue optical imaging and focusing, such as the limited resolution of AR-PAM and precise light control through complex media etc. Our research is valuable to deep-tissue PA imaging and MMF-based endoscopy, promising light delivery-related biomedical applications such as deep-tissue phototherapy and optogenetics. Through the above works, we have demonstrated that the integration of computational approaches with optical imaging technologies can unlock new potentials towards seeing deeper with greater clarity into biological tissues and even living bodies.

Publications Arising from the Thesis

Peer-reviewed journal papers:

- [1] **Shengfu Cheng**, Tianting Zhong, Chi Man Woo, Haoran Li, and Puxiang Lai*, “Lensless fiber-optic imaging via coherent light modulation and its applications (Invited)”, *Laser & Optoelectronics Progress*, 61(6), 0618002, 2024 (**Cover Paper**).
- [2] **Shengfu Cheng**[#], Xuyu Zhang[#], Tianting Zhong[#], Huanhao Li, Haoran Li, Lei Gong, Honglin Liu*, and Puxiang Lai*, “Nonconvex optimization for optimum retrieval of the transmission matrix of a multimode fiber”, *Advanced Photonics Nexus*, 2(6), 066005-066005, 2023.
- [3] Xuyu Zhang[#], **Shengfu Cheng**[#], Jingjing Gao, Yu Gan, Chunyuan Song, Dawei Zhang, Songlin Zhuang, Shensheng Han, Puxiang Lai*, and Honglin Liu*, “Physical origin and boundary of scalable imaging through scattering media: a deep learning-based exploration”, *Photonics Research*, 11(6), 1038-1046, 2023 (**Editor’s Pick**).
- [4] **Shengfu Cheng**[#], Tianting Zhong[#], Chi Man Woo, and Puxiang Lai*, “Alternating projection-based phase optimization for arbitrary glare suppression through multimode fiber”, *Optics and Lasers in Engineering*, 161, 107368, 2023.
- [5] **Shengfu Cheng**[#], Tianting Zhong[#], Chi Man Woo, Qi Zhao, Hui Hui*, and Puxiang Lai*, “Long-distance pattern projection through an unfixed multimode fiber with natural evolution strategy-based wavefront shaping”, *Optics Express*, 30(18), 32565-32576, 2022.
- [6] **Shengfu Cheng**[#], Yingying Zhou[#], Jiangbo Chen, Huanhao Li, Lidai Wang*, and Puxiang Lai*, “High-resolution photoacoustic microscopy with deep penetration through learning”, *Photoacoustics*, 25, 100314, 2022.
- [7] Zhipeng Yu[#], Tianting Zhong[#], Huanhao Li, Haoran Li, Chi Man Woo, **Shengfu Cheng**, Shuming Jiao, Honglin Liu, Chao Lu*, and Puxiang Lai*, “Long distance all-optical logic operations through a single multimode fiber empowered by wavefront shaping”, *Photonics Research* 12(3), 587-597, 2024.
- [8] Puxiang Lai^{#*}, Qi Zhao[#], Yingying Zhou, **Shengfu Cheng**, Man Woo Chi, Chuqi Yuan, Tianting Zhong et al., “Deep-Tissue Optics: Technological Development and Applications (Invited)”, *Chinese Journal of Lasers*, 51(1), 0107003, 2024.

-
- [9] Li, Huanhao[#], Zhipeng Yu[#], Qi Zhao[#], Yunqi Luo, **Shengfu Cheng**, Tianting Zhong, Chi Man Woo, Honglin Liu, Lihong V. Wang^{*}, Yuanjin Zheng^{*}, and Puxiang Lai^{*}, “Learning-based super-resolution interpolation for sub-Nyquist sampled laser speckles”, *Photonics Research*, 11(4), 631-642, 2023.
- [10] Zhipeng Yu[#], Huanhao Li[#], Tianting Zhong[#], Jung-Hoon Park[#], **Shengfu Cheng**, Chi Man Woo, Qi Zhao, YongKeun Park^{*}, Lihong V. Wang^{*}, Puxiang Lai^{*} et al., “Wavefront shaping: a versatile tool to conquer multiple scattering in multidisciplinary fields”, *The Innovation*, 3(5), 100292, 2022.
- [11] Qi Zhao[#], Huanhao Li[#], Zhipeng Yu[#], Chi Man Woo, Tianting Zhong, **Shengfu Cheng**, Yuanjin Zheng, Honglin Liu, Jie Tian^{*}, and Puxiang Lai^{*}, “Speckle-Based Optical Cryptosystem and its Application for Human Face Recognition via Deep Learning”, *Advanced Science*, 9(25), 2202407, 2022.
- [12] Tianting Zhong[#], Zhihai Qiu[#], Yong Wu, Jinghui Guo, Huanhao Li, Zhipeng Yu, **Shengfu Cheng**, Yingying Zhou, Jiejun Zhu, Jie Tian^{*}, Lei Sun^{*}, and Puxiang Lai^{*}. “Optically Selective Neuron Stimulation with a Wavefront Shaping-Empowered Multimode Fiber”, *Advanced Photonics Research*, 3(3), 2100231, 2022.

Conference presentations:

- [1] **Shengfu Cheng**, Tianting Zhong, Chi Man Woo, and Puxiang Lai, “Fast and arbitrary glare suppression through multimode fiber using alternating projection-based phase optimization”, *SPIE Photonics Asia 2023*, Beijing, China, 2023 (**Oral**).
- [2] **Shengfu Cheng**, Tianting Zhong, Chi Man Woo, and Puxiang Lai, “Arbitrary pattern projection through long and unfixed multimode fiber with wavefront shaping”, *SPIE Photonics West 2023*, San Francisco, USA, 2023 (**Poster**).
- [3] **Shengfu Cheng**, Tianting Zhong, and Puxiang Lai, “Non-convex optimization for retrieving the complex transmission matrix of a multimode fiber”, *2022 IEEE Region 10 Conference (TENCON)*, Hong Kong, 2022 (**Oral**).
- [4] **Shengfu Cheng**, Tianting Zhong, Li Jin, and Puxiang Lai, “A multimode fiber endoscopy utilizing fluorescence-based real-valued intensity transmission matrix”, *SPIE/-COS Photonics Asia 2021*, Nantong, China, 2021 (**Online**).

Patents:

- [1] Puxiang Lai, Tianting Zhong, Shengfu Cheng, and Chi Man Woo, “Method and apparatus for high-fidelity lensless multimode fiber-based photoacoustic endomicroscopy”, US patent, 2024 (**Submitted**).

Acknowledgements

My PhD study is about to finish this August, which signifies a phase of achievement resulting from the efforts of the past five years. The area of computational optical imaging is very attractive to me, as it is very comprehensive in involving theory, simulation, and experiment. I feel lucky and grateful for pursuing a PhD, from which I have greatly broadened my horizon and enriched my expertise. To me, obtaining a doctoral degree is never an end but marks a new starting point, as there are so many new things about machine vision and optical imaging to be explored in my next phase of life! At this moment, I am filled with a profound sense of gratitude towards many parties and individuals whose support, guidance, and encouragement have been my guiding stars.

First and foremost, I extend my deepest appreciation to the university and my supervisor, Dr. Puxiang Lai, for their cultivation and financial support. For Dr. Lai, his great curiosity and enthusiasm for research have impressed me and shaped my academic pursuits, his huge support and patience to me have always encouraged me to keep working especially when I got frustrated in my first project, his timely supervision on my research progress has pushed me forward in a slightly fruitful manner. Besides, I would like to thank my collaborators Dr. Lidai Wang in the City University of Hong Kong for his support and guidance on the project about photoacoustic imaging, and Dr. Honglin Liu in Shanghai Institute of Optics and Fine Mechanics for her guidance on imaging through scattering media. Also, I am deeply grateful to Dr. Cheng Wang in the City University of Hong Kong for his wonderful course on Photonics and his kindness of writing a recommendation letter for my postdoc application.

My heartfelt thanks also go to the lab mates, alumni, and fellow PhD students. Dr. Huan Hao Li, Dr. Zhipeng Yu, Dr. Yingying Zhou, and especially Dr. Tianting Zhong have always been very nice in guiding me in both study and life when I was in need. I have learnt a lot from them by working together in building optical systems, sharing ideas, and discussing research. I want to thank Miss Chiman Woo, Mr. Haoran Li, Mr. Jing Yao, and Dr. Xuandi Hou for their help in experiments and the beneficial discussions with them. Also, I am glad to have the company from other group members such as Mr. Haofan Huang and Mr. Yuandong Zheng that make my life more colorful. I cherish the bonds we have formed and look forward to witnessing the paths their bright minds will pave.

Lastly, I would like to take this chance to give my sincerest gratitude to my parents and family especially my grandpa who always support me and push me forward in my academic journey, and my dearest grandma who passed away in 2018 and should be proud of me in heaven. The love, encouragement, and belief from my family have been my stronghold, and without them I could not be who I am nowadays, for which I will do my best to reward them.

As I reflect on the past journey, I am reminded that the people we encounter, the challenges we overcome, and the support we receive truly shape our academic and personal growth. Thanks to everyone who has been a part of my PhD journey. And to myself, the attainment of a doctorate is not a symbol of prestige but reminds me to be modest and keep working. I have many shortcomings and should always have a keen awareness of time, cultivate hands-on skills, and foster innovation.

Contents

Certificate of Originality	ii
Abstract	iii
Publications Arising from the Thesis	v
Acknowledgements	viii
List of Figures	xiii
List of Tables	xix
1 Introduction	1
1.1 Light propagation in complex media	1
1.2 Optical wavefront shaping techniques	3
1.2.1 Iterative wavefront optimization	3
1.2.2 Transmission matrix	4
1.2.3 Relationship with adaptive optics	8
1.3 Deep-tissue optical imaging approaches	8
1.3.1 Photoacoustic imaging	10
1.3.2 Fluorescence imaging	12
1.3.3 Fiber-optic endoscopic imaging	13
1.4 Thesis structure	14
2 High-Resolution Deep-Penetrating Photoacoustic Microscopy through Deep Learning	17
2.1 Introduction	18
2.2 Methods	19
2.2.1 Integrated OR- and AR-PAM system	19
2.2.2 Sample preparation	20
2.2.3 Image pre-processing and data augmentation	21
2.2.4 GAN model and network training	23
2.2.5 Blind deconvolution for AR-PAM image deblurring	27
2.3 Results	28
2.3.1 Network feasibility: evaluation with <i>in-vivo</i> mouse ear photoacoustic images	28
2.3.2 Network performance: comparison and characterization	29

2.3.3	Network generalization: application for <i>in-vivo</i> mouse brain photoacoustic imaging	33
2.3.4	Network application: preliminary extension for deep-tissue OR-PAM	33
2.4	Discussion	36
2.5	Conclusion	39
3	Patterned Light Delivery through Multimode Fiber based on Wavefront Optimization	41
3.1	Introduction	42
3.2	Methods	44
3.2.1	Principle of SNES	44
3.2.2	Principle of AlterProjTM	45
3.3	Numerical simulations	48
3.3.1	Pattern projection under noise and perturbation	48
3.3.2	Comparison and analysis of suppression performance	50
3.4	Experimental results	52
3.4.1	Long-distance pattern projection through unstable MMF	53
3.4.2	Arbitrary speckle suppression through MMF	55
3.5	Discussion	56
3.6	Conclusion	57
4	Optimum Transmission Matrix Retrieval of a Multimode Fiber by Nonconvex Optimization	59
4.1	Introduction	60
4.2	Methods	62
4.2.1	Formulation of the TM retrieval problem	62
4.2.2	The modified RAF algorithm	63
4.3	Results	65
4.3.1	Accelerated gradient descent for RAF	65
4.3.2	The best two-step iteration ratio for RAF 2-1	66
4.3.3	Numerical evaluation of RAF and RAF 2-1	67
4.3.4	Numerical comparisons between RAF 2-1 and representative algorithms	68
4.3.5	Experimental results	71
4.4	Discussion	75
4.5	Conclusion	78
5	Pilot Study on Multimode Fiber-based Fluorescence Imaging with Transmission- and Reflection-mode Calibration Methods	79
5.1	Introduction	80
5.2	MMF fluorescence imaging methods	82
5.2.1	Experimental setup and techniques	82
5.2.2	Fluorescence-based TM recovery	85
5.2.3	Fluorescent object localization and reconstruction	87
5.3	Results	88
5.3.1	Scheme 1: MMF focusing performance characterization	88
5.3.2	Scheme 1: MMF point scanning fluorescence imaging	89
5.3.3	Scheme 2: NMF algorithm evaluation and rank estimation	90
5.3.4	Scheme 2: simulations of object reconstruction	92

5.3.5	Scheme 2: preliminary experimental results	94
5.4	Discussion	95
5.5	Conclusion	96
6	Summary and Future Work	98
6.1	Summary and contributions	98
6.2	Future work	100
	Reference	104

List of Figures

1.1	Illustration of light propagation in different complex media. (a) Typical distances for MFP and TMFP in biological tissues [1]. (b) Illustration of light propagation and diffusion in tissues [1]. (c) The refractive index distribution of a step-index MMF, the light propagation properties and the supported optical modes within it. (d) Schematic of light scrambling when passing through an MMF. Obj: objective.	2
1.2	Illustration of iterative wavefront optimization for focusing light through MMF with simulated results. (a) Schematic setup. SLM: spatial light modulator. (b) Output light intensity without phase modulation. (c) Output focus with optimal phase input. (d) Curves of focusing PBR versus iteration number.	4
1.3	Illustration of the TM measurement with on-axis or off-axis holography. (a) Schematic setup of on-axis TM measurement. (b) Speckle intensity produced by a flat reference phase (top-right inset). (c) Map of focusing PBR achieved by on-axis TM. (d) Schematic setup of off-axis TM measurement. (e) Spectral filtering of the -1st order for off-axis TM measurement. (f) Map of focusing PBR with an off-axis TM.	5
1.4	Illustration of solving the "dark spot" problem faced by on-axis TM using the complementary TM measured with multiple internal references. The reference speckle, PBR map and PR map by on-axis TM measured with (a) flat reference phase, (b) spiral phase (+1) and (c) spiral phase (-1). (d) The PBR map and PR map by the complementary TM, which is constructed by the above three TMs. Note η denotes focusing PBR and u denotes focusing uniformity in each PBR map.	6
1.5	Schematic diagram for the principle of (a-b) AO that corrects aberration for thin-tissue microscopy [2], and (c-d) WFS to form a focus within a thick sample through constructive interference among wavelets travelling along different optical paths [3].	9
1.6	Schematic illustrations of representative deep-tissue optical imaging approaches. (a) Imaging depth versus spatial resolution for various optical imaging modalities, reproduced from [4] with fiber-optic endoscopy (FOE) being supplemented. (b) Representative embodiments of confocal laser endomicroscopy (CLE) [5]. (c) Principle of photoacoustic imaging. (d) Typical photoacoustic imaging embodiments. (e-f) Two-photon fluorescence microscopy with (e) direct adaptive optics (AO) using a wavefront sensor and (f) indirect AO based on pupil segmentation [6].	10
1.7	Research framework of the thesis showing the research topic, challenge, imaging technologies, computational approaches, research contents and potential research values.	16

2.1	Schematic of the integrated OR- and AR-PAM system, with the optical (green) and ultrasonic (gray) beam path in the probe for OR- and AR-PAM illustrated separately. Note that the SMF and MMF are not connected to the probe at the same time but separately instead. AL, Acoustic lens; Amp, amplifier; DAQ, data acquisition; FC, fiber coupler; HWP, half-wave plate; MMF, multi-mode fiber; PBS, polarization beam splitter; SMF, single-mode fiber; UT, ultrasound transducer; WT, water tank.	21
2.2	The process of image patch extraction and alignment via correlation template matching. The OR patches were extracted successively, with each used as a template to find the highest correlated AR patch. The paired image patches were filtered with the criteria for vessel density and location before being applied with sub-pixel alignment.	23
2.3	The architecture of the WGAN model used for PAM imaging transformation. .	25
2.4	An experimentally obtained AR-PAM vascular image of mouse ear (a) is fed to the trained WGAN-GP model for imaging transformation. The resultant network output (b) is comparable to the ground truth OR-PAM image (c) of the same sample. Three ROIs marked with the white dashed boxes in (a1-a3) AR-PAM image, (b1-b3) transformed results, and (c1-c3) the ground truth OR-PAM image respectively, are enlarged and compared. Comparison of the cross-sectional profiles along the white dashed lines inside (a1, b1, c1), (a2, b2, c2), and (a3, b3, c3) are also provided in (d), (e) and (f) respectively. The blue arrow in (a-c) represents a vascular plexus that is originally blurred in AR-PAM but is now clearly resolved by the network; the green arrow shows a single capillary which is missed in the AR-PAM image and barely discernible in the network output, while clearly shown in the OR-PAM image; the purple arrow indicates a failure for the network to resolve some closely spaced parallel blood vessels that show up in the ground truth image.	29
2.5	Qualitative deblurring performance of deep learning in comparison with that using blind deconvolution. Two examples of mouse ear vascular images are presented, shown in the entire OR-PAM images (a) and (f). For each example, two ROIs marked with white dashed box are enlarged and compared. In the first example (a), (b1-e1) are for ROI 1 and (b2-e2) for ROI 2; in the second example (f), (g1-j1) are for ROI 1 and (g2-j2) for ROI 2; all correspond to AR-PAM, blind deconvolution, GAN output, and ground truth, respectively. Cross-sectional profiles along the white dashed lines inside (b1- e1), (b2-e3), (g1-j1) and (g2-j2) are provided for comparison. Metrics like PSNR and SSIM with respect to the entire OR-PAM image are also provided for reference.	31
2.6	Demonstration of lateral resolution enhancement of AR-PAM by deep learning. Lateral resolution of (a) AR-PAM, (b) network output, and (c) OR-PAM was measured to be $\sim 54.0 \mu\text{m}$, $\sim 5.1 \mu\text{m}$, and $\sim 4.7 \mu\text{m}$, respectively, using the edge response of a sharp blade. ESF, edge spread function; LSF, line spread function. The color insets are the blade images of AR-PAM, network output, and OR-PAM, respectively.	32
2.7	Application of the network on <i>in-vivo</i> mouse brain AR-PAM images. (a) is the network input (AR-PAM image) and (b) is the network output. Two ROIs in both network input (a1, a2) and output (b1, b2) are enlarged and shown. Comparison of signal intensity profiles along the horizontal dashed line in (c) the first ROI and (d) the second ROI are also given.	34

2.8	Preliminary demo for deep-penetrating OR-PAM imaging using a hair phantom covered with chicken tissues of different thicknesses. (a1-a4) Experimentally acquired AR-PAM images. (b1-b4) Network output results corresponding to tissue thicknesses of 0 (no tissue covered), 700, 1300, and 1700 μm . The white and green dashed boxes in (a1-a4) and (b1-b4) denote the ROIs for object and background, respectively, and the cyan lines indicate the positions for measuring the hair edge 10-90% rise distance. (c) CNR and (d) measured hair width versus different penetration depths (tissue thicknesses) for both AR-PAM images and the network output results.	36
3.1	Illustration and simulated results of pattern projection through an MMF with iterative wavefront optimization. (a) Schematic of focusing light into a desired pattern at the distal end of MMF. (b) Diagram of the principle of SNES. (c) Simulated progression of PCC as a function of iteration number for different optimization algorithms for the projection of "PolyU" pattern, under different relative persistence times.	45
3.2	Principle and simulated results of speckle suppression at a target region of MMF output plane. (a) Illustration of alternating projection between the phase and speckle field of an MMF. (b) Schematic of the proposed AlterProjTM for phase optimization. Simulated results of glare suppression in a target region of 30×30 , using AlterProjTM with (c) ER and (d) HIO constraints. (e) Normalized curves of the total intensity in the target regions. (f) Curves of suppression factor as a function of iteration number.	46
3.3	Theoretical comparison and performance analysis of the proposed AlterProjTM for glare suppression. Curves of suppression factor η_s as a function of iteration number in square ROIs of different side lengths, when using (a) GA, and AlterProjTM with (b) ER and (c) HIO constraints, respectively. Note that the subplots inside (b) and (c) show the zoomed-in results within the first 30 iterations. (d) Suppression factor η_s as a function of the modulation unit number for square ROIs of different side lengths, using AletrProjTM with (d) ER and (f) HIO constraint after 1000 iterations.	51
3.4	Experimental setup for glare suppression through an MMF. The complex TM of the MMF is measured via off-axis holography, and a DMD is used for phase-only modulation. CMOS, complementary metal-oxide-semiconductor camera; DMD, digital mirror device; FC, fiber collimator; L, lens; M, mirror; MMF, multimode fiber; NPBS, non-polarizing beam splitter; Obj, objective; PBS, polarizing beam splitter; $\lambda/2$, half-wave plate; $\lambda/4$, quarter-wave plate.	53
3.5	Results of pattern projection through a 15-meter-long unfixed MMF, using different optimization methods for comparison. Projection results for the target pattern of (a) 6×6 focus array, (b) pentagram, and (c) simplified Bagua are shown on the first, second, and third row, respectively, with the corresponding contrast value noted in each realization. Note that for each realization, the intensities of experimentally acquired images are individually normalized and have the same color bar for comparison of contrast. The scale bar shown in the last images of (a-c) is 10 μm	54

3.6	Normalized speckle patterns showing glare suppression in ROIs of various sizes at the distal end of the MMF, using different algorithms for comparison. The first row and second row correspond to square regions of 40×40 pixels and 100×100 pixels respectively, with (a, e) for the initial speckle pattern under zero phase modulation, (b, f) for those optimized with GA after 1200 iterations, (c, g) for those optimized with ER after 30 iterations, and (d, h) for those optimized with HIO after 30 iterations. The scale bar shown in (a-h) is $10 \mu\text{m}$	55
4.1	(a). Normalized curves of error as a function of running time for RAF with a fixed step size (μ) or an adaptive one. (b) Normalized curves of error as a function of running time for RAF with steep descent (SD), non-linear conjugate gradient (NCG), or L-BFGS.	66
4.2	(a). Normalized focusing PBR achieved by GGS 2-1 and RAF 2-1 under a series of iteration ratios during their two-step gradient iterations, respectively. . . .	67
4.3	Normalized curves of error as a function of running time for RAF and RAF 2-1 when $N = 576$ and $\gamma = 4$. Note the error bars denote the standard deviations of 30 repeated tests.	68
4.4	Theoretical comparisons of TM retrieval performances of prVBEM, GGS 2-1 and our RAF 2-1. (a) Schematic diagram of TM retrieval for a MMF. (b) Focusing efficiency achieved by different algorithms under a range of running times when $N = 576$ and $\gamma = 4$. (c) The iterations taken by different algorithms versus running times for the case of (b). (d) Focusing efficiency achieved by different algorithms under a range of γ when using $N = 576$ and a running time of 20 seconds. Note the error bars denote the standard deviations of 30 repeated tests.	70
4.5	Simulated (a) focusing efficiency and (b) focusing uniformity achieved by prVBEM, GGS 2-1, RAF, and RAF 2-1 respectively under different SNR levels when using $N = 1024$, $\gamma = 5$, and a running time of 50 s. Note the error bars denote the standard deviations of 30 repeated tests.	71
4.6	Experimental configuration for retrieving the TM of an MMF with speckle intensity measurements. C, collimator; DMD, digital mirror device; L, lens; P, polarizer; MMF, multimode fiber; Obj, objective lens. $\lambda/4$, quarter-wave plate.	72
4.7	Comparison of light focusing results through MMF with the TMs retrieved by different algorithms. (a) The histograms of normalized PBR of 20×20 foci and (b) the results of multi-spot focusing (pentagram) in the output field of MMF, both obtained by prVBEM, GGS 2-1, RAF, and RAF 2-1 respectively with $N = 1024$ and $\gamma = 5$. Note the crosses in (a) represent the mean values, the white dashed circles in (b) show the fiber region. The scale bar in (b-e) was $20 \mu\text{m}$	73
4.8	Comparison of the PBR maps of focusing on the working plane of the MMF using the TMs (a) measured by off-axis holography and (b) retrieved by RAF 2-1 under a range of with $N = 1024$. The scale bar in (a-b) was $20 \mu\text{m}$	74
4.9	Comparison of image transmission results through MMF using the retrieved TMs by RAF 2-1 under a range of γ with $N = 1024$. (a) Normalized reconstructed images for an object of Chinese character "光", with the values of PCC to the object labeled (b) The progression curves of PCC during the iterative reconstruction.	76

4.10	Simulations of the retrieval of TM corresponding to 20×20 foci under a range of input size (N) with a sampling ratio of 4 and an iteration number (T) of 500, implemented by using MvM and DFT, respectively. (a) Focusing PBR versus N . (b) Time cost versus N . Note the error bars denote the standard deviations of 5 repeated tests.	77
5.1	Experimental configuration of MMF fluorescence imaging with two schemes. Scheme 1, point-scanning imaging with transmission-mode calibration method and a single-pixel fluorescence detection module; Scheme 2, object reconstruction with reflection-mode calibration method by exploiting a wide-field fluorescence detection module. CMOS, complementary metal-oxide-semiconductor camera; DMD, digital mirror device; DM, dichroic mirror; F, filter; FPGA, field-programmable gated array; L, lens; M, mirror; MMF, multimode fiber; BS, beam splitter; Obj, objective; P, polarizer; PBS, polarizing beam splitter; PMT, photomultiplier tube; S, shutter; sCMOS, scientific CMOS; TL, tube lens; $\lambda/2$, half-wave plate; $\lambda/4$, quarter-wave plate.	83
5.2	Schematic illustration on using DMD for phase (complex-amplitude) modulation and fast CGH generation. (a) DMD and a 4f spatial filtering system can produce a desired phase (complex-amplitude) field encoded by a binary CGH. (b) Data flow of generating CGHs rapidly through MEX-CUDA programming. (c) Speed curves of generating CGHs for several different MEX functions, where the error bars denote 5 repeated tests.	84
5.3	Timing diagram of optical modulation and imaging devices in MMF fluorescence imaging during calibration and imaging phases.	85
5.4	Schematic of reflection-mode calibration of MMF and the proposed fluorescence-based RVITM. (a) Illustration of the epi-detection of fluorescence speckle I_{fluor} backscattered through an MMF. (b) Illustration of fluorescence speckle demixing via non-negative matrix factorization. (c) Fluorescence-based RVITM can be calculated with the recovered target intensities under a series of measurements. (d) Selective focusing on different targets can be achieved with the RVITM.	86
5.5	Performance characterization for MMF focusing, at a working distance of 150 μm from the distal fiber tip. (a) Map of peak-to-background (PBR) under phase-only modulation. Map of power ratio (PR) under (b) phase-only and (c) complex-amplitude modulation, respectively, both calculated via HDR imaging. (d) A focus spot in which a white dashed line denotes the intensity profile plotted in (e), which has a FWHM of $\sim 1.85 \mu\text{m}$ used for quantifying the lateral resolution. (f) The axial resolution of $\sim 35.93 \mu\text{m}$, measured by the FWHM of the foci intensity profile along axial direction.	88
5.6	MMF point-scanning imaging of 2 μm fluorescence microspheres. (a) Bright-field image of the particles under MMF speckle illumination, used for reference. (b) Fluorescence image of the particles within the same FOV as (a). (c) Zoomed-in view of the region denoted by a white dashed square in (b). (d) Curves of intensity profiles for the single particle (horizontal) and the two adjacent particles (vertical) shown in (c).	90

5.7	Improved practice for MMF calibration and the imaging results on a fluorescent labeled mouse brain slice before and after that. (a) Previous calibration for MMF in air. (b) Current calibration for both fiber and slide (in blank area). (c) A sample of brain slice imaged with a wide-field fluorescence microscope. (d) Previous imaging result. (e-f) Improved imaging results with the new calibration practice.	91
5.8	Algorithm evaluations for NMF with a rank value of 20. (a) Illustration of fluorescence speckle demixing based on NMF. (b) Root-mean-square residual versus running time for representative NMF algorithms. (c) Average focusing PBR achieved by the RVITMs constructed with representative NMF algorithms versus running time. (d) Boxplot of focusing PBR achieved by the RVITMs constructed with different NMF algorithms at a running time of 60 s. Note the results in (b-d) were all repeated for 5 tests.	92
5.9	Simulations of rank estimation for NMF. (a) Ground truth of 12 fluorescence targets. (b) Root-mean-square residual of NMF versus different values of rank estimation. (c) Focusing PBR achieved by the RVITMs constructed with NMF versus different ranks. Note results in (b-c) were repeated for 5 tests.	93
5.10	Simulations of fluorescence object reconstruction based on fingerprints. (a) Fingerprints of each individual target. (b) Deconvolution over FPs gives relative positions among the targets, which can be summed to form a sub-image centered around a target. (c) The cross-correlation map of the sub-images reveals the relative position of each target with respect to target 1, used for merging all the sub-images for (f) final reconstruction. (d) Ground truth. (e) The sum of foci achieved by the fluorescence-based RVITM.	93
5.11	Preliminary experimental results on fluorescence-based RVITM. (a) Module of fluorescence filtering and epi-detection with an sCMOS camera. (b) 2- μm fluorescence particles placed at the distal fiber end. (c) Fluorescence speckle captured at an exposure time of 1 s, in which the area denoted by a black solid box is cropped for input data. (d) Profile of high-pass filter for fluorescence data filtering. (e) Filtered fluorescence speckle. (f) FPs of each target obtained by NMF. (g) Recovered target intensities under a series of measurements. (h) Focusing on the targets (failed) with the constructed RVITM.	95

List of Tables

1.1	Comparison of various embodiments of PA imaging, adapted from Ref.[7]. . .	11
2.1	Quantitative comparison between deep learning and blind deconvolution in evaluating image enhancement, in which the metrics are represented in the form of mean \pm standard deviation.	30
5.1	Major progress of MMF fluorescence endomicroscopy.	81

Chapter 1

Introduction

1.1 Light propagation in complex media

Light is an excellent carrier for delivering energy and information since it propagates in a nearly lossless manner in transparent media such as air and glass fiber, which has promoted widely used technologies such as photolithography, fiber-optic communications and so on nowadays. However, light will be strongly scattered when it comes to complex media like smoke, fog, turbid liquids, biological tissues, or multimode fiber (MMF), which disrupt the distribution of energy or information it has carried. Light scattering in complex environments has been almost universal in nature and largely unavoidable, affecting applications from remote sensing through atmosphere to deep-tissue optical imaging.

Biological tissues are opaque to light due to its heterogeneous spatial distribution of refractive index, causing light to diffuse rapidly and preventing us from seeing deep within tissues. Optical scattering is prominent in light extinction when travelling in tissues, compared to other factors like absorption, as the scattering coefficient μ_s is much larger than the absorption coefficient μ_a in most tissues. Since light scattering may come with a change in the direction of photons, multiple scattering can gradually lead to a randomization of photon propagation direction. The mean free path (MFP) describes the average distance a photon can travel between two scattering events, given as $MFP = 1/\mu_s$. MFP for visible light is typically $\sim 100 \mu\text{m}$ in tissues (Figure 1.1a) and it varies with tissue types. The scattering coefficient is reduced when considering an anisotropy factor $g \in (0.8, 1)$ describing the probability of forward scattering such that, $\mu'_s = \mu_s(1 - g)$. Note that $\mu'_s = \mu_a(1 - g)$ still holds for most tissue. The transport

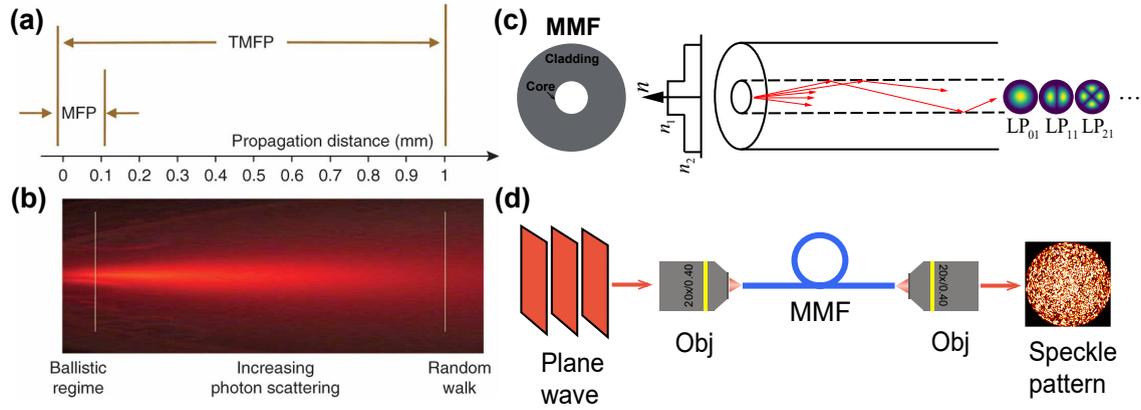


Figure 1.1: Illustration of light propagation in different complex media. (a) Typical distances for MFP and TMFP in biological tissues [1]. (b) Illustration of light propagation and diffusion in tissues [1]. (c) The refractive index distribution of a step-index MMF, the light propagation properties and the supported optical modes within it. (d) Schematic of light scrambling when passing through an MMF. Obj: objective.

mean free path (TMFP) is defined as $TMFP = 1/\mu'_s$. The larger the g , the longer the TMFP will be due to increased forward scattering. TMFP is typically ~ 1 mm for most biological tissues, beyond which light becomes totally diffusive with the random walk of photons, as shown in Figure 1.1b.

Another important complex medium in this thesis is MMF. Similar to biological tissue, it also scrambles incident light into a seemingly random speckle pattern (Figure 1.1d), but the principle behind that is different. As seen in Figure 1.1c, an MMF is a cylindrical waveguide, composed of a core and cladding on the periphery with different refractive indices (n_1 and n_2), as well as a coating layer on the outside. The fiber core diameter is usually 50-200 μm , with a limited numerical aperture (NA). For step-index MMF, it is defined as $NA = \sqrt{n_1^2 - n_2^2}$. An MMF supports the propagation of light with different incident angles (within the NA), while only these under certain reflection angles can form stable optical fields within the fiber, namely the modes, for output correspondingly. The number of modes an MMF can support is determined by the optical wavelength λ , the core radius a , and the NA. By defining the V-number as $V = \frac{2\pi}{\lambda}NA$, it can be calculated as $M = \frac{4V^2}{\pi^2}$. An MMF usually supports hundreds to thousands of modes that have different propagation velocities, causing a distortion by mode mixing and coupling, known as modal dispersion. The inherent modal dispersion prevents direct spatially-resolving imaging application of MMF.

1.2 Optical wavefront shaping techniques

Optical wavefront shaping (WFS) was first proposed by I. M. Vellekoop and A. P. Mosk in 2007 [8], which demonstrated focusing light through a highly scattered medium (thickness of ~ 20 times TMFP) and has revolutionized the field of imaging through complex media. In essence, what WFS does is to apply a pre-compensated wavefront by a spatial light modulator (SLM) for controlling the multiply scattered light within or through a scattering medium or the multiple modes in an MMF to produce constructive interference. This process can be described as

$$E_m^{\text{out}} = \sum_{n=1}^N T_{mn} E_n^{\text{in}}, \quad (1.1)$$

where T_{mn} is the element of a transmission matrix (TM) of the complex medium relating the n -th of input mode E_n^{in} to the m -th output mode E_m^{out} . Under phase-only modulation, the optimal input phase shall be $\hat{E}_n^{\text{in}} = \exp(-j\angle T_{mn})$, $n = 1, 2, \dots, N$ for achieving a focus in the m th output mode. The theoretically achieved enhancement η , often referred as peak-to-background-ratio (PBR) [9], that is, the intensity of the focus over that of the background, is determined by

$$\eta = \alpha(N - 1) + 1. \quad (1.2)$$

Here, N is the number of modulation units, and $\alpha \in [0, 1]$ is the enhancement efficient related to the modulation type: for phase-only modulation $\alpha = \pi/4$, for binary amplitude modulation $\alpha = 1/2\pi$, and for complex amplitude modulation, $\alpha = 1$ [9]. There have been two main feedback-based WFS techniques, iterative WFS and TM method, both entailing the output measurements as the feedback for determining the optimal wavefront.

1.2.1 Iterative wavefront optimization

As shown in Figure 1.2a, the essential spirit of iterative WFS is updating the wavefront on a SLM towards optimal modulation for controlling light transmission through a complex medium based on the feedback signals related to the localized intensity of the output. If there is no WFS, under all-zero or random phase pre-compensation, the MMF would output a speckle pattern (Figure 1.2b); with WFS, the MMF would produce a sharp focus or other desired pattern (Figure 1.2c), with the PBR gradually approaching the theoretical value of 452 under 24×24 phase-only modulation (Figure 1.2d).

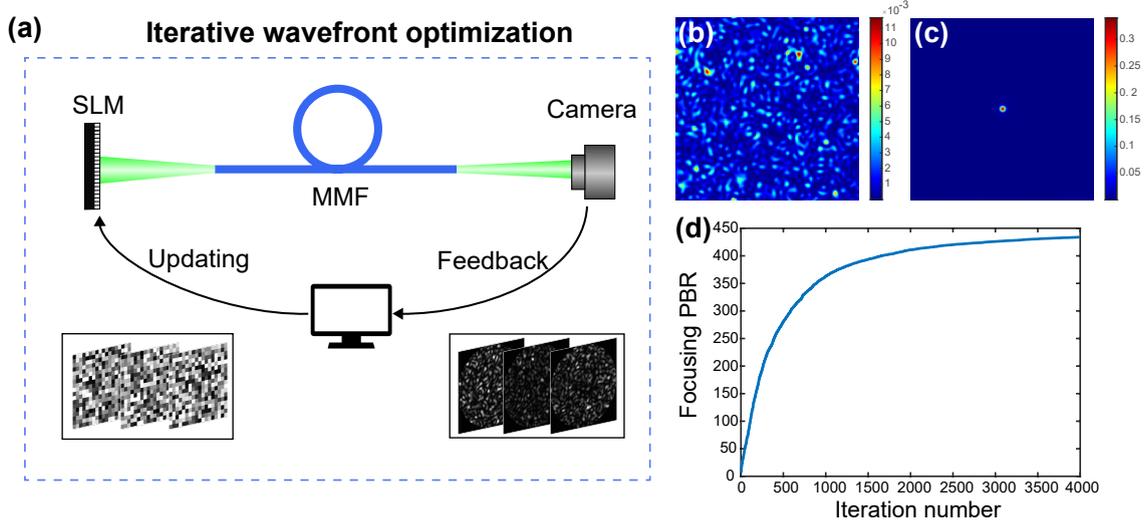


Figure 1.2: Illustration of iterative wavefront optimization for focusing light through MMF with simulated results. (a) Schematic setup. SLM: spatial light modulator. (b) Output light intensity without phase modulation. (c) Output focus with optimal phase input. (d) Curves of focusing PBR versus iteration number.

Iterative WFS requires feedback to indicate the result of focusing, which usually adopts a monitoring camera behind the medium in an invasive manner. Noninvasive schemes would resort to an internal "guide-star" to provide the feedback, such as photoacoustic signal [10], two-photon fluorescence signal [11], etc. The optimization speed of iterative WFS is usually slow (on the scale of minutes), mainly throttled by the refreshing rate of SLM (e.g., 60-200 Hz for liquid-crystal SLM), the acquisition speed of camera, the data processing and transfer speed among multiple hardware, as well as the optimization algorithm. Early phase optimization algorithms included stepwise sequential algorithm, continuous sequential algorithm, and partitioning algorithm [12]. Later, intelligent optimization algorithms such as genetic algorithm (GA) [13] and natural gradient strategy [14, 15] have been adopted for focusing through complex media that show faster searching speed and stronger anti-noise capability.

1.2.2 Transmission matrix

The measurement of the TM of a complex medium offers the knowledge of spatial light transformation from the input to the output, which was pioneered by S. M. Popoff et al in 2010 [16]. The TM allows focusing at any desired position behind the medium and has gained wide applications such as optical imaging, manipulation, and data communication in complex environments. For TM measurement, a series of complete input bases are pre-loaded on a SLM, which can be based on Hadamard matrix [16], discrete Fourier transform (DFT) matrix [17]

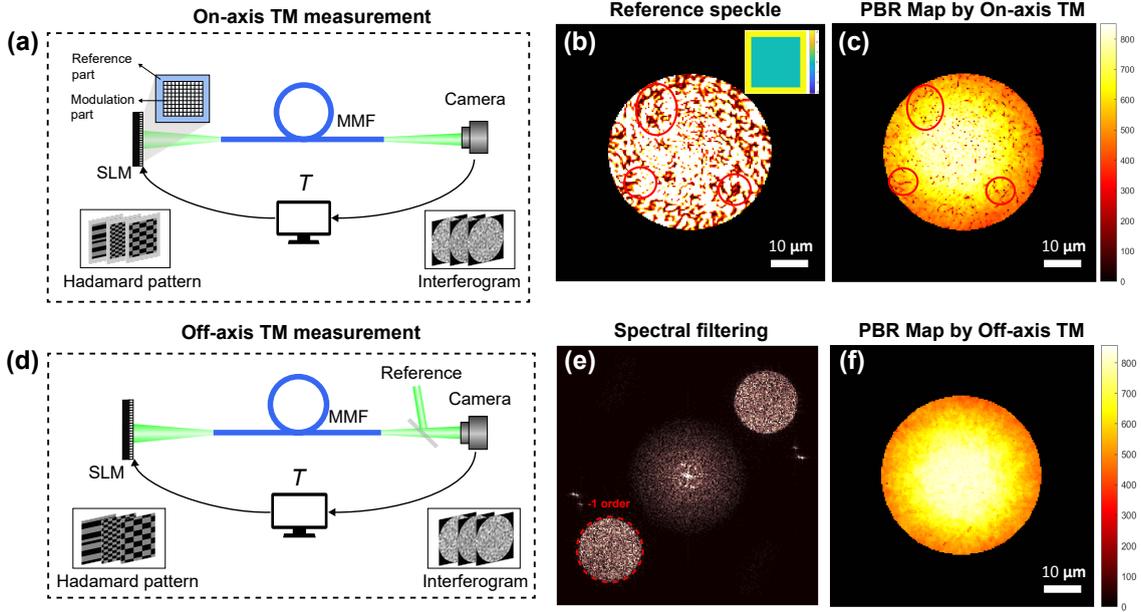


Figure 1.3: Illustration of the TM measurement with on-axis or off-axis holography. (a) Schematic setup of on-axis TM measurement. (b) Speckle intensity produced by a flat reference phase (top-right inset). (c) Map of focusing PBR achieved by on-axis TM. (d) Schematic setup of off-axis TM measurement. (e) Spectral filtering of the -1st order for off-axis TM measurement. (f) Map of focusing PBR with an off-axis TM.

or focal spots [18, 19]. It usually requires full-field interferometric measurements to acquire a TM, such as on-axis holography with an internal reference [16] or off-axis holography with an external reference [18, 20].

In on-axis holography, as shown in Figure 1.3a, the internal reference part and the modulation part co-propagate through a complex medium, producing interferograms for the measurement of TM. If we apply a phase shift β on the reference, the interferometric intensity on the m th output mode I_m^β reads

$$I_m^\beta = |s_m + \sum_n e^{j\beta} T_{mn} E_n^{\text{in}}|^2 = |s_m|^2 + |\sum_n e^{j\beta} T_{mn} E_n^{\text{in}}|^2 + 2 \text{Re}\{\bar{s}_m e^{j\beta} \sum_n T_{mn} E_n^{\text{in}}\}, \quad (1.3)$$

where s_m is the m th mode of the reference speckle, $m = 1, \dots, M$, M is the number of output modes, and $\text{Re}\{\cdot\}$ takes the real-valued part. With the four-phase-shifting method, in which holographic measurements are taken when $\beta = 0, \pi/2, \pi, 3\pi/2$, we have the complex-amplitude output field E_m^{out} (mixed with the reference speckle conjugate \bar{s}_m), given by

$$\frac{(I_m^0 - I_m^\pi)}{4} + \frac{j(I_m^{3\pi/2} - I_m^{\pi/2})}{4} = \bar{s}_m \sum_n T_{mn} E_n^{\text{in}} = \bar{s}_m E_m^{\text{out}}, m = 1, \dots, M. \quad (1.4)$$

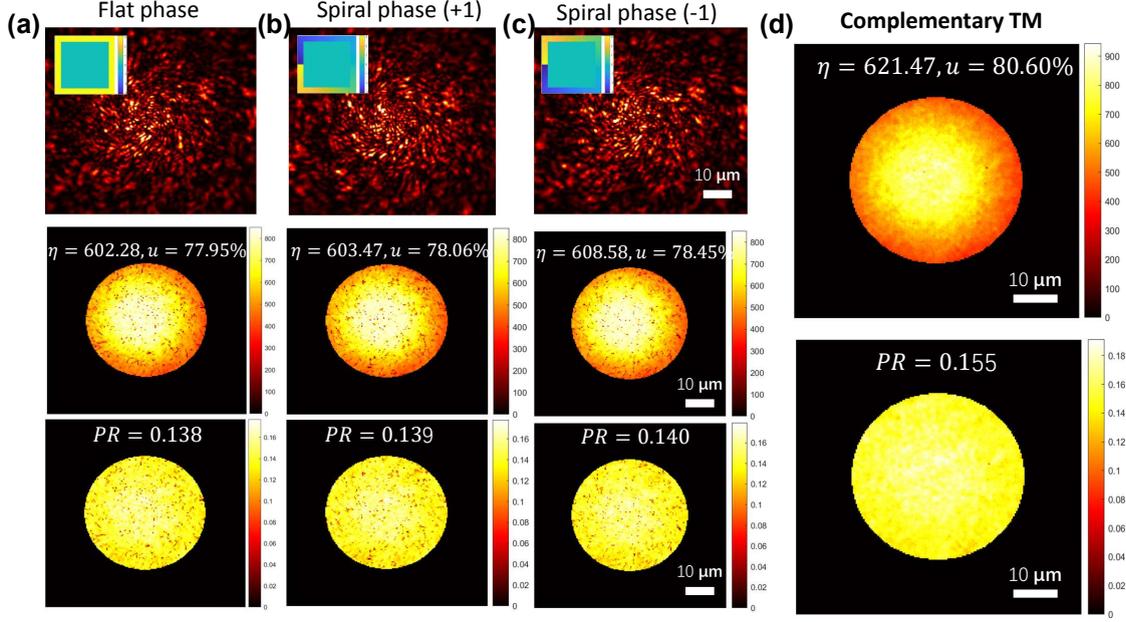


Figure 1.4: Illustration of solving the "dark spot" problem faced by on-axis TM using the complementary TM measured with multiple internal references. The reference speckle, PBR map and PR map by on-axis TM measured with (a) flat reference phase, (b) spiral phase (+1) and (c) spiral phase (-1). (d) The PBR map and PR map by the complementary TM, which is constructed by the above three TMs. Note η denotes focusing PBR and u denotes focusing uniformity in each PBR map.

Let the diagonal matrix composed of the reference speckle fields be $\mathbf{S}_{\text{ref}} = \text{diag}(\bar{s}_1, \bar{s}_2, \dots, \bar{s}_M)$, the observed output field reads $\mathbf{E}_{\text{out}}^{\text{obs}} = \mathbf{S}_{\text{ref}} \times \mathbf{E}_{\text{out}}$ where $\mathbf{E}_{\text{out}} = [E_1^{\text{out}}, E_2^{\text{out}}, \dots, E_M^{\text{out}}]^T$ is the ideal output field vector. As such, the observed TM is given by $\mathbf{T}_{\text{obs}} = \mathbf{E}_{\text{out}}^{\text{obs}} \times E_{\text{in}}^{-1} = \mathbf{S}_{\text{ref}} \times \mathbf{T}$ where \mathbf{T} is the ideal TM. The measurement of on-axis TM is influenced by the reference speckle, not only in the form of coupling with it. In fact, the TM cannot be accurately measured in the positions where the intensities of the reference speckle are too weak to form effective interference with the signal light [21, 22]. As an experimental proof, Figure 1.3 (b-c) show a good correspondence between the dark areas of a reference speckle and those of the map of PBR (denoted by the red circles) when focusing through an MMF with an on-axis TM.

In off-axis holography, as shown in Figure 1.3d, an external reference beam is employed for interference with the speckle beam output by the complex medium, in which phase-shifting or spatial filtering method can be used to obtain the complex amplitude of the output field. Off-axis holography with spatial filtering only requires N measurements, in contrast to the four-phase-shifting method. Figure 1.3e shows the frequency spectrum of an off-axis interferogram where the -1st order (down-left) is selected for spectral filtering. Compared to on-axis TM, off-axis TM does not suffer from the "dark spot" problem and produce more homogeneous

foci across the output plane (Figure 1.3f). However, off-axis TM measurement necessitates an external reference beam and very stable environment for interferometry, which could be challenging in some scenarios.

Apart from the above two TM measurement methods, recent efforts also resort to retrieving the complex-valued TM directly from the output intensity measurements using advanced phase retrieval algorithms [23–25]. The TM retrieval method waives the need of an external reference beam, and the result can approach the golden standard off-axis holography. However, it usually entails more measurements and a large amount of computation.

The "dark spot" problem of on-axis TM can be mitigated by measuring a complementary TM with multiple internal references [21, 22]. The idea behind it is the reference speckle fields produced by distinct references can be complementary and form a homogeneous reference field virtually. We have confirmed that experimentally by using three internal references, flat phase, spiral phases with topological charge of +1 and -1, with results given in Figure 1.4. Mathematically, suppose we have three on-axis TMs $r'\mathbf{T}$, $r' = 1, 2, 3$, then the phase difference ${}_{m'}^r\Delta\varphi$ between the r' th TM and the 1st TM (as the reference TM) at the m th output mode is

$${}_{m'}^r\Delta\varphi = \arg \sum_n^N {}^1\mathbf{T}_{mn}^* {}^{r'}\mathbf{T}_{mn}, \quad (1.5)$$

where $(\cdot)^*$ denotes the phase conjugate. The three TMs are combined after correcting the phase differences to form a complementary TM,

$$\hat{\mathbf{T}}_{mn} = \sum_{r'=1}^3 {}^{r'}\mathbf{T}_{mn} \exp(-j {}_{m'}^{r'}\Delta\varphi). \quad (1.6)$$

We experimentally confirm the complementary TM guarantees more homogeneous focusing results than these achieved by on-axis TM with a single reference. As shown in Figure 1.4d, both the focusing PBR and power ratio (PR) are improved, almost without obvious dark spots, which is important for the quality of MMF-based point-scanning imaging. Also note that the maps of PR were not calculated by high-dynamic-range (HDR) imaging, which were lower than the actual values and used for relative comparison only.

1.2.3 Relationship with adaptive optics

Adaptive optics (AO) refers to a more general category of techniques for aberration correction in the imaging path of an optical telescope or microscope [2, 26], where the distortions mainly come from turbulent atmosphere or biological specimen, respectively. As illustrated in Figure 1.5, in an AO microscopy, a fast deformable mirror (DM) is usually employed that locates at the input pupil of an objective lens, for rapidly compensating the incident wavefront so that a corrected focus can be formed through the specimen. In the meanwhile, the DM also eliminates the aberrations in the detection path, allowing for high-resolution imaging at tissue depths.

Although sharing similar principles, there are some key differences between WFS and AO in the context of biophotonics. AO mainly works in the realm of ray optics, aiming to correct weak aberrations caused by the optical system and the refractive index mismatch of specimen, for improved imaging quality over a large field of view (FOV). But for WFS, it overcomes the strong multiple scattering induced by a much thicker sample via producing constructive interference at desired spots in the FOV. That said, AO and WFS also share a cross-over regime for deep-tissue optical imaging at the depths ranging from a scattering MFP to a 1 TMFP [9].

1.3 Deep-tissue optical imaging approaches

Optical imaging holds many advantages in biomedical application, such as high spatial resolution, non-radiative safety, high sensitivity, and high molecular specificity etc. However, as light scattering increases exponentially with the penetration depth, it seriously degrades the focusing of light and fundamentally limits high-resolution imaging capability at depths in tissue. Conventional optical imaging techniques, such as wide-field microscopy confocal microscopy, primarily utilize ballistic photons that are un-scattered to ensure imaging resolution, thus limited to shallow imaging depth (usually tens to hundreds of micrometers). There is an inherent tradeoff between imaging depth and spatial resolution achieved by optical microscopy, as illustrated in Figure 1.6a.

Deep-tissue optical imaging approaches have been developed for preclinical or clinical applications, such as near-infrared-II (NIR-II) imaging, diffuse optical imaging, photoacoustic tomography etc. NIR-II imaging operates in the second near-infrared window (1000-1700 nm) and detects fluorescence signals from deep tissues labelled by photoluminescent contrast agent

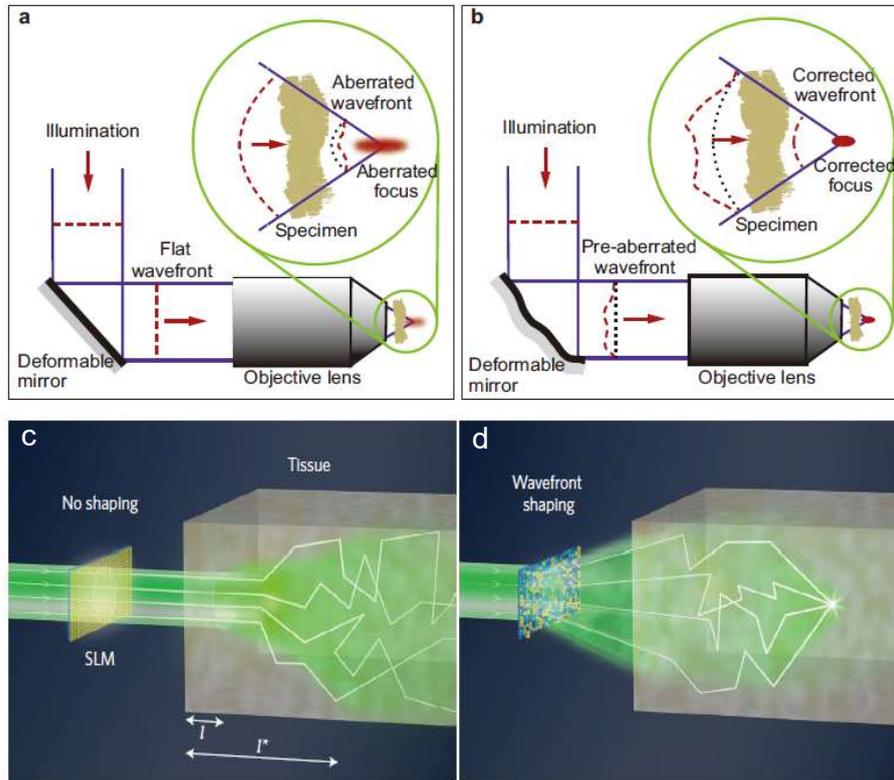


Figure 1.5: Schematic diagram for the principle of (a-b) AO that corrects aberration for thin-tissue microscopy [2], and (c-d) WFS to form a focus within a thick sample through constructive interference among wavelets travelling along different optical paths [3].

like carbon nanotubes. NIR-II imaging greatly reduces light scattering and autofluorescence, which is ideal for non-invasive inspection of deep tissues with high resolution and contrast, with application to cancer theranostics and brain imaging etc. Diffuse optical imaging detects the diffused photons that are scattered by tissues for modelling the light propagation through tissues and monitoring the molecular structures and component contents (e.g., hemoglobin, fat, water) within the tissues. Diffuse optical imaging is advantageous in label-free tissue function imaging and clinical diagnosis by selecting different wavelengths (e.g., NIR) for illumination and imaging modalities such as tomography or 2D imaging, with an imaging depth of 5~10 cm. Photoacoustic tomography is hybrid in optical absorption and ultrasound detection. By selecting different acousto-optic alignment, imaging methods and ultrasound transducer types, it achieves multi-scalable spatial resolution and imaging depth, gaining much attention for structural and functional imaging applications. Herein, several relevant deep-tissue optical imaging methods to the study of the thesis will be deeply discussed, including photoacoustic imaging, fluorescence imaging and fiber-optic endoscopic imaging, which are all valuable to biomedicine.

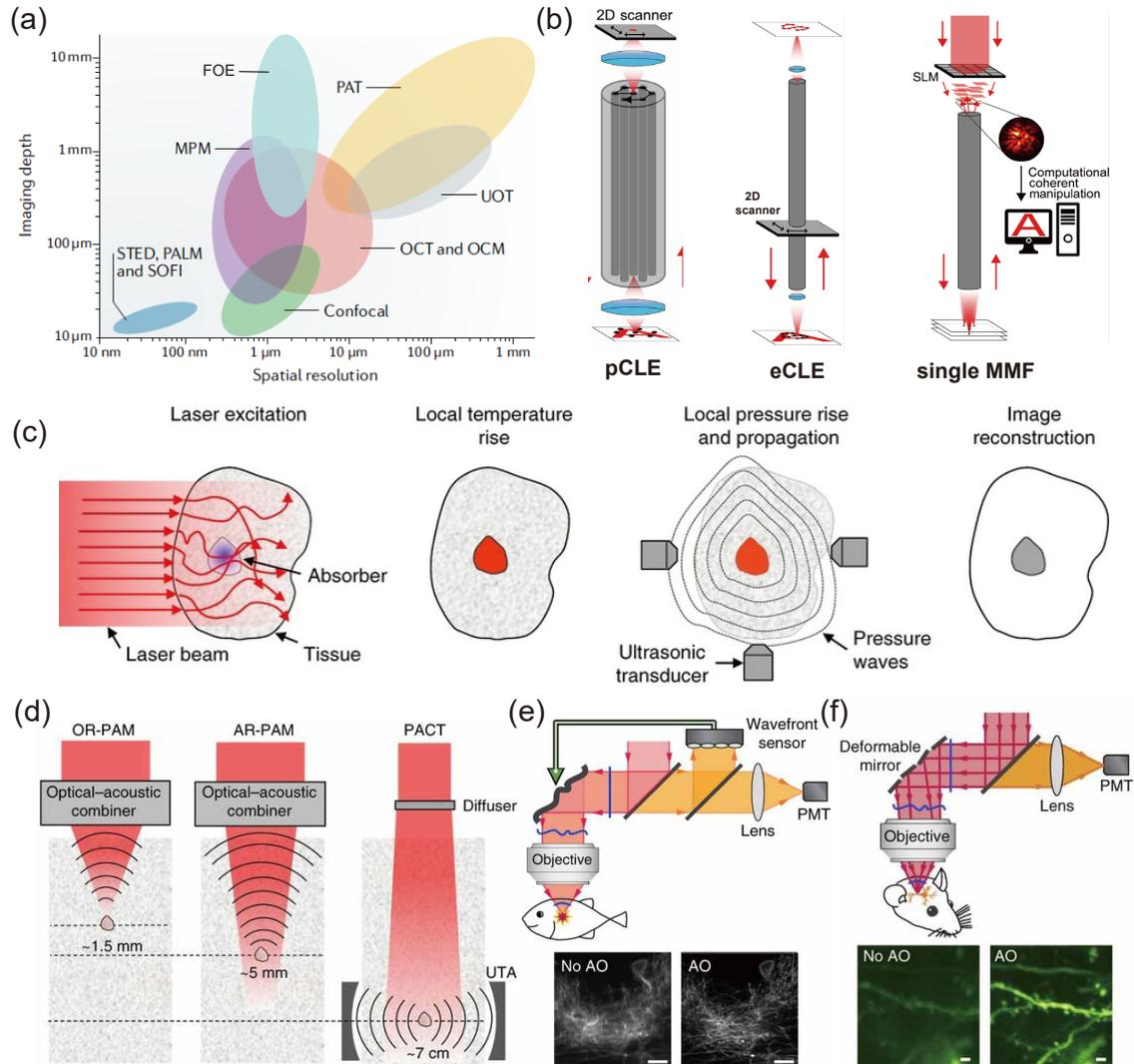


Figure 1.6: Schematic illustrations of representative deep-tissue optical imaging approaches. (a) Imaging depth versus spatial resolution for various optical imaging modalities, reproduced from [4] with fiber-optic endoscopy (FOE) being supplemented. (b) Representative embodiments of confocal laser endomicroscopy (CLE) [5]. (c) Principle of photoacoustic imaging. (d) Typical photoacoustic imaging embodiments. (e-f) Two-photon fluorescence microscopy with (e) direct adaptive optics (AO) using a wavefront sensor and (f) indirect AO based on pupil segmentation [6].

1.3.1 Photoacoustic imaging

Figure 1.6c shows the principle of photoacoustic (PA) imaging which is based on PA effect. When shining a pulsed laser into a tissue sample, light is absorbed and converted into heat, producing ultrasonic waves due to thermoelastic expansion, which are detected by an ultrasonic transducer for the reconstruction of an image that depicts the tissue's optical absorption coefficient distribution. Thanks to optical absorption and ultrasound detection, PA imaging combines the advantages of optical rich contrast and ultrasonic penetration depth. PA imaging

Table 1.1: Comparison of various embodiments of PA imaging, adapted from Ref.[7].

	OR-PAM	AR-PAM	PACT
Lateral resolution	Optical (0.2~10 μm)	Acoustic (15~50 μm)	Acoustic (~0.5 mm)
Axial resolution	Acoustic (~30 μm)	Acoustic (15~30 μm)	Acoustic (~0.5 mm)
Imaging depth	Optical quasi-ballistic regime (~1 mm)	Optical diffusive (1-10 mm)	Optical diffusive regime (≥ 10 mm)
Imaging mode	Raster scanning	Raster scanning	Parallel (1D or 2D array-based)

is intrinsically capable of structural and functional imaging of blood vessels since it is sensitive to hemoglobin, which is abundant in blood vessels and shows dominant optical absorption at typical visible and near-infrared spectrum. PA imaging is also naturally capable of volumetric imaging and tomography thanks to the detection of A-line PA signals from targets of different depths, thus is sometimes referred as PA tomography (PAT).

PA imaging is scalable in spatial resolution, imaging depth, and imaging contrast, by selecting different laser sources, ultrasonic transducer types, acousto-optic alignment, and scanning methods. Generally, PA imaging can be categorized as optical-resolution photoacoustic microscopy (OR-PAM), acoustic-resolution photoacoustic microscopy (AR-PAM), and photoacoustic computed tomography (PACT), with decreased spatial resolution yet increased imaging depth in turn, as shown in Figure 1.6d. Both OR- and AR-PAM adopt raster-scanning across the sample to produce signals, in which OR-PAM works in a quasi-ballistic range while AR-PAM usually works in an optical diffusive regime. Specifically, the size of focus spot in OR-PAM dictates the lateral resolution $R_{(\text{lat,OR})}$, given by

$$R_{(\text{lat,OR})} = 0.5\lambda_o/NA_o, \quad (1.7)$$

where λ_o is the optical wavelength and NA_o is the effective NA of optical focusing [27]. In comparison, the lateral resolution of AR-PAM is determined by the acoustic focus of the ultrasonic transducer instead, given by

$$R_{(\text{lat,AR})} = 0.71\lambda_A/NA_A, \quad (1.8)$$

where λ_A is the acoustic wavelength and NA_A represents the NA of the ultrasonic transducer

[27]. For PACT, it utilizes a transducer array (1D or 2D) to simultaneously detect PA signals emitted from tissue at multiple positions, achieving faster imaging speeds and larger penetration depth but lower spatial resolutions. The main specificities and imaging mode of the above PA imaging embodiments are summarized in Table 1.1.

1.3.2 Fluorescence imaging

Optical imaging with the contrast mechanism of fluorescence has been commonly utilized in biomedicine as it achieves single molecular specificity and high sensitivity, thus allowing monitoring molecular dynamics and signals in complex biological environments. However, due to strong optical scattering, fluorescence imaging is usually limited to thin samples or superficial depths. In the past decades, the major workhorse for deep-tissue fluorescence imaging is the development of two- or multi-photon microscopy (MPM), which are based on the nonlinear effect of two- or multi-photon excitation [28]. MPM has several advantages that contribute to enhanced deep-tissue imaging capability. First, MPM usually employs near-infrared laser (700-1300 nm) for excitation, which has a longer penetration depth as it encounters less optical scattering in tissue. Second, thanks to the nonlinear excitation where the emission light is proportional to the square or cube of the intensity of excitation light, the fluorescence emission is confined in the focus, increasing the signal-to-background ratio (SNR), contrast, and axial resolution for imaging in deep tissue.

That said, the imaging depth of MPM could still be insufficient for maintaining high-resolution imaging beyond the tissue depth of, say 500 μm in a mouse brain cortex [29]. In the past two decades, AO has been introduced for active wavefront control to further correct the tissue-induced aberrations for improving the imaging quality of deep-tissue fluorescence microscopy. Basically, there have been direct AO and indirect AO. For direct AO, as shown in Figure 1.6e, usually a Shack-Hartmann wavefront sensor is adopted for direct wavefront sensing by measuring the local slope of the wavefront. The detected information can be used for the DM to rapidly apply the wavefront modulation to the excitation light. Direct AO is only applicable to imaging at superficial depth as the wavefront sensor requires enough ballistic photons for direct wavefront measurement. By contrast, indirect AO is more practicable as it does not involve a wavefront sensor. It includes pupil-segmentation based method (Figure 1.6f) and

metric-based method. Indirect AO requires wavefront reconstruction which limits its correction speed, but on the other hand, it can be applied to opaque tissues with a simpler hardware implementation [6].

1.3.3 Fiber-optic endoscopic imaging

Conventional deep-tissue optical imaging methods inherently encounter the tradeoff between the imaging depth and the imaging quality. One feasible way to overcome it is to introduce light sources inside tissue and make observation *in situ*, thus bypassing the challenge of optical scattering. As such, fiber-optic endoscope (FOE) has been developed that allows imaging at arbitrary depth with optical resolution (Figure 1.6a). FOE, in other words, confocal laser endomicroscopy (CLE) [30, 31], mainly includes probe-based CLE (pCLE) and endoscope-based CLE (eCLE), as shown in Figure 1.6b. Generally speaking, pCLE scans and couples light into individual optical fiber of an optical fiber bundle for laser-scanning imaging, with a micro-objective lens attached at the end for improved imaging resolution. pCLE allows flexible endoscopy into body tissues such as gastrointestinal tract. eCLE usually integrates a scanning mirror at the distal tip of a conventional endoscope, which is typically a single-mode fiber (SMF), to perform raster scanning for diffraction-limit imaging. eCLE is excellent in imaging quality and robustness, which has been actively developed to achieve PA imaging [32] or optical coherence tomography [33], used for optical biopsy *in situ*. However, due to the usage of the distal lens or scanning device, the diameters of both CLE probes are still relatively large, ranging from millimeters to centimeters, which may cause serve trauma when imaging onto tissues without natural tract.

In the past decade, thanks to the invention of WFS technique, the field has seen exciting development of single MMF-based endomicroscopy [34–36], which is ultra-minimally invasive since the size of fiber probe is typically 50–200 μm . Assisted by WFS, it is feasible to control light propagation through MMF for selective excitation of the optical modes with similar group velocities to produce diffraction-limited focus, allowing for MMF laser-scanning imaging. Combined with different signal detection modules, many kinds of optical microscopy have been realized through a single MMF, including fluorescence [37], photoacoustic [38], reflectance [20], and nonlinear [39] imaging, etc. Ultrafine MMF-based endomicroscopy gains special interests for imaging at sensitive tissue regions such as deep brain, and it has been recently reported to observe neuronal activity, connectivity, and blood flow dynamics *in vivo*

throughout the whole mouse brain [40]. But still, MMF-based endomicroscopy faces challenges like limited imaging performance in FOV and resolution and especially the sensitivity to external disturbance, which limits it to the realm of rigid endoscope. More comprehensive reviews on the methods, applications, and challenges of MMF-based endomicroscopy can be referred to refs.[35, 36].

1.4 Thesis structure

The research in this thesis centers around deep-tissue optical imaging and focusing, utilizing PA imaging and MMF-based endoscopy as two primary tools, with a particular emphasis on developing computational methods. Specifically, these studies are divided into three major parts: Part I explores deep-penetrating PAM with high resolution by a data-driven method; Part II achieves deep-tissue patterned optical delivery through MMF based on wavefront optimization or TM reconstruction; Part III further investigates MMF-based fluorescence endomicroscopy using transmission- or reflection-mode calibration schemes. Towards this end, several projects have been conducted during my PhD study to address the major research gaps, as summarized in the following outline.

In Chapter 2 (Part I), to extend OR-PAM towards deep-tissue imaging, we propose a deep learning-based imaging method that instantly transforms the blurry images acquired with an AR-PAM to match the OR-PAM results, while maintaining the imaging depth of AR-PAM. Using an integrated OR- and AR-PAM system, we have trained an improved generative adversarial network (GAN) to learn the imaging transformation relation between AR- and OR-PAM images. The trained GAN can improve the mouse cerebrovascular image that it hasn't seen before and retain the high-resolution results at tissue depths beyond one TMFP.

In Chapter 3 (Part II), in the scope of controlling light delivery through MMF for deep-tissue applications, both focusing and suppressing light in a patterned region at the output of MMF have been studied, i.e., patterned light delivery. To achieve long-distance pattern projection with high fidelity through an MMF in an unstable environment, we have adopted a natural gradient ascent-based strategy guided with a new fitness function. It offers faster convergence and stronger robustness over existing iterative WFS schemes, showing promise for energy delivery-related biomedical applications such as phototherapy. Besides that, we also propose

an alternating projection-based phase optimization method for fast and arbitrary glare suppression through an MMF, attractive for deep-tissue super-resolution imaging and optical trapping that involves a customizable speckle field.

In Chapter 4 (Part II), to reconstruct the TM of an MMF from merely output intensity measurements, which waives the need of an external reference beam and avoids the "dark spot" problem induced by internal reference, we develop a two-step reweighted amplitude flow (RAF 2-1) algorithm for TM retrieval with optimum efficiency. Compared with existing representative TM retrieval methods, it requires less running time or sampling ratio to reach the theoretical focusing efficiency with the retrieved TM. Experimentally, RAF 2-1 can approach the golden standard off-axis holography method under a sampling ratio of 8, which facilitates deep-tissue optical imaging-related applications with the use of an MMF.

In Chapter 5 (Part III), to promote MMF-based fluorescence endomicroscopy, especially by an online calibration of MMF, we conduct a pilot study with both transmission-mode (Scheme 1) and reflection-mode (Scheme 2) TM calibration. In Scheme 1, we have validated the performance of point scanning fluorescence imaging with samples of microspheres and mouse brain slice. In Scheme 2, we have proposed a fluorescence-based intensity TM based on the unmixing of the epi-detected fluorescence speckle. In principle, it offers a feasible solution for the re-calibration of MMF in the absence of distal access during fluorescence imaging, while further experimental validation is needed.

In Chapter 6, we briefly summarize all the research work in Chapters 2-5 and point out the major contributions of the thesis. We also discuss several aspects of our imaging technologies and the future directions from an overall perspective, with the aim towards more practicable deep-tissue optical imaging and focusing with the use of PA imaging, MMF imaging and computational optics.

The research encapsulated in this thesis is depicted in Figure 1.7. Faced with the contradiction between imaging resolution and penetration depth mainly due to the strong light scattering, our studies aim to achieve high-resolution PA imaging and MMF-based endoscopy within deep tissues. Specifically, computational methods have been developed for either postprocessing scattering-degraded AR-PAM images (Part I) or controlling light delivery through MMF for endoscopic applications (Part II and III). All the studies serve the common goal of advancing deep-tissue, high-resolution optical imaging and focusing. The studies in Part II on light control of MMF have laid a good foundation for MMF-based fluorescence imaging in Part III. Although

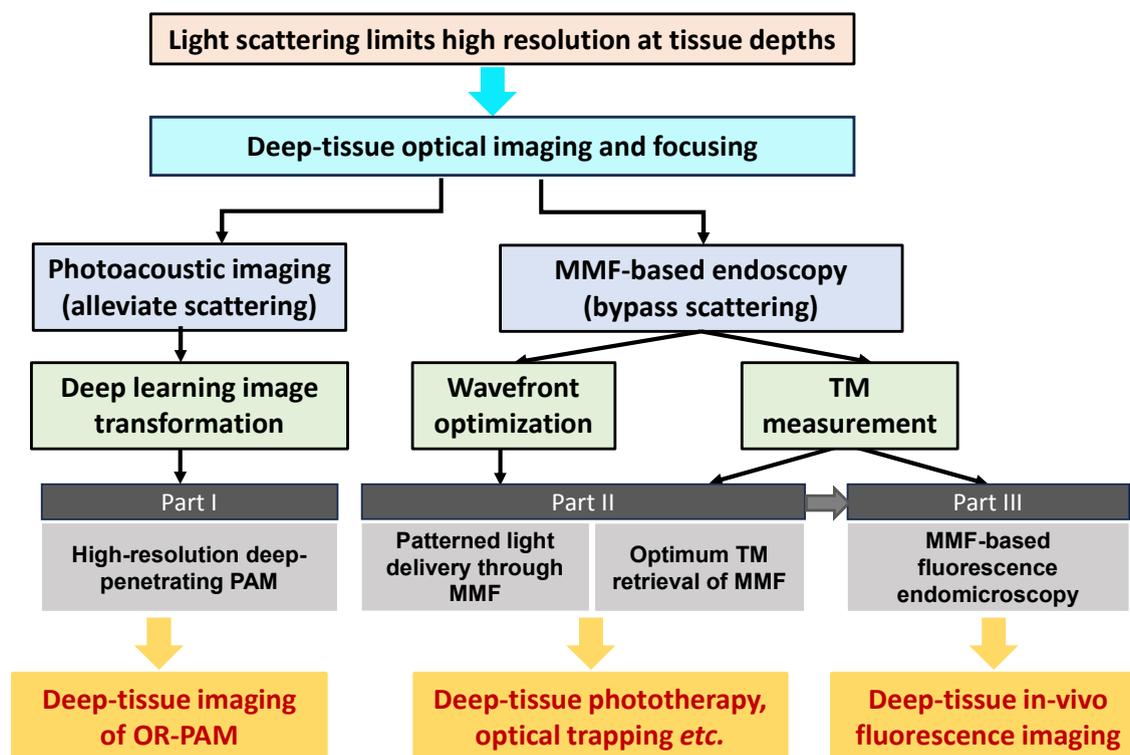


Figure 1.7: Research framework of the thesis showing the research topic, challenge, imaging technologies, computational approaches, research contents and potential research values.

Part I, which is about PA imaging, seems less relevant to the other parts, the wavefront modulation methods introduced in Part II can also be applied to enhance focusing quality and AR-PAM imaging physically in deep tissues [41]. Besides, the idea of MMF-based fluorescence imaging in Part III can also be translated to realize endoscopic PAM with high resolution [38, 42]. Regarding biomedical applications, our research may open new avenues for deep-tissue optical imaging, phototherapy, optical manipulation and more.

Chapter 2

High-Resolution Deep-Penetrating Photoacoustic Microscopy through Deep Learning

This chapter is reproduced with some adaptations according to a published journal paper: Shengfu Cheng, Yingying Zhou, Jiangbo Chen, Huanhao Li, Lidai Wang, and Puxiang Lai, "High-resolution photoacoustic microscopy with deep penetration through learning", Photoacoustics 25:100314 (2022).

In Part I of this thesis, we investigate the possibility of achieving a deep-tissue OR-PAM. OR-PAM, renowned for its superior spatial resolution and has garnered significant attention recently. However, its application remains limited to shallow depths due to strong multiple light scattering in biological tissues. In Chapter 2, we propose to achieve a performance of deep-penetrating OR-PAM by harnessing deep learning on blurry mouse vascular images acquired *in vivo* using an AR-PAM setup. Our trained GAN improved AR-PAM's lateral resolution from 54.0 μm to 5.1 μm , approaching that of a typical OR-PAM (4.7 μm). We evaluated the network's feasibility with living mouse ear data, which generated superior microvasculature images to blind deconvolution. We also generalized the network by testing on *in-vivo* mouse brain data. Furthermore, our deep-learning method could maintain high resolution at tissue depths beyond one optical transport mean free path in the phantom experiment. Although there is room for improvement, the proposed deep learning image transformation method may open new avenues for expanding the scope of OR-PAM for deep-tissue imaging in biomedicine.

2.1 Introduction

PA microscopy offers high-resolution imaging of rich optical-absorption contrasts *in vivo* and provide structural, functional, and molecular information of biological tissues [7, 43]. OR-PAM uses tightly focused laser beam for excitation and thus has diffraction-limited resolution to resolve single capillaries and monitor microvascular level biological processes. OR-PAM has gained intense attention in the past decade [44–47] and has seen many preclinical and clinical applications in neuroscience [48], tumor angiogenesis [49], histology [50, 51], dermatology [52], and many others [53–55].

Limited by strong scattering in biological tissue, the penetration depth of OR-PAM is within one optical TMFP (~ 1 mm for biological tissues). It would be impactful if OR-PAM can see deeper into tissue. One attempt is to explore whether OR-PAM performance can be inferred or constructed through computation based on deep-penetrating, albeit low-resolution, PA signals. AR-PAM does not focus light tightly and thus can extend to several millimeters to centimeters deep [7, 56]. AR-PAM also waives the necessity for a costly single-mode laser to produce high-quality focused beam. Thus, lower cost multimode pulse laser like laser diode or light-emitting diode can be used as the light source [57–59]. Imagine that if a relationship can be built or learned between superficial AR-PAM and OR-PAM data sets, and the validity of the relationship remains for deeper tissue regions or different organs, then deep-penetrating OR-PAM could be achieved through learning the acoustic-resolution PA signals at that depth.

Here, we propose a deep learning method to transform low-resolution AR-PAM images into high-resolution ones that are comparable to the corresponding OR-PAM results. This allows us to combine the advantages of deep penetration of AR-PAM and high resolution of OR-PAM. Apart from that, OR-PAM imaging with lower cost could also be achieved based on the usage of AR-PAM apparatus, as there will be less restrictions on the laser source. The idea of optimizing low-resolution PA imaging by a high-resolution one is also found elsewhere. As an example, Shi *et al.* transform low-resolution mid-infrared (MIR) PA signals into the difference of two adjacent ultraviolet PA signals for detection and high-resolution MIR PAM imaging [60]. Herein, we utilize a deep learning image transformation method. In the past few years, there have been a number of deep learning applications aimed at enhancing the performance of PA imaging [61, 62], such as increasing the contrast [63] or penetration depth [64] under low fluence illumination, improving the lateral resolution for out-of-focus region in AR-PAM [65], and enhancing OR-PAM images acquired under low laser dosage or sampling rate [66, 67].

Besides, deep learning has seen applications in PACT that mainly involve image enhancement from suboptimal reconstruction [68, 69] and artifact removal [70, 71]. Several related deep learning applications include single image super-resolution [72–75], microscopic image enhancement [76–78] and microscopic imaging transformation [79, 80]. It has been shown that conventional convolutional neural network (CNN) trained with pixel-wise loss tends to output over-smoothed results [75]. In contrast, the GAN model with residual blocks, trained with perceptual loss, performs particularly well (with better details) for these problems.

In this study, we adopt Wasserstein GAN with gradient penalty (WGAN-GP) [81] as the training network to transform low-resolution AR-PAM images to match high-resolution OR-PAM images obtained at the same depth. In the following sections, we first describe the integrated OR- and AR-PAM system for data acquisition and the WGAN-GP model used for PAM imaging transformation. The trained network was first validated with *in vivo* mouse ear vascular images and the performance was compared with that of a typical blind deconvolution method. We further apply the network to *in vivo* mouse brain AR-PAM data to verify its validity for different tissue regions. After that, the performance of the network on deep-tissue imaging is evaluated with a hair phantom. We show that, with the proposed PAM imaging transformation, deep-penetrating OR-PAM imaging could be achieved at depths that are way beyond the depth limit of traditional OR-PAM. Whilst it can be further improved, the proposed method provides new insights to expand the scope of OR-PAM towards deep-tissue imaging and wide applications in biomedicine.

2.2 Methods

2.2.1 Integrated OR- and AR-PAM system

An integrated OR- and AR-PAM system was built in this study to acquire PA data, as shown in Figure 2.1. The laser source is a 532 nm wavelength pulsed laser whose pulse width is 7 ns (VPFL-G-20, Spectra-Physics). The laser output is directly delivered into the PAM probe [47] by a 2-m SMF (P1-460B-FC-2, Thorlabs Inc) for OR-PAM imaging, or by a 1-m multi-mode fiber (MMF, M105L01-50-1, Thorlabs Inc) to support AR-PAM imaging. In the experiment, the pulse energy for the OR- and AR-PAM was ~ 80 nJ and ~ 2000 nJ, respectively. The fiber coupling efficiencies of the SMF and the MMF were measured to be $\sim 60\%$ and $\sim 90\%$, respectively. Noted that the optical/acoustic beam combiner in the probe reflects the optical beam to the sample

and, in the meanwhile, transmits the produced ultrasound wave to the piezoelectric transducer (V214-BC-RM, Olympus-NDT). The central frequency and bandwidth of the ultrasound transducer used in the experiments are 50 MHz and 40 MHz, respectively. The optical-resolution and acoustic-resolution settings are switched by using different fibers only; usually after the entire scanning of OR-PAM for an image is finished, we switch the single mode fiber to a multi-mode fiber manually for AR-PAM imaging. This has endowed our integrated PAM system with the ability to yield automatically co-registered OR and AR imaging data sets [82]. The switch is controlled by the combination of a half-wave plate (HWP, GCL-060633, Daheng Optics) and a polarizing beam splitter (PBS, PBS051, Thorlabs Inc). When most light is reflected by the PBS to the MMF, light becomes diffusive in the sample so that the imaging resolution is determined acoustically by the acoustic lens (#45-697, Edmund optics), which collimates the PA waves. When most light transmits through the PBS to the SMF, light is tightly converged onto tissue sample, producing an optical focus coaxially and confocally aligned with the acoustic focus to optimize the detection sensitivity. The detected PA signals by the ultrasound transducer are amplified (ZFL-500LN+, Mini-circuits) and then transferred to the data acquisition card (DAQ, ATS9371, Alazar Tech), which is connected to the computer. Two-axis linear stage (L-509.10SD, Physik Instrument) is used to mount the scanning probe, which creates two-dimensional raster scanning to obtain volumetric A-line data. In our system, the lateral resolution of OR-, AR-PAM modules are about 4.5 μm and 50 μm , respectively.

2.2.2 Sample preparation

Several 6-week healthy ICR mice were anesthetized with isoflurane. Before imaging, the sample (e.g., mouse ear) was applied with ultrasound gel (Aquasonic 100) and fixed on a glass platform, beneath the water tank. The PAM probe was put above the target and immersed in the water tank to ensure acoustic coupling. All procedures involving animal experiments were approved by the Animal Ethical Committee of the City University of Hong Kong. An area of $5 \times 5 \text{ mm}^2$ of the mouse ear was imaged by OR-PAM at a step size of 2.5 μm , and then the same FOV was scanned by AR-PAM at the same step size with OR-PAM. 14 pairs of AR- and OR-PAM vascular images of different mouse ears were acquired. Apart from ears, PAM imaging of mouse brain vasculature was also conducted. The skin hairs of mouse brain were removed by using hair removal cream (Veet, Hong Kong) before the experiment. Then the scalp was disinfected and cut with surgical scissors. The exposed cerebral vessels were scanned at a step size of 2.5 μm within a FOV of $5 \times 5 \text{ mm}^2$, using AR-PAM only.

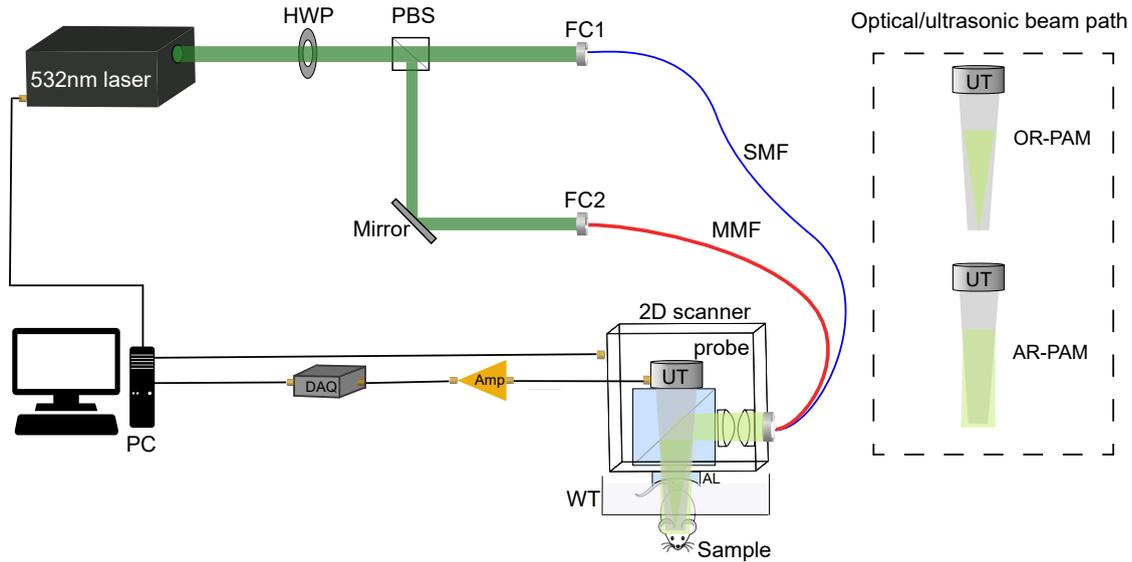


Figure 2.1: Schematic of the integrated OR- and AR-PAM system, with the optical (green) and ultrasonic (gray) beam path in the probe for OR- and AR-PAM illustrated separately. Note that the SMF and MMF are not connected to the probe at the same time but separately instead. AL, Acoustic lens; Amp, amplifier; DAQ, data acquisition; FC, fiber coupler; HWP, half-wave plate; MMF, multi-mode fiber; PBS, polarization beam splitter; SMF, single-mode fiber; UT, ultrasound transducer; WT, water tank.

To evaluate the imaging transformation performance at different depths, chicken breast tissues were sliced into different thicknesses to cover a few human hairs for AR-PAM imaging, which was used to mimic optical targets imaged at different tissue depths. We acquired AR-PAM images over a FOV of $5 \times 5 \text{ mm}^2$ of human hairs that were not covered or covered with tissues of thickness of 700, 1300, and 1700 μm , respectively. The pulse energy for AR-PAM in the phantom experiment was increased with increasingly thick chicken breast slices covered above the hair pattern.

2.2.3 Image pre-processing and data augmentation

The acquired PAM images in this study are maximum amplitude projections (MAP) of volumetric acquisitions, that is, 3D A-line data that are typically sized of (2000,2000,512) in which 2000 is the image size along each direction and 512 is the number of samples for one A-line. The A-line data needs to be processed before conducting MAP, which is based on the actual condition of raster scanning. Usually, we need to flip the A-line data of even columns, and sometimes to translate upwards or downwards the A-lines at some positions to avoid image dither or ghosting caused by motor sweeping dislocations. It is almost inevitable for the acquired PAM images to contain noise, such as isolated bright spots that compress the image

grayscale level or stain noises especially in OR-PAM images. Thus, all PAM images were first normalized to 0-1 before applying a 5×5 median filter to remove the extremely bright spots and mitigate the stain noises. After that, considering the training of deep neural network requires a large data set but the collected data was limited from experiment, data augmentation [83–85] was conducted using a Python library Albumentations [86]. There were mainly geometry and grayscale image transformation operations to teach the deep networks the desired invariance properties [85]. For geometry transformation, we conducted flipping along different directions (horizontal, vertical, and diagonal), random affine transformation (including translation, scaling and $\pm 15^\circ$ rotation), random cropping and padding, as well as elastic deformation, to mimic different spatial distributions of blood vessels. Also, 10% synthetic AR-PAM images were further blurred using a random kernel or Gaussian filter. For grayscale transformations, we had random gamma (gamma value ranging from 0.6 and 1.4) adjustment to tune the image grayscale range. 10% synthetic AR-PAM images were further adjusted on random brightness and contrast, for modeling the illumination intensity discrepancies in the imaging system. These techniques aimed to artificially increase the data distribution of available PAM images for training, with the hope for the networks to learn the robustness against deformation and gray value variations [85] and to gain better generalization ability. In this study, 14 pairs of PAM vascular images of the mouse ear were acquired experimentally. Among them, 11 pairs were used to synthesize 528 image pairs that constitute the training set. The remaining three PAM image pairs were used for network tests, without any image augmentation.

Since the acquired PAM images are of large size that our network cannot process directly, an entire PAM image is thus cut into small image patches, which also greatly increases the amount of training data. Noted that regular image patch extraction (and stitching) is enough for network evaluation on a test PAM image, while different strategy that combines accurate image patch alignment was adopted for generating the training set. This is mainly because the pixel-wise loss would be used to guide the neural network to learn a statistical PAM imaging transformation. As illustrated in Figure 2.2, a template-matching algorithm based on image intensity correlation is employed, implemented in MATLAB. The image patches of size 390×390 are first extracted successively from an entire OR-PAM image with an overlap of 64 pixels in both horizontal and vertical directions, which work as the templates to find the highest-correlation matched patches in the corresponding AR-PAM image. This is done by calculating the 2D cross-correlation matrix between the OR patch and the entire AR image, in which the maximum value indicates the most likely matched AR patch. The cropped patches in AR-PAM

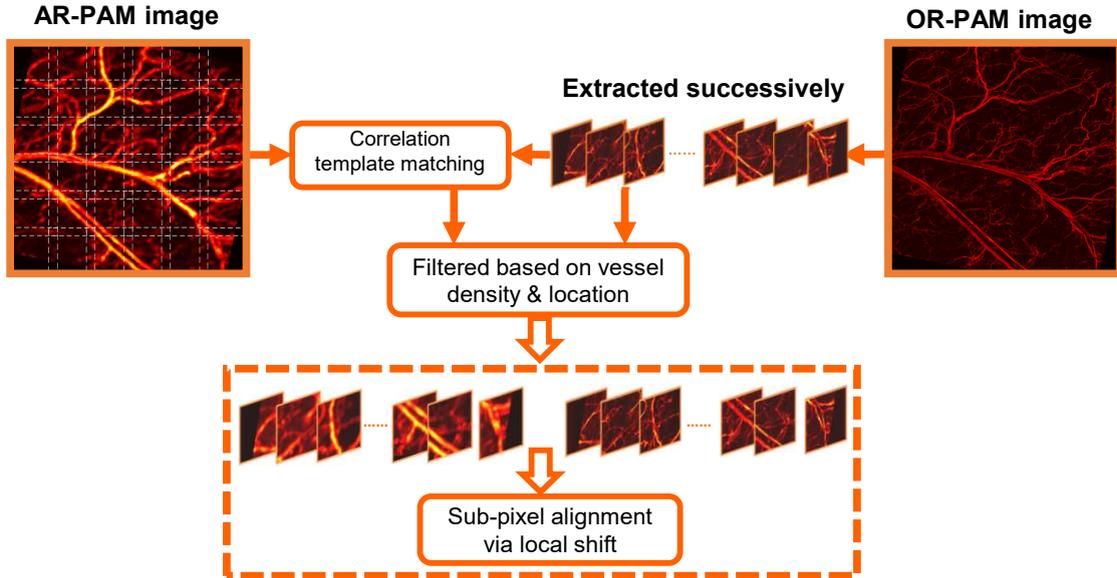


Figure 2.2: The process of image patch extraction and alignment via correlation template matching. The OR patches were extracted successively, with each used as a template to find the highest correlated AR patch. The paired image patches were filtered with the criteria for vessel density and location before being applied with sub-pixel alignment.

image will be filtered based on two criteria before forming pairs with their OR templates: if the vessel density is not enough (less than half of the mean) or the location differs greatly (more than 10 pixels in any direction), the cropped AR-OR patch would be abandoned. Note that the matched image patches are still not accurately aligned at the sub-pixel level. Thus, additional local shift (shift amount is determined by the traversal search) between the extracted image patches is applied by bilinear interpolation. Eventually, the precisely registered images are cropped with three pixels on each side to avoid registration artifacts, forming the input-label pairs of size 384×384 for network training. Also noted that the image patch size 384 is the sum of two powers of 2 (*i.e.*, $256 + 128$), which may also suit the graphics processing unit (GPU) allocation and speed up training.

2.2.4 GAN model and network training

To achieve PAM imaging transformation, we adopted a GAN-based framework for a deep neural network in this study. GAN was initially introduced by Goodfellow et al. in 2014 and has been proven a powerful generative model for super-resolution [73, 75] and many other imaging-related applications [79, 80]. There are two sub-networks in a GAN, namely the Generator and the Discriminator, being trained simultaneously. The Generator takes an AR image

as the input and produces a resolution-enhanced image, which is then passed onto the Discriminator to determine its similarity to the ground truth OR image. There is an adversarial training between the Generator G and the Discriminator D : G tries to fool D by generating an image that closely resembles its OR label, while D tries to distinguish the generated fake data from the real one. Conventionally a GAN is trained to minimize the cross-entropy error (also referred to as the Jensen–Shannon divergence) between the generated and real data distribution. However, it has been observed a GAN inclines to be unstable and difficult to converge during such training, mainly owing to the vanishing gradient problem of the Generator and model collapse [81, 87–89]. To cure the problem, Wasserstein GAN was proposed [88, 89]; it uses Wasserstein distance to replace the Jensen-Shannon divergence as the objective to be optimized. The min-max game between the two sub-networks G and D within a GAN that adopts Wasserstein distance can be formulated By

$$\min_G \max_{D \in \mathcal{L}} \mathbb{E}_{x \sim \mathbb{P}_{\text{OR}}} [D(x)] - \mathbb{E}_{z \sim \mathbb{P}_{\text{AR}}} [D(G(z))], \quad (2.1)$$

where D is subject to 1-Lipschitz function, \mathbb{P}_{OR} denotes the real OR image distribution, and the generated data distribution is implicitly defined by $G(z)$ with z following AR image distribution \mathbb{P}_{AR} . Hereby, Wasserstein GAN was used for our PAM imaging transformation.

The WGAN model used for the imaging transformation from AR- to OR-PAM is illustrated in Figure 2.3. The Generator network follows the U-Net architecture [85] that is composed of an encoder and a decoder path. The network can process an input AR image in a multi-scale fashion, enabling the network to learn the imaging transformation at various scales. The encoder path comprises four residual convolutional blocks [74, 90] that are connected by a down-sampling block. Each convolutional block is composed of two 1×1 convolutional and a 3×3 convolution, with a Leaky Rectified Linear Unit (LReLU) layer (slope 0.2) following every convolutional layer. The down-sampling block consists of a convolutional layer with a kernel of size 4 and stride 2, an LReLU layer and a Group Normalization (GN) layer [91]. After four down-sampling blocks, a 3×3 convolutional layer is bridged to the decoder path, in which the feature maps are up-sampled. The decoder, similar to the encoder, is also composed of four convolutional blocks (but no residual structure) that are connected by the up-sampling block. The up-sampling is performed with transposed convolution (also referred to deconvolution) layer, which forms the block together with an LReLU layer and a GN layer. Finally, a convolutional layer following the last convolutional block output a clearly resolved image of the same

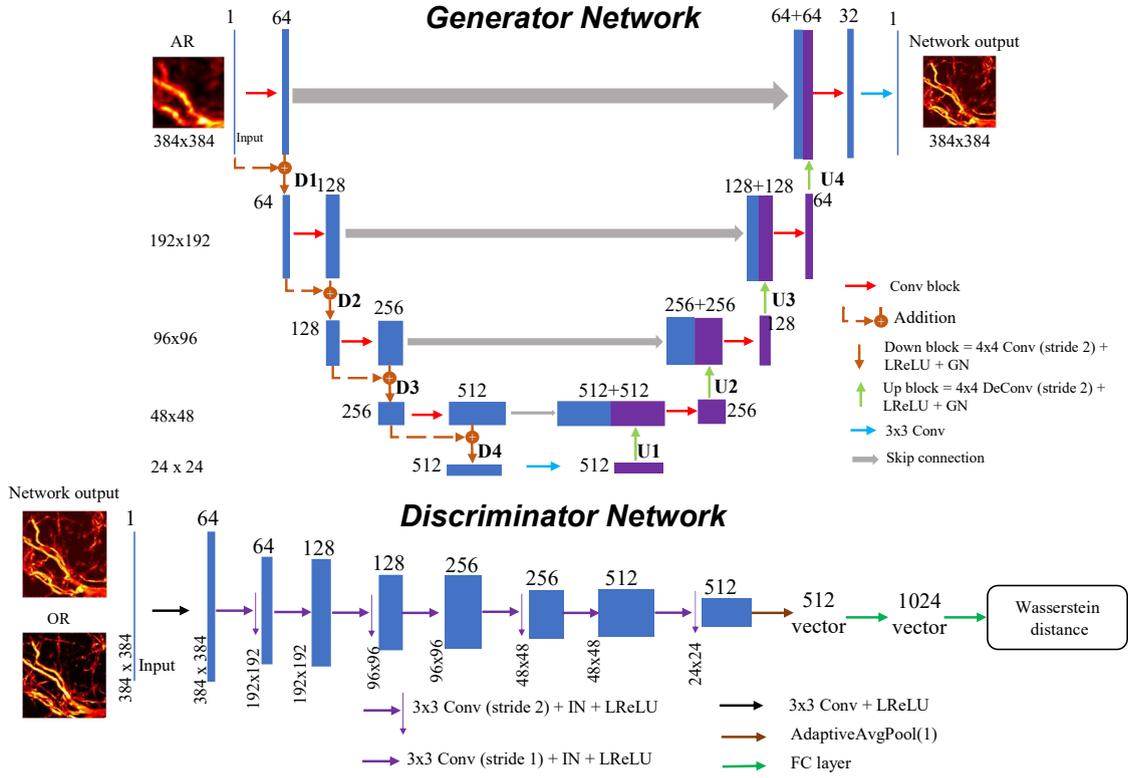


Figure 2.3: The architecture of the WGAN model used for PAM imaging transformation.

size and channel as the input. The Discriminator is a typical CNN used for image classification, except for the removal of sigmoid activation in the output layer. Starting with a convolutional layer (with LReLU activation), seven convolutional blocks are followed, in which a feature map decreases its spatial size while increasing the number of channels. Each convolutional block consists of a convolutional layer with a kernel size of 3 and stride of either 2 or 1, a Instance Normalization (IN) layer [92] and a LReLU layer (slope 0.2). Note that the down-sampling of size and the increasing of channel is conducted alternately in the convolutional blocks through the control of kernel stride and number. The output of the last convolutional block is applied with adaptive average pooling and outputs a feature map of size 1×1 . With two full-connected layers followed, the final output of the Discriminator is the scalar denoting the Wasserstein distances of input from OR image data distribution.

The behavior of optimization-based imaging transformation is principally driven by the choice of the objective/loss function. For the Generator, the primary objective is to minimize the pixel-wise loss, which is represented by the mean absolute error (MAE) between the network output image $G(I^{AR})$ and the ground truth OR image I^{OR} , given by

$$L_G^{MAE} = \frac{1}{M \times M} \sum_{i=1}^M \sum_{j=1}^M \|I_{i,j}^{OR} - G_{i,j}(I^{AR})\|, \quad (2.2)$$

where M is the image patch size. Besides, MAE in the frequency domain (FMAE) calculated from the magnitude of the 2D Fourier transform of $G(I^{AR})$ and I^{OR} is also employed, which provides the optimizer information about the vessel orientation [66], given by

$$L_G^{FMAE} = \frac{1}{M \times M} \sum_{i=1}^M \sum_{j=1}^M \left| |\mathcal{F}_{i,j}(I^{OR})| - |\mathcal{F}_{i,j}(I^{AR})| \right| \quad (2.3)$$

The perceptual loss of the Generator is defined as a weighted sum of the above two items, with the weighting factor of 1 for L_G^{FMAE} and a small weighting factor of 10^{-4} for the FMAE loss since it may contribute to training instability [93]. In addition to perceptual loss, the adversarial loss returned by the critic network D is crucial to achieving PAM imaging transformation, which provides an adaptive loss term and may help the Generator jump out of local minima. We define the Generator loss as the weighted combination of perceptual loss and adversarial loss (with coefficient γ), given by

$$L_G = L_G^{MAE} + 10^{-4} \times L_G^{FMAE} + \gamma \times (-\mathbb{E}_{z \sim \mathbb{P}_{AR}}[D(G(z))]) \quad (2.4)$$

In an attempt to enforce the Lipschitz constraint, in this study we adopt an improved Wasserstein GAN, that is, WGAN-GP [81], in which the gradient norm of the Discriminator's output with respect to its input is constrained to 1. In this case, the Discriminator loss with gradient penalty is given by

$$L_D = \mathbb{E}_{z \sim \mathbb{P}_{AR}}[D(G(z))] - \mathbb{E}_{x \sim \mathbb{P}_{OR}}[D(x)] + \lambda \times \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (2.5)$$

where $\mathbb{P}_{\hat{x}}$ denotes the random sampling distribution and λ is the penalty coefficient.

The training of our WGAN-GP model was implemented in Pytorch (v1.8.0) on Microsoft Windows 10 operating system, using a graphics workstation based on an Intel Xeon CPU, a NVIDIA 3070 GPU, and 64 GB RAM. There were 16,849 aligned pairs of PAM image patches in the training set. A small weight initialization method was adopted for the GAN, in which the initialization parameters of all convolution and deconvolution layers of the GAN, calculated by

MSRA initialization (also known as Kaiming initialization [94]), were multiplied by 0.1. Both two sub-networks were optimized using AdamW [95], *i.e.*, Adam optimizer with decoupled weight decay regularization of $\beta_1 = 0.5$ and $\beta_2 = 0.9$, and were trained with the same initial learning rate of 10^{-4} . For the loss function of the GAN model, the weight γ of adversarial loss term in the Generator loss was set to 10^{-3} and 10 for the gradient penalty coefficient in the Discriminator loss. It should be noted that to seek an adversarial equilibrium between the two sub-networks for GAN training, we can tune their learning rate or adjust the optimization times for the Generator or the Discriminator within each iteration. The total training epochs were 12 and the batch size was set as 2 for the GAN to be trained with mini-batch gradient descent, which took about 0.804 seconds for each iteration.

2.2.5 Blind deconvolution for AR-PAM image deblurring

Compared with high-resolution OR-PAM vascular images, images acquired with AR-PAM *in situ* have lower spatial resolution and are visually blurry. From the perspective of image deconvolution, it is reasonable to treat the OR-PAM image as the object itself while the corresponding AR-PAM image as the result of a convolution of the object and the system point spread function (PSF). As it is infeasible to model the PSF of such a conceptual PAM imaging system, we turn to use statistical blind deconvolution to iteratively recover the object and improve the estimation of PSF with an initial guess from a blurry AR-PAM image. This functions as the baseline, for a beneficial comparison with the deep learning enabled PAM imaging transformation regarding the performance of deblurring or resolution improvement.

Note that blind image deconvolution, as a highly ill-posed inverse problem, requires estimating both the blur kernel and object from a degraded image. Currently, most blind deconvolution methods fall into the variational Bayesian inference framework [96], with main differences coming from the form of the likelihood, the choice of priors on the object, and the blur kernel and the optimization methods to find the solutions [97]. Here, we used a general blind deconvolution method that adopts expectation-maximization optimization, to find the maximum posterior solution with flat priors. Besides that, a fractional-order total variation image prior was also tried [98], as the total variation is a popular regularization technique in image deconvolution. The blind deconvolution was implemented with 30 iterations for each input blurry image using MATLAB.

2.3 Results

2.3.1 Network feasibility: evaluation with *in-vivo* mouse ear photoacoustic images

The feasibility of the WGAN-GP network was evaluated with the PAM image pairs of living mouse ear vasculature that were not included in network training. The results are shown in Figure 2.4. Visually, the improvement of resolution is obvious with the network transformation, and quite a few small vessels that have been hidden in the AR-PAM image are now resolved by the network. To better evaluate the improvement, three regions of interest (ROIs) indicated by a white dashed box in AR-PAM, network output, and OR-PAM images are chosen and compared. Taking a close look at them, clearly resolved blood vascular details are presented in the network output, which matches well with the ground truth (OR-PAM image) in the same region. Moreover, the signal intensity profiles along the cyan dashed lines within each ROI are compared. As seen, the AR-PAM imaging tends to generate overly smoothed signal intensity profiles due to its low resolution, while the network is capable of distinguishing vessels hierarchically. The sharp signal intensity profiles inside each ROI of the network output shares good consistency with those of the ground truth OR-PAM image, verifying the feasibility and reliability of our PAM imaging transformation. To explore the deblurring effect of the AR-to-OR network deeply, we here give specific analyses with three examples. The first example is that a blurred vascular plexus denoted by the blue arrow has been clearly resolved by the network, which matches well with the ground truth. Next, the green arrow shows a single capillary that is missed in the AR-PAM image and barely discernible in the network output but is clearly shown in the OR-PAM image. Another example described by the purple arrow is the network find a limitation in resolving some closely spaced parallel blood vessels. It suggests that given extremely blurry pixels, the network may fail to reconstruct the full feature of the target. That said, this is also the point to indicate the feasibility and capability of the network of enhancing AR images while maintaining fidelity without generating fake features. Our findings reveal that the deblurring performance of the network is highly dependent on the quality, such as signal-to-noise ratio, of the given AR-PAM image. It is the case that most of the blurry blood vessels could be clearly resolved by the network, which possess good fidelities and meanwhile, some subtle distinctions from the ground truth OP-PAM images.

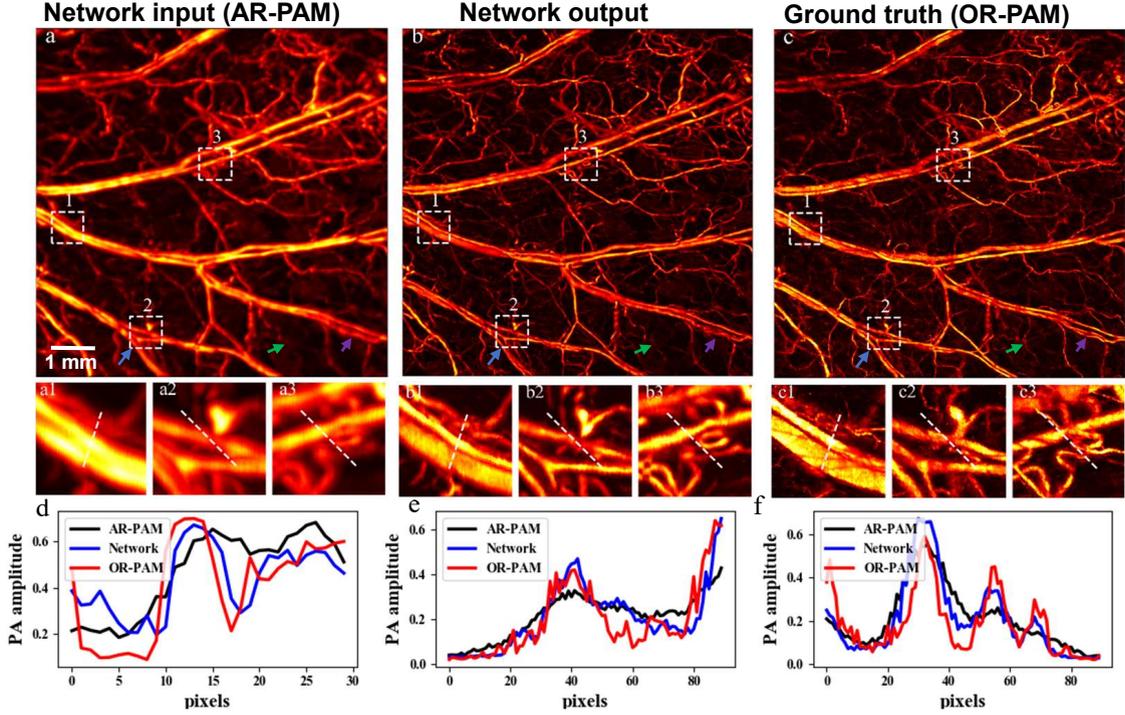


Figure 2.4: An experimentally obtained AR-PAM vascular image of mouse ear (a) is fed to the trained WGAN-GP model for imaging transformation. The resultant network output (b) is comparable to the ground truth OR-PAM image (c) of the same sample. Three ROIs marked with the white dashed boxes in (a1-a3) AR-PAM image, (b1-b3) transformed results, and (c1-c3) the ground truth OR-PAM image respectively, are enlarged and compared. Comparison of the cross-sectional profiles along the white dashed lines inside (a1, b1, c1), (a2, b2, c2), and (a3, b3, c3) are also provided in (d), (e) and (f) respectively. The blue arrow in (a-c) represents a vascular plexus that is originally blurred in AR-PAM but is now clearly resolved by the network; the green arrow shows a single capillary which is missed in the AR-PAM image and barely discernible in the network output, while clearly shown in the OR-PAM image; the purple arrow indicates a failure for the network to resolve some closely spaced parallel blood vessels that show up in the ground truth image.

2.3.2 Network performance: comparison and characterization

The performance of the network transformation is characterized by two aspects. First, the deblurring performance on mouse ear AR-PAM images of the network is compared with that using a blind deconvolution method, as shown in Figure 2.5. Apart from perceptual quality, two representative metrics including peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [99] are also provided for comparison. PSNR is defined via the mean squared error (MSE) between an image to be evaluated and its ground truth OR image and the logarithmic form is given by:

$$PSNR = 20 \log_{10} \frac{MAX_{OR}}{\sqrt{MSE}} \quad (2.6)$$

Table 2.1: Quantitative comparison between deep learning and blind deconvolution in evaluating image enhancement, in which the metrics are represented in the form of mean \pm standard deviation.

Method	PSNR	SSIM	PCC
AR-PAM	16.77 \pm 2.61 dB	0.54 \pm 0.06	0.76 \pm 0.08
Blind Deconvolution	18.05 \pm 1.71 dB	0.27 \pm 0.07	0.76 \pm 0.09
Network output	20.02 \pm 1.51 dB	0.61 \pm 0.05	0.78 \pm 0.08

in which MAX_{OR} denotes the maximum of OR image. SSIM evaluates an image quality perceptually, which also incorporate luminance and contrast information, defined as:

$$SSIM(I, I^{OR}) = \frac{(2\mu\mu_{OR} + c_1)(2\sigma_{cov} + c_2)}{(\mu^2 + \mu_{OR}^2 + c_1)(\sigma^2 + \sigma_{OR}^2 + c_2)} \quad (2.7)$$

where $\mu(\sigma^2)$ and $\mu_{OR}(\sigma_{OR}^2)$ are the mean (variance) of an image to be evaluated and its OR label, respectively; σ_{cov} denotes the covariance between the two images; c_1, c_2 are the constants to stabilize the division, respectively. In Figure 2.5, two ROIs marked in a white dotted box are enlarged and compared. For ROI 1 in the first ear vasculature image (Figure 2.5a), the single capillaries are resolved well for both blind deconvolution and network output, while the latter resembles better to the OR-PAM result. In ROI 2, the vessel bifurcation contains different types of vessels that the deconvolution method fails to produce rich vessel details, but the network output can distinguish vessels hierarchically. In the second case (Figure 2.5b), we show that the WGAN-GP model can easily separate the large arteries and the small veins attached, as shown in both two ROIs, while the deconvolution method can merely distinguish them partially. It is worth noting that neither deconvolution nor the network could match the ground truth in every detail because some subtle capillaries were missed in the original AR-PAM image, as mentioned earlier. Also, note that compared with the AR-PAM image that appears overly smooth, the network output gets sharp with latent capillary artifacts generated sometimes (see d2 and e2 in Figure 2.5), which might undermine the tricky metric like SSIM. In short, the network significantly outperforms blind deconvolution in deblurring blood vessels that the clearly resolved images are perceptually comparable to the experimentally acquired OR-PAM images.

Apart from qualitative analyses, enhancements on mouse ear data are compared and quantified by calculating metrics, including PSNR, SSIM, and PCC (Pearson correlation coefficient). PCC

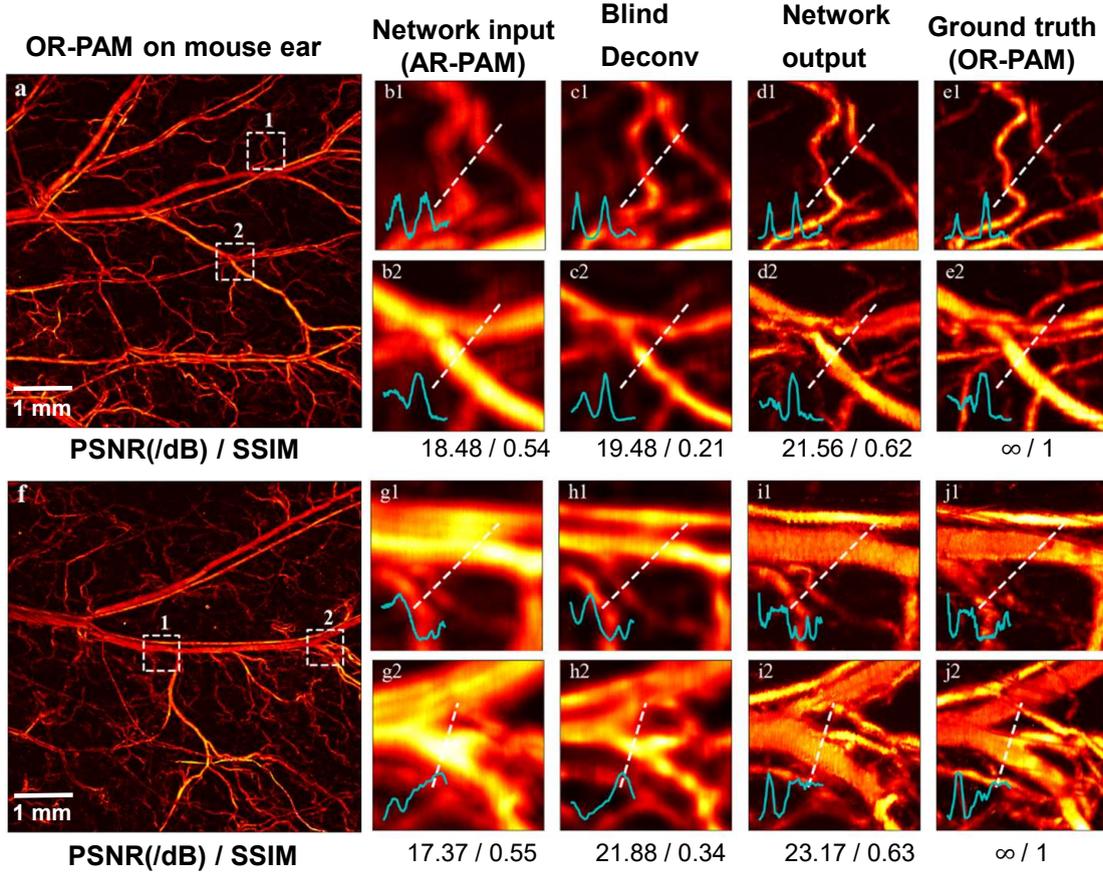


Figure 2.5: Qualitative deblurring performance of deep learning in comparison with that using blind deconvolution. Two examples of mouse ear vascular images are presented, shown in the entire OR-PAM images (a) and (f). For each example, two ROIs marked with white dashed box are enlarged and compared. In the first example (a), (b1-e1) are for ROI 1 and (b2-e2) for ROI 2; in the second example (f), (g1-j1) are for ROI 1 and (g2-j2) for ROI 2; all correspond to AR-PAM, blind deconvolution, GAN output, and ground truth, respectively. Cross-sectional profiles along the white dashed lines inside (b1-e1), (b2-e3), (g1-j1) and (g2-j2) are provided for comparison. Metrics like PSNR and SSIM with respect to the entire OR-PAM image are also provided for reference.

is expressed by:

$$PCC(I, I^{OR}) = \frac{\sigma_{cov}}{\sigma \cdot \sigma_{OR}} \quad (2.8)$$

where σ_{cov} denotes the covariance between an image to be evaluated and its OR label, σ and σ_{OR} are the standard deviation of the image and its OR label, respectively. The comparison was performed between test image patches, with the results given in Table 2.1 The network produces overall better results than the blind deconvolution method in improving all three metrics. The small variance in all the metrics for the network outputs also indicates the robust performance of our AR-to-OR network. To be specific, the mean PSNR improves from ~ 16.77 to ~ 20.02 dB, while SSIM (improved by 13% averagely) and PCC (averagely by 3%) only see modest improvement. Noted that the blood vessel image is overall sharpened by the

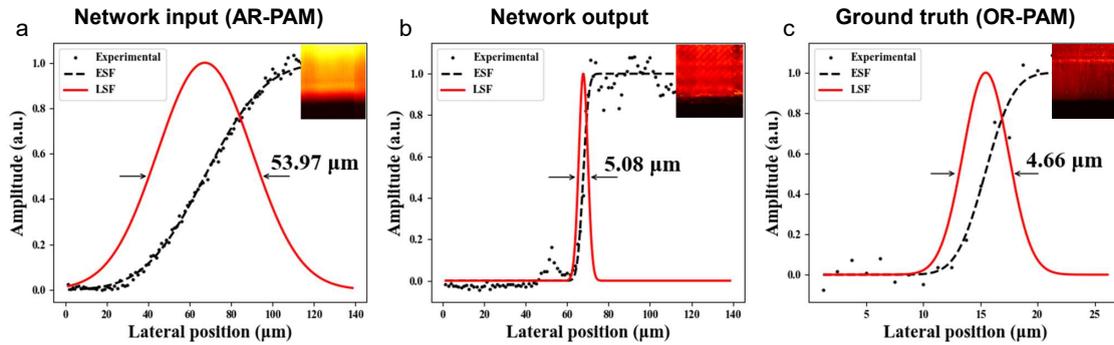


Figure 2.6: Demonstration of lateral resolution enhancement of AR-PAM by deep learning. Lateral resolution of (a) AR-PAM, (b) network output, and (c) OR-PAM was measured to be $\sim 54.0 \mu\text{m}$, $\sim 5.1 \mu\text{m}$, and $\sim 4.7 \mu\text{m}$, respectively, using the edge response of a sharp blade. ESF, edge spread function; LSF, line spread function. The color insets are the blade images of AR-PAM, network output, and OR-PAM, respectively.

network, some blurred capillary discrepancies also become sharper. As metrics like SSIM and PCC are very sensitive to these artifacts, only modest improvements are found in these metrics, although the image resolution improvement is significant, as can be seen from Figure 2.5. Regarding the deconvolution method, it improves PSNR slightly for the test set, while failing to improve the other metrics, even degrading SSIM. It should be mentioned that the results were obtained using flat prior, a general image prior in the blind deconvolution method. Even worse results were found with the popular total variation regularization [98]. All these suggest that the enhancement of AR-PAM images towards high-resolution OR-PAM images is challenging with blind deconvolution. In contrast, deep learning enabled PAM imaging transformation may help solve this tricky inverse problem, with perceptual satisfactory deblurred results and improved metrics.

To further characterize the spatial resolution improvement after the network transformation, the edge response of a sharp blade was measured with the OR- and AR-PAM settings. As illustrated in Figure 2.6, the normalized experimental PA data were fitted by edge spread function (ESF, black dashed line); whose derivative gave the line spread function (LSF, in red line). The full width at half maximum (FWHM) of LSF was used to represent the system's lateral resolution. We can see that in the network output, the edge response curve in situ has become much steeper, which means that the blurred edge in AR-PAM imaging is now clearly resolved. The lateral resolution has been accordingly improved from $\sim 54.0 \mu\text{m}$ in AR-PAM to $\sim 5.1 \mu\text{m}$, which is quite comparable to that of the ground truth (OR-PAM), $\sim 4.7 \mu\text{m}$. Although the measurements might vary slightly at different locations of the blade edge, this exemplifies the significant resolution enhancement via our network transformation.

2.3.3 Network generalization: application for *in-vivo* mouse brain photoacoustic imaging

Thus far, the feasibility and effect of the proposed network to achieve OR-PAM resolution based on AR-PAM images have been demonstrated, where both training and test data sets were obtained from *in-vivo* mouse ears. To validate the generalization of the network, *in-vivo* mouse cerebrovascular images were acquired and fed into the network; only AR-PAM images were available in this group of experiments, following a realistic application scenario without labeling. Figure 2.7 shows the original AR-PAM and network output images of mouse brain vasculature. It can be observed that the network output has sharper vascular patterns and enhanced vascular signals. Two ROIs indicated by white dashed boxes in both AR-PAM image and network output are enlarged and compared. Significantly improved image quality was achieved by our network, free from noticeable artifacts. The vascular signal intensity profiles for the same region along the horizontal direction are also used to assess the transformation performance. We can see that the network output follows basic trends of vascular signals in the AR-PAM image but yields many finer details and can clearly distinguish different vascular signals. This is consistent with the fact that enhanced intensities and sharper cerebral vessels were produced. Even without a ground truth OR-PAM image, the above comparisons could, to some degree, verify the reliability of our approach and the significant improvement it achieves. More importantly, it is worth mentioning that even the given mouse brain data has quite different vascular structures from the ear and some cerebral vessels are within at deeper tissues, the trained network can still cope with them, which verifies the universal applicability of the proposed method to the brain vasculature.

2.3.4 Network application: preliminary extension for deep-tissue OR-PAM

Thus far, the feasibility and effect of the proposed network to achieve OR-PAM resolution based on AR-PAM images have been demonstrated, where both training and test data sets were obtained from *in-vivo* mouse ears. To validate the generalization of the network, *in-vivo* mouse cerebrovascular images were acquired and fed into the network; only AR-PAM images were available in this group of experiments, following a realistic application scenario without labeling. Figure 2.7 shows the original AR-PAM and network output images of mouse brain vasculature. It can be observed that the network output has sharper vascular patterns and enhanced vascular signals. Two ROIs indicated by white dashed boxes in both AR-PAM image

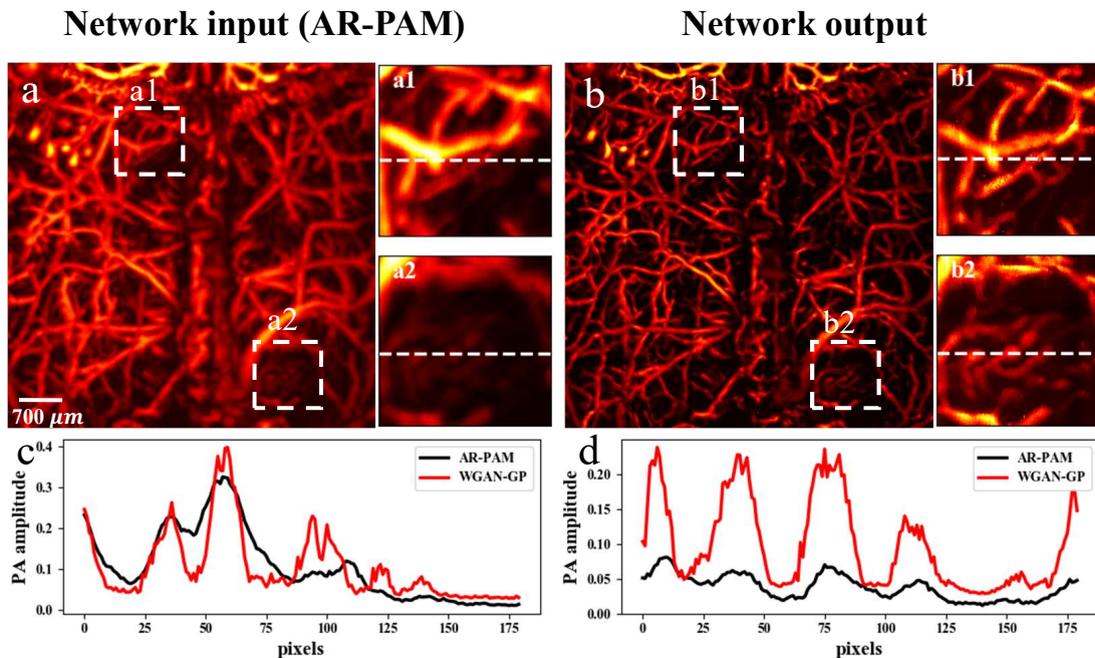


Figure 2.7: Application of the network on *in-vivo* mouse brain AR-PAM images. (a) is the network input (AR-PAM image) and (b) is the network output. Two ROIs in both network input (a1, a2) and output (b1, b2) are enlarged and shown. Comparison of signal intensity profiles along the horizontal dashed line in (c) the first ROI and (d) the second ROI are also given.

and network output are enlarged and compared. Significantly improved image quality was achieved by our network, free from noticeable artifacts. The vascular signal intensity profiles for the same region along the horizontal direction are also used to assess the transformation performance. We can see that the network output follows basic trends of vascular signals in the AR-PAM image but yields many finer details and can clearly distinguish different vascular signals. This is consistent with the fact that enhanced intensities and sharper cerebral vessels were produced. Even without a ground truth OR-PAM image, the above comparisons could, to some degree, verify the reliability of our approach and the significant improvement it achieves. More importantly, it is worth mentioning that even the given mouse brain data has quite different vascular structures from the ear and some cerebral vessels are within at deeper tissues, the trained network can still cope with them, which verifies the universal applicability of the proposed method to the brain vasculature.

To further explore the application of the proposed network for deep-tissue PA imaging, we prepared a hair phantom by covering human hairs with chicken breast slices of different thicknesses. As the thickness of the tissue sample can be adjusted gradually, it is possible to find out the maximum imaging depth that our network transformation could handle in the experiment.

Figure 2.8 shows the evaluation results of our network based on AR-PAM images of hair pattern that were not covered or covered with tissue slice of a thickness of 700, 1300, and 1700 μm . Note that there was slight position shift of the hairs beneath when changing the tissue slices of various thicknesses, which, however, does not affect the evaluation of image enhancement at different depths. Since ground truth OR-PAM images were no longer available in these tissue depths, we thus used contrast-to-noise ratio (CNR) and hair edge 10-90% rise distance [100], to indicate the imaging SNR and resolution performance under different penetration depths. The logarithmic CNR using the decibel scale is given by:

$$CNR = 20 \log_{10} \frac{\mu_{\text{object}} - \mu_{\text{background}}}{\sigma_{\text{background}}} \quad (2.9)$$

where μ_{object} and $\mu_{\text{background}}$ denote the mean intensity of hair object and background noise respectively, and $\sigma_{\text{background}}$ the standard deviation of background noise. Practically, the white dashed boxes in both the AR-PAM images (Figure 2.8 a1-a4) and network outputs (Figure 2.8 b1-b4) were selected as the ROIs to measure the object signal, while the bigger yellow dashed boxes were denoted as the backgrounds. Note that altogether 10 different areas of object and background were used for average CNR calculation at each penetration depth, but only one ROI is marked in Figure 2.8 for conciseness. Regarding the measured hair edge sharpness under different imaging depths, the 10-90% rise distance of hair edge at the positions marked by the cyan dashed lines were used. Again, multiple such positions were selected to obtain an averaged metric at each depth.

From Figure 2.8 (a-b), we can see that for AR-PAM images, overall both hair signal and the contrast decrease with increasing penetration depth, due to the weak and diffusive optical excitation in situ under increasingly strong scattering. Note that the signals in some target areas at the thickness of 1700 μm case might be a bit stronger, mainly due to the increased laser pulse energy and the possible existence of microporous structure in the chicken tissue slices. Nevertheless, the overall image signal in 1700 μm is weaker than that at 1300 μm , in which the hair signal incompleteness caused by strong optical scattering was more serious. The network outputs follow a similar trend, but enjoys enhanced contrast, especially for penetration depths of 1300 and 1700 μm that have been sufficiently beyond the diffusion limit. Quantitatively, as shown in Figure 2.8 (c-d), the CNR drops from ~ 34.9 dB to ~ 10.6 dB and the measured 10-90% rise distance increases from ~ 57 μm to ~ 95 μm for AR-PAM images in the range of 1700 μm tissue depth. With deep learning, the CNR declines only to ~ 17.1 dB. More strikingly, the

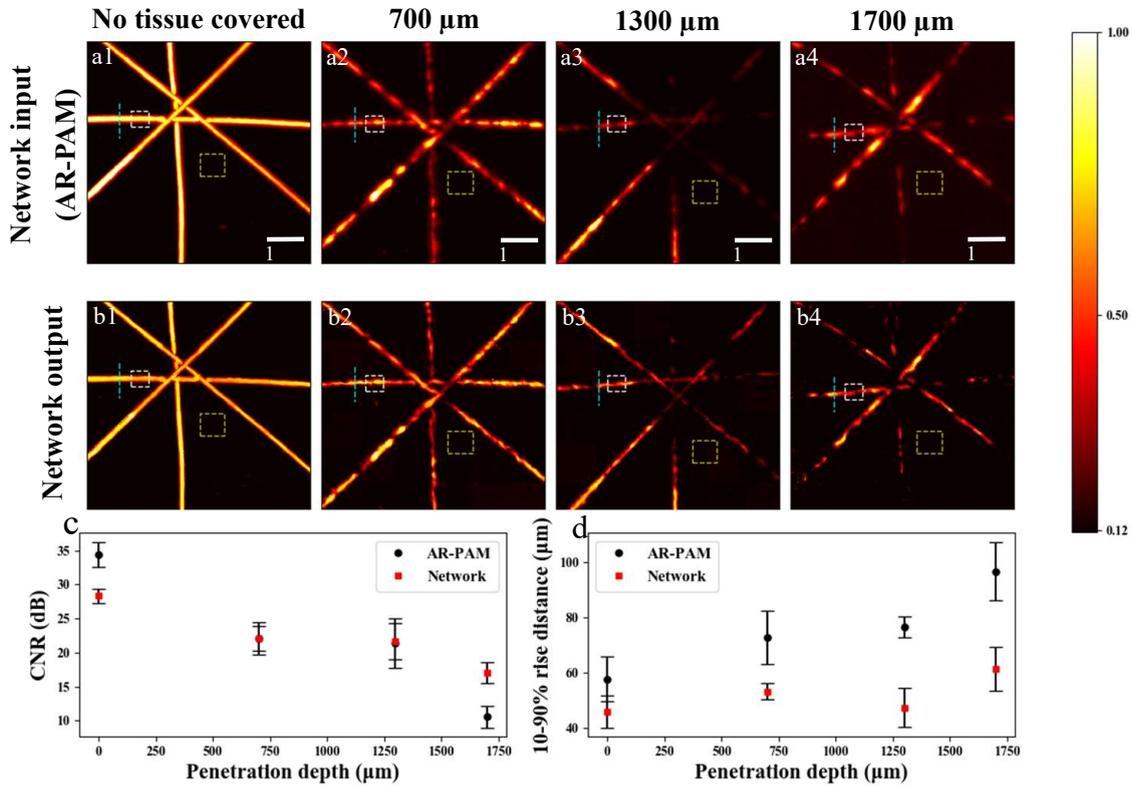


Figure 2.8: Preliminary demo for deep-penetrating OR-PAM imaging using a hair phantom covered with chicken tissues of different thicknesses. (a1-a4) Experimentally acquired AR-PAM images. (b1-b4) Network output results corresponding to tissue thicknesses of 0 (no tissue covered), 700, 1300, and 1700 μm . The white and green dashed boxes in (a1-a4) and (b1-b4) denote the ROIs for object and background, respectively, and the cyan lines indicate the positions for measuring the hair edge 10-90% rise distance. (c) CNR and (d) measured hair width versus different penetration depths (tissue thicknesses) for both AR-PAM images and the network output results.

measured 10-90% rise distances merely see a slight increase from $\sim 46 \mu\text{m}$ to $\sim 60 \mu\text{m}$, suggesting a greatly improved lateral resolution for tissue depth up to 1700 μm . To sum up, this hair phantom experiment reveals potentials for deep-tissue OR-PAM with our approach, which can remarkably enhance imaging resolution and SNR.

2.4 Discussion

To achieve high-resolution PAM imaging in deep tissue, we adopt the WGAN-GP model to transform blurry images acquired with an AR-PAM to match the corresponding OR-PAM results, thanks to an integrated PAM setup. The network was first trained with aligned AR- and OR-PAM data sets experimentally obtained from *in-vivo* mouse ears. Then, the feasibility of the network was validated with the same type of data but was not used in the network training,

yielding superior lateral resolution comparable to the ground truth OR-PAM images. It should be emphasized that the successful applications to *in-vivo* mouse brain AR-PAM imaging verify the universal applicability of the developed network, as it was trained with data solely from the mouse ear. With an AR-PAM setting boost by the network transformation capability, our method could thus be extended to deep-tissue phantoms where experimental OR-PAM is not possible. Apart from the benefit for deep penetration, it may also initiate other potentials: an inexpensive multimode pulsed laser source can be adopted to reduce the system cost, as the beam quality requirement in AR-PAM is way less demanding; a larger scanning step size may also be adopted for AR-PAM imaging which could significantly reduce the time of raster scanning. All these, in combination, empower conceivable all-round boost of penetration depth, cost control, as well as imaging speed to OR-PAM. In particular, the image transformation network could be easily integrated with a regular AR-PAM setting for image enhancement. Our method is in principle more generic than a similar work of PAM super-resolution where deep learning is used for single-molecule localization PAM reconstruction with much fewer frames [101]. Essentially, the method can recover the high-frequency components of blurry AR-PAM images, producing more clearly resolved vascular patterns thanks to the end-to-end, pixel-level image processing of the network.

That said, a few more aspects need to be discussed herein. First, the OR-PAM imaging is limited to superficial tissues (depth within several hundred micrometers) in our setup as a laser of 532 nm wavelength is used. Although using a near-infrared laser may promise a larger penetration depth, the lateral resolution of OR-APM is worse than ours, which is less desired as the target resolution to be learnt by AR-PAM. Second, when applying the image transformation network to deep tissue regions where the ground truth OR-PAM is not available, the results of network output can be ensured by comparing the changes of signal profiles on a target region, which should be consistent and get enhanced reasonably, as done in Figure 2.7 (c-d). Besides that, we can also use two label-free metrics including CNR and 0-90% rise distances (edge sharpness) for evaluation, as done in Figure 2.8 (c-d). Third, since the training dataset included those obtained under artificial geometry and grayscale transformation, the trained model can reasonably learn the robustness against variations in imaging conditions like laser dosage change, mechanical drift or other noises. However, we did not conduct specific tests for quantitative analysis because the data acquired from different samples (mouse ear, brain, hair phantom) might encounter small differences of conditions but all showed enhanced results.

There are mainly two limitations of our approach. The first is related to the artifacts, which

might be generated by the network due to several reasons. It should be noted that not all the blood vessels inside the paired AR- and OR-PAM images are consistent due to the relatively large discrepancy of object-resolving capability. Some physically existing local capillaries might be missing in AR-PAM, with only blurry and dispersive pixels instead. Such inconsistency could be an obstacle for the network to conduct pixel-to-pixel transformation in the training phase. In the network test phase, the image enhancement performance of the network is highly dependent on the SNR of the AR-PAM input. The network could significantly enhance the blood vessel image of *in-vivo* mouse brain (with scalp removed) and moderately improve the thick-tissue hair phantom results. It makes sense that inaccurate predictions may be produced where the input pixels are extremely blurry. Hence, there may be some subtle distinctions between the network output and the ground truth OR-PAM image, mainly because of the latent capillary missing and artifacts. These subtle distinctions lead to limited improvements of metrics like SSIM (averagely by 13%) and PCC (averagely by 3%) that are sensitive to the perfect pixel-by-pixel match between the network output and ground truth. The artifacts can be distinguished visually if there are additional or missing capillaries in the network output, compared to the OR-PAM result. Besides, the artifacts also exist if the signal profiles of the same tissues are inconsistent with those of OR-PAM.

To minimize the latent artifacts and improve the metrics, we adopted data preprocessing and network training methods carefully. Accurate registration in the process of image patch extraction is of crucial importance, in which the sub-pixel alignments between the AR and OR image pairs allow the network to optimally learn a pixel-to-pixel transformation. Besides, the design and training of the network also play an important role. Different from conventional CNN where the optimization is solely driven by a pixel-wise loss that tends to produce overly smoothed results, the WGAN-GP model we used benefits from the adversarial training where the adaptive adversarial loss is crucial to guide the Generator network to generate images resembling OR labels. Further improvement could be made by acquiring more PAM image data used for training, to equip the network with stronger and more general imaging transformation ability. Lastly, it should also be emphasized that the aim of the network is not to perfectly transform AR-PAM images in all details, but to approximate the resolution of OR-PAM as possible as it can.

The other limitation is our transformation method still finds a distance to ideal deep-tissue OR-PAM imaging. As mentioned earlier, the quality of the network output is highly dependent on that of the input AR-PAM data; the low SNR and spatial resolution of AR-PAM imaging at

depths in tissue pose a physical limit for our image-based postprocessing method. Nevertheless, with the recent developments towards faster and more efficient photoacoustically guided wavefront shaping [10, 102, 103], we believe ideal OR-PAM imaging at depths in the tissue will be possible soon.

Recently, a similar research was reported [104], where simulated rather than experimentally acquired AR-PAM data generated by blurring the corresponding OR-PAM images were used and did not experimentally demonstrate its ability of deep-penetration imaging. In addition, its conventional network model is also different from the WGAN-GP model in our study. Also note that the proposed PAM imaging transformation method in this work should be distinguished from single image super resolution and blind deconvolution, although they are closely related. For image super resolution, the aim is to reconstruct the baseline resolution when given an input solely down-sampled from the baseline image; but for our network, it is to approximate the resolution of OR-PAM from another poorer imaging modality. For the blind deconvolution method, it has been shown earlier that it failed to recover high-quality images, whether in flat priors or fractional-order priors. The deblurring performance is inclined to be suboptimal, as it is often difficult to determine the PSF of the conceptual imaging system from ground truth OR to AR images. What's more, the computation cost of deconvolution is demanding due to the multiple iterations it requires for parameter updates. In our example, it took about one and half minutes to produce a deblurred image with 30 iterations on average. In comparison, the trained network could rapidly output a high-resolution image from the blurry input within seconds.

2.5 Conclusion

In this study, a WGAN-GP neural network is designed and trained based on co-registered AR-to-OR PAM images experimentally acquired from *in-vivo* mouse ears. The feasibility and reliability of the proposed network to improve imaging resolution are demonstrated *in vivo*. The network outputs have also been compared with those obtained with a blind deconvolution method, showing perceptually better image deblurring results and improved evaluation metrics. Moreover, the transformation capability can be extended to other types (*e.g.*, the brain vessel) or deep tissues (*e.g.*, chicken breast slice of 1700 μm thickness) that are not readily accessible by OR-PAM. Note that the proposed method has its limitations, such as the existence of artifacts in the network output and the performance dependence on the input data quality

(AR-PAM images). As an extension to the imaging depth of OR-PAM via computation, the proposed method potentially provides new insights to expand the scope of OR-PAM towards deep-tissue imaging and wide applications in biomedicine.

Chapter 3

Patterned Light Delivery through Multimode Fiber based on Wavefront Optimization

This chapter is reproduced with some adaptations according to two published journal papers:

*(1) Shengfu Cheng, Tianting Zhong, Chi Man Woo, Qi Zhao, Hui Hui, and Puxiang Lai, "Long-distance pattern projection through an unfixed multimode fiber with natural evolution strategy-based wavefront shaping", *Optics Express* 30(18):32565-32576 (2022).*

*(2) Shengfu Cheng, Tianting Zhong, Chi Man Woo, and Puxiang Lai, "Alternating projection-based phase optimization for arbitrary glare suppression through multimode fiber", *Optics and Lasers in Engineering* 161:107368 (2023).*

In this chapter (Part II of the thesis), the focus shifts to the control of light delivery through MMF for deep-tissue high-resolution applications, including focusing or suppressing light into an arbitrary pattern, which are both based on computational WFS. By adopting a natural gradient ascent-based wavefront optimization strategy, we first achieve high-quality pattern projection through complex media. This approach, coupled with a novel fitness function based on cosine similarity, enables long-distance projection of arbitrary patterns through a 15-meter MMF, showcasing enhanced performance in focusing contrast and pattern fidelity in unstable environments. This advancement holds promise for deep-tissue applications requiring precise energy delivery, such as phototherapy and optogenetics. Apart from focusing application,

WFS can also be applied to glare suppression that reduces speckle background through complex media. For glare suppression in a target region, current methods are either slow or not sufficiently generic. Here, an alternating projection method that fully exploits the knowledge of TM is proposed for fast and arbitrary glare suppression. Parallely, multiple phase masks corresponding to various target regions can be computationally optimized without iterative hardware feedback. Further, we experimentally demonstrate effective glare suppression for target regions of various shapes and sizes at the distal end of an MMF with fast frame rate. This method may enable MMF-based imaging, sensing, and speckle optical tweezer in complex environments.

3.1 Introduction

The control of light propagation through complex media, such as an MMF, has been desired with profound implications for deep-tissue biomedical applications. The inherent challenge lies in the medium's propensity to light, transforming a coherent beam into seemingly random speckle pattern, thereby obscuring the potential for precise light delivery. However, advancements in WFS techniques [8, 105, 106] have opened new avenues for controlling light in such media, enabling both the focusing of light into any desired pattern and the suppression of unwanted scattered light in a target region. This chapter synthesizes findings from two of our published research articles about light control through an MMF, highlighting the dual capabilities of focusing light and suppressing glare based on wavefront optimization.

The first study delves into the realm of focusing light through MMFs. Compared with typical single-spot focusing, focusing light into an arbitrary pattern is often favored in energy delivery-related biomedical applications, yet more demanding to achieve satisfactory pattern fidelity and focusing contrast, especially in noisy and instable environments. There have been TM [18, 107, 108] and iterative [109–113] methods to overcome the above obstacles. Compared to TM, which needs additional optimization of the SLM hologram to mitigate the interference effect in multi-spot focusing, iterative methods are more straightforward and robust. Existing works for patterned projection through fixed complex media are mostly based on intelligent optimization algorithms such as GA [109–112] or improved ant colony algorithm (IACO) [114]. Hence, they rely more or less on random variations of the population, which might be easily affected by external interference and hence limited by the optimization speed and efficiency.

Conversely, the second study addresses the challenge of suppressing scattered light, namely glare suppression. It originates from the reflection-geometry scenarios [115, 116] where glare from backscattered light degrades the SNR, such as in LiDAR detection and deep-tissue optical imaging. Traditional methods for glare suppression, including coherence gating [117] and time-of-flight differentiation [116], are limited by technical challenges and temporal resolution. It should be noted that suppressing the scattered light intensity in a transmission geometry is also valuable for imaging and sensing applications, such as speckle optical tweezer [118] or blind structured illumination microscopy [119]. Existing progress on glare suppression through complex media employed GA [115] or Hadamard encoding algorithm [120], which could be time-consuming because of their nature of feedback-based iterative optimization and the use of a slow SLM. There is also an aperture-target TM method by constructing the low-transmittance eigenchannel [121], which requires complex amplitude modulation and might not be generic enough, as a new aperture-target TM for decomposition is required for each target region.

In this chapter, we first suggest using a natural evolution strategy (NES) for the wavefront optimization in projecting a pattern through MMF. NES is based on natural gradient ascent for parameter optimization [122], and has demonstrated excellent anti-interference ability, which is essential for iterative WFS [14, 123] in perturbed environment. Specifically, we adopted the separate NES (SNES), coupled with a novel fitness function, vector cosine similarity (VecCos), for enabling the projection of arbitrary patterns through a 15-meter-long, unfixed MMFs with enhanced contrast and speed. As for glare suppression, employing a TM-based alternating projection (AlterProjTM) algorithm is proposed here for phase optimization that achieves rapid and effective suppression of scattered light in targeted regions. With the knowledge of TM, multiple phase masks that correspond to different target regions can be optimized in parallel, which is further accelerated by GPU execution. It shows by simulation that a suppression factor of $\sim 10^{-3}$ was achieved with only 30 iterations even for a large target region that contains over a hundred speckles, under phase-only modulation. Experimentally, light is effectively suppressed in target regions of various shapes and sizes at the distal end of an MMF by using a DMD at a frame rate of ~ 50 frame per second (fps).

3.2 Methods

3.2.1 Principle of SNES

In the scheme of SNES, samples are drawn from the standard multivariate Gaussian distribution $\mathbf{s}_k \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$ with a diagonal covariance matrix. The distribution parameters $\theta = (\boldsymbol{\mu}, \boldsymbol{\sigma})$ where $\boldsymbol{\mu} \in \mathbb{R}^N$ and $\boldsymbol{\sigma} \in \mathbb{R}_+^N$ denote the vectors of the mean and standard deviation respectively, map a population of N_p samples into the search distribution (*i.e.*, input wavefronts) such that $\mathbf{z}_k = \boldsymbol{\sigma} \cdot \mathbf{s}_k + \boldsymbol{\mu}$ ($k = 1, \dots, N_p$). Suppose we have $\mathbf{z}_k \sim \pi(\mathbf{z}|\theta)$, which is evaluated by a fitness function f , the expected fitness of the search distribution is

$$J(\theta) = \mathbb{E}_\theta [f(\mathbf{z})] = \int f(\mathbf{z})\pi(\mathbf{z}|\theta)d\mathbf{z}. \quad (3.1)$$

To maximize $J(\theta)$ with gradient ascent, the plain gradient regarding the parameters is derived

$$\nabla_\theta J(\theta) = \nabla_\theta \int f(\mathbf{z})\pi(\mathbf{z}|\theta)d\mathbf{z} = \mathbb{E}_\theta [f(\mathbf{z})\nabla_\theta \log \pi(\mathbf{z}|\theta)]. \quad (3.2)$$

According to Ref. [122], plain gradient ascent tends to be unstable with either oscillation or premature convergence. Natural gradient is thus put forward by multiplying the plain gradient with the inverse of Fisher information matrix $\mathbf{F} = \mathbb{E}_\theta [\nabla_\theta \log \pi(\mathbf{z}|\theta)\nabla_\theta \log \pi(\mathbf{z}|\theta)^T]$,

$$\tilde{\nabla}_\theta J(\theta) = \mathbf{F}^{-1}\nabla_\theta J(\theta). \quad (3.3)$$

Note the natural gradient can be computed in a discrete manner by estimation from the search distribution \mathbf{z}_k ($k = 1, \dots, N_p$). The search distribution is updated by natural gradient ascent

$$\theta \leftarrow \theta + \eta \tilde{\nabla}_\theta J(\theta) = \theta + \eta \mathbf{F}^{-1}\nabla_\theta J(\theta). \quad (3.4)$$

where η is the learning rate of parameter θ . The above derivations show the general principle of NES that employs global search and natural gradient ascent. With regards to its variant SNES, several crucial modifications have been adopted to improve the performance and robustness [122]. These include fitness shaping that replaces the rank-based fitness with a set of ordered utility values w_k ($k = 1, \dots, N_p$), the introduction of a local "natural" coordinate system

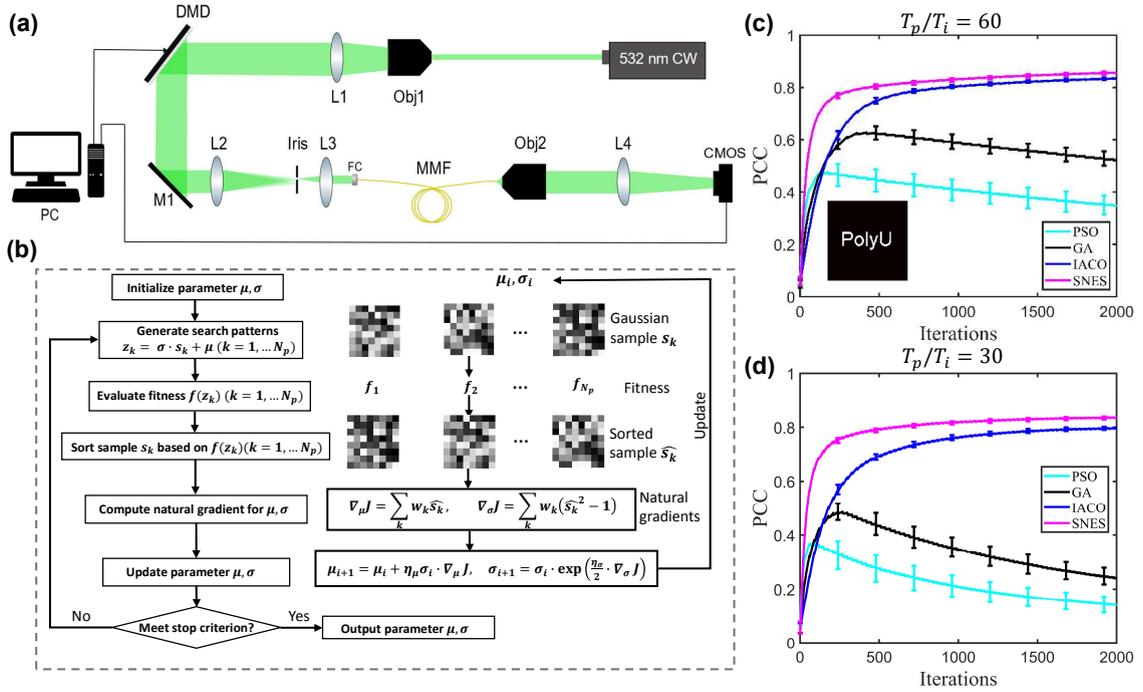


Figure 3.1: Illustration and simulated results of pattern projection through an MMF with iterative wavefront optimization. (a) Schematic of focusing light into a desired pattern at the distal end of MMF. (b) Diagram of the principle of SNES. (c) Simulated progression of PCC as a function of iteration number for different optimization algorithms for the projection of "PolyU" pattern, under different relative persistence times.

that simplifies the Fisher matrix to an identity matrix and so on. Figure 3.1 (a-b) illustrates the use of SNES for pattern projection by updating the search parameters μ and σ .

3.2.2 Principle of AlterProjTM

We get inspiration from the problem of phase retrieval solved by using the Gerchberg-Saxton (GS) algorithm [124, 125]. The basic idea of GS algorithm is to retrieve the phase of an object from the intensity measurement of its diffraction pattern (*i.e.*, the Fourier transform of the object) via alternating projection between the object plane and the Fourier plane. Here, we propose to optimize the phase mask for the suppression of light intensity at a target output region by alternating projection between the phase and speckle field of an MMF, with the TM as the propagation operator.

Figure 3.2 (a-b) depicts the principle of alternating projection-based phase optimization for glare suppression. The algorithm starts with an initial speckle pattern produced under zero-phase modulation. At the k -th iteration, the inverse TM transforms the estimated speckle field $\mathbf{I}_k(u, v)$ into the phase field $\mathbf{A}_k(x, y)$,

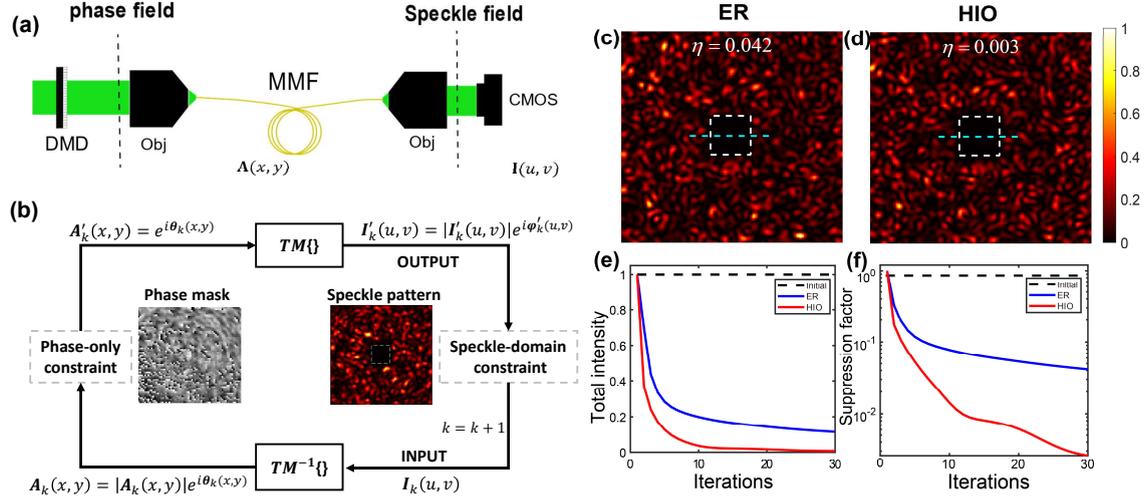


Figure 3.2: Principle and simulated results of speckle suppression at a target region of MMF output plane. (a) Illustration of alternating projection between the phase and speckle field of an MMF. (b) Schematic of the proposed AlterProjTM for phase optimization. Simulated results of glare suppression in a target region of 30×30 , using AlterProjTM with (c) ER and (d) HIO constraints. (e) Normalized curves of the total intensity in the target regions. (f) Curves of suppression factor as a function of iteration number.

$$\mathbf{A}_k(x, y) = |\mathbf{A}_k(x, y)| e^{i\theta_k(x, y)} = TM^{-1} \{ \mathbf{I}_k(u, v) \}. \quad (3.5)$$

where $\theta_k(x, y)$ is the phase of the resultant input wavefront. Here, the Tikhonov inversion [107] is adopted, which helps mitigate the measurement noise of TM and allows more accurate modelling of the back-propagation of light. The phase-only constraint sets the amplitudes of the input wavefront all to be one, such that

$$\mathbf{A}'_k(x, y) = e^{i\theta_k(x, y)}. \quad (3.6)$$

The new phase field $\mathbf{A}'_k(x, y)$ then propagates back to the speckle domain $\mathbf{I}'_k(u, v)$,

$$\mathbf{I}'_k(u, v) = |\mathbf{I}'_k(u, v)| e^{i\phi'_k(u, v)} = TM \{ \mathbf{A}'_k(x, y) \}. \quad (3.7)$$

where $\phi'_k(u, v)$ is the phase of the resultant speckle field. Similarly, a constraint is applied on the speckle domain to force an intensity reduction in the target region. The updated speckle field $\mathbf{I}_{k+1}(u, v)$ continues the above processes iteratively, with the goal of estimating a solution that conforms to both constraints simultaneously.

One natural choice of the speckle-domain constraint is setting the values of the points within the target region (represented by T) to be 0, given by

$$\mathbf{I}_{k+1}(u, v) = \begin{cases} 0, & (u, v) \in T \\ \mathbf{I}'_k(u, v), & (u, v) \notin T \end{cases} \quad (3.8)$$

This constraint is consistent with the error-reduction (ER) algorithm [124, 126], a variant of GS algorithm. However, it has been observed that the ER algorithm tends to reach a plateau after the first few iterations. A modification is the hybrid input-output (HIO) algorithm [125, 126]. The alternating projection streaming can be regarded as an input-output system, with the produced speckle field as the output and the updated one as the system input for the next iteration, as illustrated in Figure 3.2. For the pixels within the target region, HIO applies negative feedback upon the previous input $\mathbf{I}_k(u, v)$ using the iteration output $\mathbf{I}'_k(u, v)$ to generate the next input, such as

$$\mathbf{I}_{k+1}(u, v) = \begin{cases} \mathbf{I}_k(u, v) - \gamma \cdot \mathbf{I}'_k(u, v), & (u, v) \in T \\ \mathbf{I}'_k(u, v), & (u, v) \notin T \end{cases}, \quad (3.9)$$

where γ is the feedback coefficient. The output in the target region is driven to a value of zero due to a decrease in the input. HIO is supposed to have faster convergence as it could avoid a stagnation problem occurred in the ER algorithm.

The convergence property of the proposed alternating projection method for glare suppression is proved below. The mean absolute error of light intensities in the target area for k -th iteration is described by

$$E_{I,k} = \frac{1}{M_T^2} \sum_{(u,v) \in T} |I_{k+1}(u, v) - I'_k(u, v)| \quad (3.10)$$

where M_T denotes the side length of the square target region. In the phase field, both A'_k and A'_{k+1} satisfy the phase-only constraint by definition. According to the principle of alternating projection, at any point (x, y) , A'_{k+1} is the nearest value to A_{k+1} . Hence, in all points (x, y) , we have

$$\begin{aligned} |TM \{A_{k+1}(x, y) - A'_{k+1}(x, y)\}| &\leq |TM \{A_{k+1}(x, y) - A'_k(x, y)\}| \\ &= |I_{k+1}(u, v) - I'_k(u, v)|. \end{aligned} \quad (3.11)$$

Therefore, with Eq. (3.10), and Eq. (3.11), we have

$$\|TM\{A_{k+1} - A'_{k+1}\}\|_1 \leq M_T^2 \cdot E_{I,k}, \quad (3.12)$$

where $\|\cdot\|_1$ denotes the 1-norm for vector. In the speckle field, both I_{k+1} and I_{k+2} satisfy the speckle-domain constraint by definition. Similarly, at any point (u, v) , I_{k+2} is the nearest value to I'_{k+1} . Hence, in all points (u, v) , we have

$$\begin{aligned} |I_{k+2}(u, v) - I'_{k+1}(u, v)| &\leq |I_{k+1}(u, v) - I'_{k+1}(u, v)| \\ &= |TM\{A_{k+1}(x, y) - A'_{k+1}(x, y)\}|. \end{aligned} \quad (3.13)$$

Therefore, with Eq. (3.10), and Eq. (3.13), we have

$$M_T^2 \cdot E_{I,k+1} \leq \|TM\{A_{k+1} - A'_{k+1}\}\|_1. \quad (3.14)$$

Combining Eq. (3.12), and Eq. (3.14), eventually we have

$$E_{I,k+1} \leq E_{I,k}. \quad (3.15)$$

This means the absolute errors of light intensities in the target region declines iteratively, which shows a good convergence property of our alternating projection method for glare suppression.

3.3 Numerical simulations

3.3.1 Pattern projection under noise and perturbation

We first numerically test the performance of SNES for pattern projection through a complex medium under noisy and disturbed conditions to mimic the real experiment. Considering that there are shot noise, dark electricity noise of camera and read-out noise *etc.*, we apply a multiplicative noise to the output light intensity $\mathbf{I} \in \mathbb{R}_+^M$ during the simulation, described by

$$\mathbf{I}_{\text{noise}} = \mathbf{I} + \alpha \mathbf{I} \cdot \boldsymbol{\varepsilon}. \quad (3.16)$$

Here, ε is a vector with the same size as \mathbf{I} , which is randomly drawn from a uniform distribution $U(-0.5, 0.5)$, and $\alpha > 0$ denotes the noise level. In real conditions, the complex medium may be unstable, and its TM may fluctuate over time. According to the model in Ref. [12], a small perturbation ξ drawn from a complex Gaussian distribution $\mathcal{N}(0, \delta)$ can be added to each of the TM elements t_{mn} for simulation,

$$t_{mn} \rightarrow \frac{1}{\sqrt{1 + \delta^2}} (t_{mn} + \xi) \quad (3.17)$$

Note that δ is relative to the standard deviation of the TM, which denotes the perturbation level. If we let T_p be the persistence time of the complex medium when its TM remains unchanged and T_i be the duration of one iteration, the perturbation level can be predicted by $\delta = 1/\sqrt{T_p/T_i}$ [29], where T_p/T_i is denoted as relative persistence time.

For pattern projection through complex media, the correlation between the input and output intensity distributions, *i.e.*, PCC, is an important metric to evaluate pattern fidelity. Also, the focusing contrast, that is, the ratio between the mean light intensities in the target region and that of the non-target region, is a key metric of pattern projection. previous research usually adopted PCC as the fitness function to evaluate the performance [110, 113, 114]. We found that using PCC as the feedback is not ideal for increasing the focusing contrast. In this work, we propose to replace it with VecCos as the fitness function, given by

$$VecCos = \frac{\sum_i^M (\mathbf{I}_i \cdot \mathbf{I}_i^t)}{\sqrt{\sum_i^M (\mathbf{I}_i)^2} \sqrt{\sum_i^M (\mathbf{I}_i^t)^2}} \quad (3.18)$$

In simulation, we controlled the noise level to be 60% and set the relative persistence time of the medium, T_p/T_i , to be 60 and 30, which reasonably mimicked empirical observations. We compared the performance of SNES with other algorithms in projecting a binary pattern of "PolyU". The detailed parameters are reasonably set for fair comparison, with details referred to [15]. Each of the algorithms was repeated 30 times for averaging and a new TM was simulated each time, and they all used PCC as the fitness function for guiding the wavefront optimization. Figure 3.1c shows the progression curves of PCC between the simulated outputs and the target pattern for various algorithms, including GA, particle swarm optimization (PSO), IACO [114] and SNES. The common parameters for all the algorithms are controlled to be the same: the population size of each generation is 18; the number of modulation units is 64×64 ; and the

number of iterations is 2000 to ensure convergence. All the algorithms are tuned to their best states for a fair comparison. It can be seen that SNES evolves fastest and reaches the highest PCC (larger than 0.8) under both conditions, and it is least affected by the variations in perturbation levels. For IACO, it evolves slowest initially, but exhibits strong immunity to noise and perturbation and grows gradually, achieving PCCs that are slightly smaller than those of SNES at the end. The other two algorithms, GA and PSO, however, are more susceptible to the measurement noise and furthermore the perturbation to the medium. These results prove that SNES has superior features such as fast convergence and robust immunity to noise and perturbation.

3.3.2 Comparison and analysis of suppression performance

We also investigated AlterProjTM for glare suppression via numerical simulation. As a proof-of-principle study, AlterProjTM optimized with the two speckle-domain constraints, ER and HIO, was explored for the comparison of suppression performance. Herein, the number of modulation units N was set to be 64×64 , and the feedback coefficient γ of HIO was empirically set as 0.8 to ensure quick and stable convergence. To evaluate the suppression effect, a suppression factor η_s is defined

$$\eta_s = \mu(|\mathbf{I}_i|)_{|i \in T} / \mu(|\mathbf{I}_i|)_{|i \notin T}, \quad (3.19)$$

Where $\mu(\cdot)$ denotes the mean operation, T represents the target region, and \mathbf{I}_i is the speckle field at the position index i . Figure 3.2(c-f) shows the simulated glare suppression results with an iteration number of 30 for achieving an acceptable suppression result at a short time cost. The ROI was a square of 30×30 pixels in the middle of the speckle field, marked with white dashed boxes. The light intensities within the ROI are suppressed for those optimized with both ER (Figure 3.2, $\eta_s = 0.042$) and HIO (Figure 3.2, $\eta_s = 0.003$). The normalized total intensity of the ROI declines progressively during the first 30 iterations for both ER and HIO, and HIO can already converge to approximate zero, as seen in Figure 3.2e. The progression curves of suppression factor as a function of iteration number are shown in Figure 3.2f, which also reveals that HIO declines faster and achieves superior suppression effect than ER.

Numerical comparisons were also made between the AlterProjTM and the state-of-the-art GA to test their performances of glare suppression in square regions of various side length L . Still,

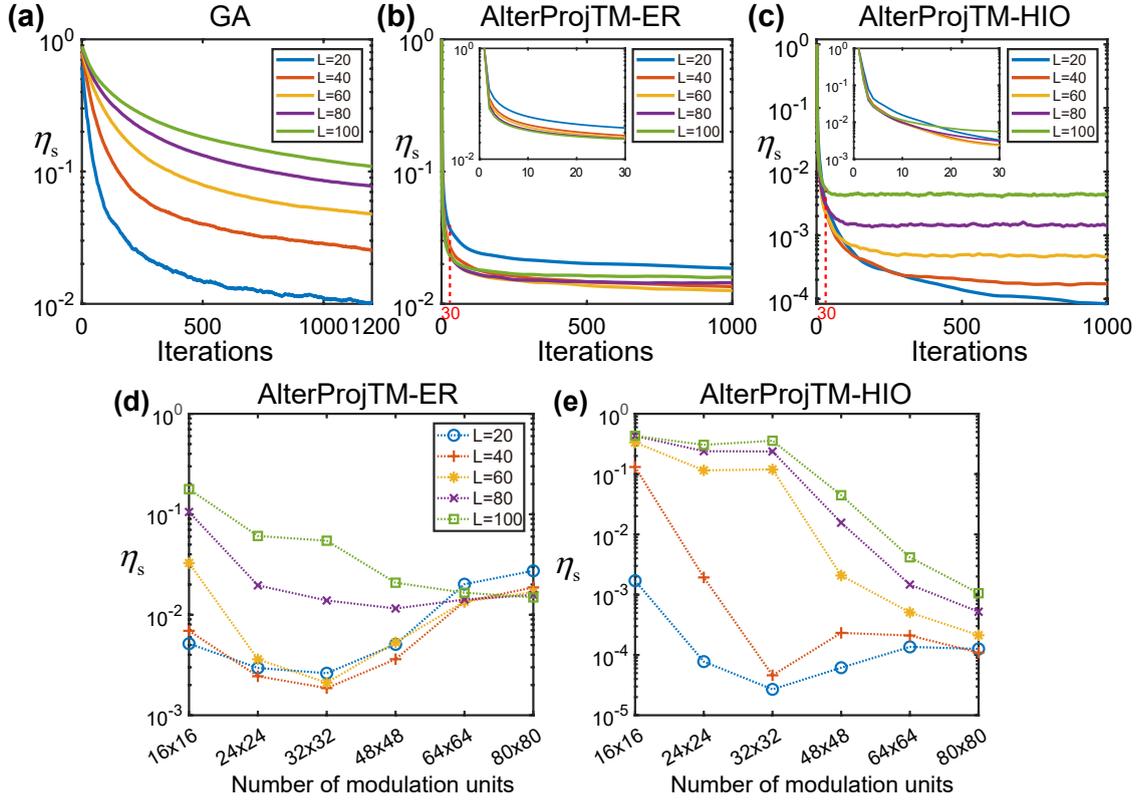


Figure 3.3: Theoretical comparison and performance analysis of the proposed AlterProjTM for glare suppression. Curves of suppression factor η_s as a function of iteration number in square ROIs of different side lengths, when using (a) GA, and AlterProjTM with (b) ER and (c) HIO constraints, respectively. Note that the subplots inside (b) and (c) show the zoomed-in results within the first 30 iterations. (d) Suppression factor η_s as a function of the modulation unit factor η_s as a function of the modulation unit number for square ROIs of different side lengths, using AletrProjTM with (d) ER and (f) HIO constraint after 1000 iterations.

for each L the result of each algorithm was averaged over 30 repetitions. The parameters of AlterProjTM remained unchanged, and those of GA were set as follows: the initial and final mutation rates were 0.01 and 0.001, respectively; the decay factor was 250; the population size was 30; and the iteration number was 1200. Figure 3.3 (a-c) shows that AlterProjTM with 30 iterations can approach or surpass GA with 1200 iterations in terms of suppression factor. Specifically, it took an average of 12.03 s for GA, while only 0.54 s for ER and 0.46 s for HIO, when running on GPU in a regular PC. Figure 3.3a reveals the suppression effect of GA enhances (η declined from ~ 0.1 to ~ 0.01) as the side length of the ROI decreases from 100 to 20, which is reasonable as the number of phase control segments remains unchanged. For AlterProjTM with ER constraint (Figure 3.3b), it converges prematurely after several hundred iterations for different ROIs, with relatively poor performance ($\eta_s > 0.01$) in intensity suppression. Also, it seems that there is no obvious rule between the ROI sizes and the achieved suppression factors. By contrast, when using the HIO constraint (Figure 3.3c), the suppression

ability of AlterProj can be fully exerted, achieving hierarchical suppression results for ROIs of increasing sizes that are all considerably superior to those of GA and ER. In particular, the optimization can achieve a suppression factor at a level of 10^{-3} theoretically with merely 30 iterations, with no obvious differences among different ROIs.

Further analyses on the suppression performance of AlterProjTM in ROIs of various sizes, under a series of numbers of modulation units are briefly given below. For AlterProj optimized with ER constraint, simply setting the signal values within ROI to be 0 can easily cause the algorithm to be stuck in local minima and converge prematurely. Thus, the suppression ability of the method is not fully exerted by using the constraint of ER. According to Figure 3.3d, for a certain ROI, the suppression result is not always enhanced with increasing N especially for ROI with small sizes, mainly due to the premature convergence of ER. When it comes to HIO constraint, due to its suppression mechanism of negative feedback, there is no such issue of premature convergence, so the suppression ability of our method is fully exerted. As shown in Figure 3.3e, under a fixed N , the suppression factor usually gets worse as the ROI size increases, showing a good inversely proportional relationship. That said, the "over-saturation" phenomenon still exists to some extent. For example, $N = 32 \times 32$ is optimal for ROIs with smaller sizes, including the cases of $L = 20$, and 40. The reason for this may be partly attributed to the choice of feedback coefficient γ , as the optimal value of γ varies slightly with different ROI sizes and different N . $\gamma = 0.8$ is adopted herein as it is suitable for most cases, while it may not be optimal for intensity suppression in small ROIs when the modulation capability is relatively strong.

3.4 Experimental results

The experimental configuration for controlling light delivery through an MMF is illustrated in Figure 3.4. The laser from a 532 nm continuous-wave laser (EXLSR-532-300-CDRH, Spectra Physics, USA) was first divided into two beams by a polarizing beam splitter (PBS) after passing through a half-wave ($\lambda/2$) plate. One beam was used as the reference beam, which was expanded by a 4f lens system (L5 and L6), with its polarization state being adjusted by a $\lambda/2$ plate. The other beam was expanded by a 40 \times objective lens (Obj1) and collimated by a convex lens (L1) subsequently. After turning to be circularly polarized light by a quarter-wave plate ($\lambda/4$), the beam was reflected by a DMD (V-7001, VIALUX, Germany) and relayed into

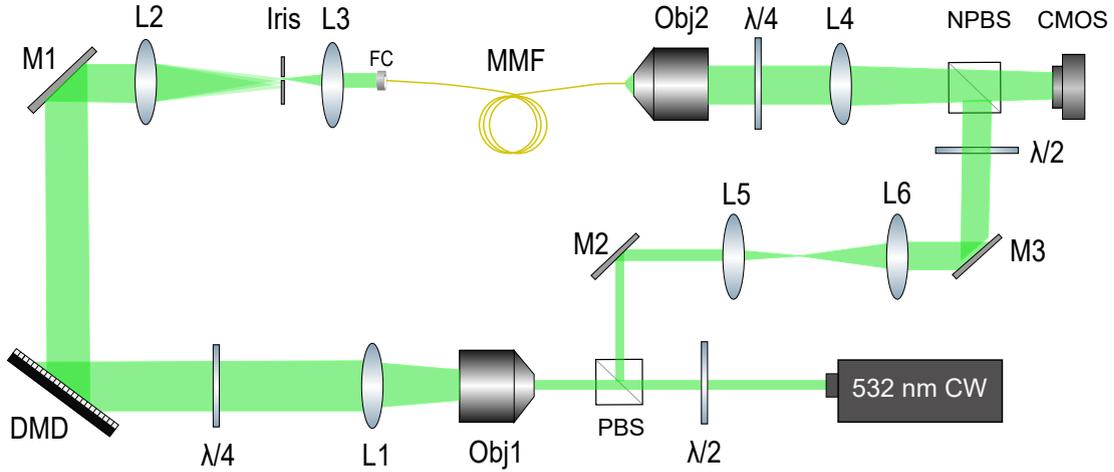


Figure 3.4: Experimental setup for glare suppression through an MMF. The complex TM of the MMF is measured via off-axis holography, and a DMD is used for phase-only modulation. CMOS, complementary metal-oxide-semiconductor camera; DMD, digital mirror device; FC, fiber collimator; L, lens; M, mirror; MMF, multimode fiber; NPBS, non-polarizing beam splitter; Obj, objective; PBS, polarizing beam splitter; $\lambda/2$, half-wave plate; $\lambda/4$, quarter-wave plate.

another 4f system (L2, iris, and L3). After spectral filtering, the beam encoded with the desired phase distribution [127, 128] was coupled into an MMF (0.22 NA, 105 μm core, SUH105, Xinrui, China). The beam after propagating through the MMF was magnified by a 40 \times objective (Obj2) and collimated by a tube lens L4. A $\lambda/4$ plate was used to recover the linearly polarized components of the disordered beam. The speckle and the reference light paths were combined by a non-polarized beam splitter for interference before being captured by a CMOS camera (BFS-U3-04S2M, FLIR, USA). Note a shutter was adopted in the path of external reference beam, which controlled whether the reference beam is needed (for TM measurement) or not (for iterative WFS).

3.4.1 Long-distance pattern projection through unstable MMF

We have conducted the experiment of long-distance pattern projection through an MMF, which was 15-meter-long and loosely placed in an optical table. As such, the MMF was considered as a dynamically changing medium since weak air flow and platform vibrations could cause a drift of position or deformation to it. The speckle decorrelation time (the time it took for the correlation between the captured speckle patterns with the initial one first dropped below $1/e$, using the same input wavefront) was characterized to be $\sim 35\text{s}$, showing a quick decorrelation status. In the experiment, we have compared SNES with other wavefront control algorithms,

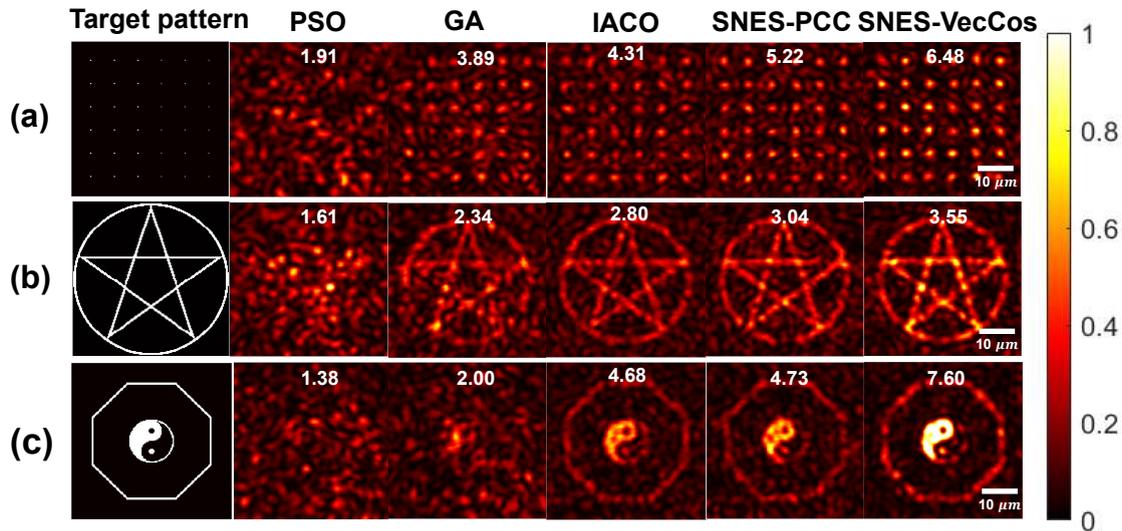


Figure 3.5: Results of pattern projection through a 15-meter-long unfixed MMF, using different optimization methods for comparison. Projection results for the target pattern of (a) 6×6 focus array, (b) pentagram, and (c) simplified Bagua are shown on the first, second, and third row, respectively, with the corresponding contrast value noted in each realization. Note that for each realization, the intensities of experimentally acquired images are individually normalized and have the same color bar for comparison of contrast. The scale bar shown in the last images of (a-c) is $10 \mu\text{m}$.

using PCC as the fitness function. Besides that, SNES was also tested by being optimized with the proposed VecCos, namely SNES-VecCos, which was expected to enhance the performance.

Figure 3.5 shows the experimental results of pattern projection through the long unstable MMF for different samples, including a 6×6 focus array, a pentagram, and a simplified Bagua pattern. In the experiment, wavefront optimization of various methods all ran for 1,000 iterations in the meanwhile for the above output results. Qualitatively, the performances of different methods are in good agreement with those observed in simulations (Figure 3.1 c-d). The pattern fidelity achieved by IACO and SNES were comparable and considerably better than those obtained by GA and PSO. It is also obvious that SNES-VecCos achieved the highest contrast in all cases. Particularly for the Bagua pattern, the contrast yielded by SNES-VecCos was $\sim 60\%$ higher than those achieved by SNES-PCC and IACO. Note that since the number of output modes to be controlled varied case by case, the optimal contrast of projected patterns that can be attained was naturally at different levels. Nevertheless, the high-quality results of pattern projection through the long unstable MMF confirmed the superior performance of the proposed SNES-VecCos method through experiments.

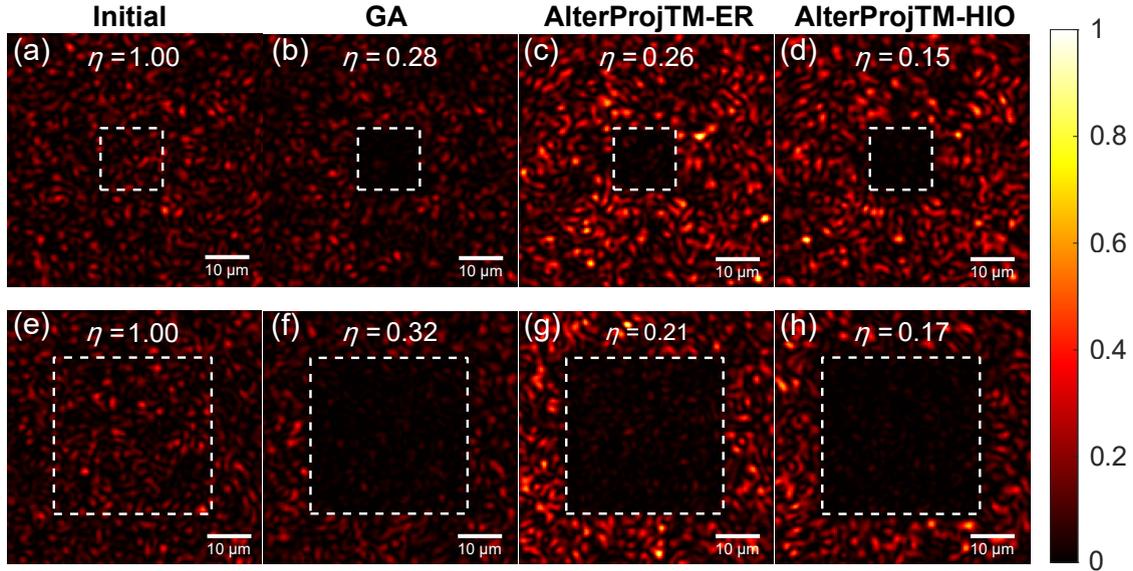


Figure 3.6: Normalized speckle patterns showing glare suppression in ROIs of various sizes at the distal end of the MMF, using different algorithms for comparison. The first row and second row correspond to square regions of 40×40 pixels and 100×100 pixels respectively, with (a, e) for the initial speckle pattern under zero phase modulation, (b, f) for those optimized with GA after 1200 iterations, (c, g) for those optimized with ER after 30 iterations, and (d, h) for those optimized with HIO after 30 iterations. The scale bar shown in (a-h) is $10 \mu\text{m}$.

3.4.2 Arbitrary speckle suppression through MMF

In the experiment of glare suppression, the 15 m MMF was replaced with a 1 m one and was fixed in the platform. Similar to the previous simulations, GA and the proposed AlterProjTM optimized with ER and HIO were all conducted for performance comparison, with the same parameter settings except that GA ran for 1,200 iterations while AlterProjTM merely ran for 30 iterations. The TM of the MMF was measured via off-axis holography first before being used for phase optimization with different algorithms that were implemented by GPU. It should be noted that no hardware feedback was involved, and the operations for GA and AlterProjTM were both based on the same TM for fair comparison. One obvious advantage of our method is that multiple phase masks that correspond to different ROIs can be optimized in parallel. For example, it only took ~ 2.3 s to optimize 120 frames of phase masks simultaneously, equivalent to ~ 22.5 ms/frame. In the experiment, a maximum frame rate of ~ 50 Hz was achieved for real-time glare suppression through MMF.

Figure 3.6 shows the normalized speckle images before and after phase optimization by different methods, for ROIs of 40×40 pixels (first row) and 100×100 pixels (second row) respectively. Compared to the simulated results in Figure 3.3, where GA achieved $\eta_s = 0.025$ after 1200 iterations and ER achieved $\eta_s = 0.028$ after 30 iterations, the suppression factors

realized in experiment were almost one order-of-magnitude worse in both cases. For HIO, the suppression effect ($\eta_s = 0.15$) was undoubtedly better than GA and ER, although still almost two order-of-magnitude worse than the simulated one ($\eta_s = 0.0024$). When it comes to the case of 100×100 ROI that contains ~ 625 speckles, the performance of glare suppression was slightly worse for GA ($\eta_s = 0.32$) and HIO ($\eta_s = 0.17$), but slightly better for ER ($\eta_s = 0.21$). In summary, the obtained experimental results agreed reasonably with the simulated ones in Figure 3.3. Although the case of ER seems to be counterintuitive that the 100×100 region yielded a slightly better suppression factor than the 40×40 region. It could be understandable as ER was prone to premature convergence and the optimization was not yet fully converged with 30 iterations. Apart from the results shown in Figure 3.6, fast and effective glare suppression in arbitrary regions of different shapes and sizes at the distal end of MMF was demonstrated experimentally (Media 1) using AlterProjTM with HIO constraint. The technique could be promising for speckle optical tweezer [118] or super-resolution imaging [119] inside deep tissues through an MMF, where a customizable speckle field is required.

3.5 Discussion

The first part of our discussion revolved around the implementation and comparison of various wavefront optimization methods for projecting patterns through a long and unstable MMF. Among these, SNES optimized with VecCos emerged as the most efficient, demonstrating rapid evolution and strong robustness. This superiority is attributed to its principle of parameter optimization via natural gradient ascent, which, unlike other algorithms that more or less rely on random variations, ensures robustness against medium disturbance. The scheme of SNES-VecCos significantly enhances the contrast of projected patterns, making it a promising tool for biomedical applications requiring high-precision optical delivery such as phototherapy or optogenetics. However, the time consumption for wavefront optimization poses a limitation, suggesting the need for further system improvements to accommodate real-time applications in dynamic biological tissues. For instance, faster feedback on pattern formation, more efficient data transfer, processing, and control through a customized field-programmable gate array framework may be employed.

In the second part, we present an alternating projection-based phase optimization method utilizing the TM for effective glare suppression at arbitrary regions through MMF. In practice, this

method requires only a few iterations for phase optimization, offering a quick and parallel optimization process. It extends the capability of WFS to achieve arbitrary intensity suppression and the formation of focal patterns, by modifying the speckle-domain constraint. However, the study also has some limitations such as the complexity of measuring a complex TM and discrepancies between experimental and simulated suppression results. The reasons may include: the inaccuracy of the TM measurement for MMF which could be affected by the quality of the reference beam and the noise during image acquisition; the inaccuracy of phase modulation through a DMD and 4f system, the small perturbation to MMF and so on.

Notably, it seems the proposed AlterProjTM showed an insensitivity to the size of ROI to be suppressed, which can be observed from Figure 3.6 and Media 1. Several reasons may account for that. The first is only 30 iterations were adopted in the experiment for phase optimization with both ER and HIO constraints, at which the optimizations did not fully converge, so the differences among different ROIs were not obvious. Then, for ER, it suffers from premature convergence accompanying with the problem of "oversaturation" which means under a certain input size, it is not always the case that a larger ROI would have a worse suppression effect. Finally, the weakened suppression performance of AlterProj in the experiment made it less apparent for the distinctions among different ROIs. It is emphasized that under appropriate speckle-domain constraint (HIO) and sufficient convergence, AlterProjTM has confirmed an inversely proportional relationship between the ROI size (*i.e.*, number of speckles included) and the suppression effect achieved.

3.6 Conclusion

In conclusion, this chapter has thoroughly studied controlling light delivery through an MMF based on wavefront optimization, including focusing light and suppressing light in an arbitrary patterned region at the output of MMF. For the first part of pattern projection, SNES optimized with VecCos is proposed in this study with the best performance to control light focusing into an arbitrary pattern in a noisy and perturbed environment. As such, a pattern can be effectively generated at the distal end of the MMF within 1000 iterations with high fidelity and focusing contrast. And for the second part of suppressing scattered light, a TM-based phase optimization method is proposed for fast and effective glare suppression through MMF, with speckle-domain constraint of both ER and HIO being compared. The simulations and experiments collectively indicate that AlterProjTM with HIO constraint is superior in terms of fast convergence and

strong suppression. For example, a suppression factor of 0.17 was realized by HIO in a large target region containing ~ 625 speckles at the distal end of an MMF. The dual capability of precise light delivery through MMF may promise innovative deep-tissue optical applications, including optogenetic, phototherapy and imaging or sensing with speckles, where controlled light delivery can significantly impact outcomes.

Chapter 4

Optimum Transmission Matrix Retrieval of a Multimode Fiber by Nonconvex Optimization

*This chapter is reproduced with some adaptations according to two published journal papers: Shengfu Cheng, Xuyu Zhang, Tianting Zhong, Huanhao Li, Haoran Li, Lei Gong, Honglin Liu, and Puxiang Lai, "Nonconvex optimization for optimum retrieval of the transmission matrix of a multimode fiber", *Advanced Photonics Nexus* 2(6):066005 (2023).*

Continuing Part II for studying the light control of MMF, we delve into the reconstruction of the TM of MMF in a reference-less setup, which waives the need of an external reference beam and also avoids the "dark-spot" problem suffered by internal reference method. Although efforts to directly retrieve a complex-valued TM from intensity measurements using phase retrieval algorithms have been taken, they are still faced with limitations such as slow or suboptimum recovery, particularly in noisy conditions. In Chapter 3, we propose a modified non-convex optimization approach to tackle the TM retrieval problem. Numerical evaluations demonstrated that our method achieves optimum focusing efficiency with reduced running time or sampling ratio. Comparative tests across different signal-to-noise levels confirmed its improved robustness. In the experiment, our algorithm's superior focusing performance was validated cohesively through single- and multi-spot focusing. Notably, at a sampling ratio of 8, it realized 93.6% efficiency compared to the gold standard holography method. Parallel operation and GPU acceleration enable our nonconvex approach to efficiently retrieve an 8685×1024

TM (sampling ratio=8) with 42.3 seconds averagely on a regular computer. Leveraging the recovered TM, we also achieved high-fidelity image transmission through an MMF. With optimum efficiency and fast execution for TM retrieval, our method will facilitate deep-tissue applications such as optical imaging, manipulation, and treatment.

4.1 Introduction

Different from ordinary ballistic optics, light propagation in complex media is highly disordered [129, 130] due to the multiple scattering occurring in media like biological tissues or mode dispersion in an MMF. Finding an order out of such disorders has been long pursued. Over the past decade, enormous progresses have been made via wavefront shaping [8, 15, 131–134] and especially the TM method [16, 23, 106, 127, 135] in controlling light to focus and image through complex media. The TM of a disordered medium describes the complex output responses for an arbitrary point-source input, which is regarded as the transfer function of the medium under the shift-invariance assumption [135]. The measurement of TM offers a versatile tool to control light delivery in spite of scattering [106, 133], as well as recovering object information from acquired speckle patterns [136, 137]. The TM method has spurred a wide range of MMF-based applications, such as focusing [18, 108], glare suppression [121, 138], endoscopic imaging [139–141], manipulation [142], optogenetics [143], and communication [136, 144, 145].

TM measurement of a scattering medium was first introduced by Popoff *et al.* [16, 106] using coaxial holography with internal reference. Since then, various forms of TM measurement have been developed. Typically, the TM is measured by recording the complex output fields under a sequence of input modulations. The modulation basis is usually orthogonal, which can be of diverse forms, including Hadamard matrix [16, 106], DFT matrix [146], point source [20, 147], and random phase [23]. Regardless of the form, the measured TM relates all input modes to each output mode by linear superposition. Depending on the type of input modulation and output measurement, the TM could be complex-valued [16, 20], real-valued [148, 149] or even binary [150]. Among them, complex TM is used most as it supports both amplitude and phase modulation of light, which, however, usually entails holographic setup. Off-axis holography based on either phase-shifting [151] or spatial filtering [152] can acquire the complex TM accurately. Nevertheless, effective off-axis interferometry with an additional reference beam and high system stability it demands could be unpracticable in some scenarios. As an example, the coherence length of pulse laser could be too short to be used for interferometry-based TM

measurement. With coaxial holography, the above issues might be alleviated, but the dark spot problem with the measured TM caused by speckle reference field [22] is still unsatisfying.

Recent efforts have sought to accurately retrieve the complex TM from intensity-only measurements by using advanced phase retrieval algorithms [23, 24, 146, 153–158], which started with a Bayesian inference approach (*i.e.*, prVBEM) [23]. This was followed by prSAMP [153] and prVAMP [22]. Although robust to noise, a prior knowledge of noise statistics is a must for these approaches. Semidefinite programming (SDP) that uses convex relation has also been introduced for solving the TM retrieval problem [155], but it usually requires $N \ln N$ (N is the input size) measurements and tends to be computationally inefficient. Additionally, works based on extended Kalman filter (EKM) [156] or generalized Gerchberg-Saxton (GGS) algorithm [24] claim retrieving TM with $4N$ measurements could be enough. That said, EKM is computationally burdened and hard for parallelization. GGS is efficient in computation, but its performance is still suboptimum in real practice. Most recently, the area also sees the birth of a smoothed Gerchberg-Saxton algorithm [158] and a nonlinear optimization method [146] for TM retrieval.

To overcome the aforementioned limitations, in this study a state-of-the-art nonconvex optimization approach is adopted and modified for TM retrieval with optimum performance. Compared with existing TM retrieval algorithms, the proposed nonconvex method can achieve optimal efficiency numerically with less running time or sampling rate. In the experiment, by focus-scanning across the FOV of an MMF with the acquired TM, the performance of the proposed method is validated to approach the golden standard, *i.e.*, off-axis holography with a sampling rate of 8. Moreover, with the assistance of parallel operation and GPU acceleration, multiple rows of TM can be recovered rapidly. Our method for TM retrieval is featured with optimum efficiency and fast implementation in a reference-less and robust setting, which may gain special attention for many deep-tissue imaging and focusing applications with the usage of MMF.

4.2 Methods

4.2.1 Formulation of the TM retrieval problem

The theoretical model of retrieving a TM under a sequence of input modulations is formulated as follows. Suppose the number of discrete modulation units (*i.e.*, input size) and speckle field pixels (*i.e.*, output size) is N and M , respectively. Given P calibration patterns such that the probe matrix $\mathbf{X} \in \mathbb{C}^{N \times P}$ and the amplitude measurements $\mathbf{Y} \in \mathbb{R}_+^{M \times P}$, the TM $\mathbf{A} \in \mathbb{C}^{M \times N}$ that needs to be estimated shall follow

$$\mathbf{Y} = |\mathbf{A}\mathbf{X}|, \quad (4.1)$$

where $|\cdot|$ takes the absolute value for the elements inside. By taking the conjugate transpose of both sides of Eq. (4.1), we have

$$\mathbf{Y}^H = |\mathbf{X}^H \mathbf{A}^H|, \quad (4.2)$$

where $(\cdot)^H$ is the element-wise conjugate transpose operator. Column-wisely, $\mathbf{Y}^H = [\mathbf{y}_1, \dots, \mathbf{y}_i, \dots, \mathbf{y}_M]$, where $\mathbf{y}_i \in \mathbb{R}_+^P$ constitutes the measurements associated with the i^{th} output mode, and $\mathbf{A}^H = [\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_M]$ where $\mathbf{a}_i \in \mathbb{C}^N$ denotes the conjugate transpose of the i^{th} row of TM, $i = 1, \dots, M$. In this case, the TM retrieval problem is decomposed into M independent sub-problems given as

$$\mathbf{y}_i = |\mathbf{X}^H \mathbf{a}_i|, i = 1, \dots, M. \quad (4.3)$$

Due to the operation of taking absolute values in Eq. (4.3), the above problem of estimating one row of TM is nonlinear and falls in the category of phase retrieval.

Phase retrieval problem has been studied intensively in mathematics as it is commonly encountered in practice, with representative algorithms including alternating projection [159] (*e.g.*, Gerchberg-Saxton and Fineup), SDP [160] (*e.g.*, PhaseLift, PhaseMax, PhaseCut), approximate message passing (*e.g.*, GAMP [161], VAMP [162]), and nonconvex optimization [163–167] *etc.* Among these methods, nonconvex approaches are proven to be superior and have been

developed rapidly in the past years. There are mainly two categories of nonconvex approaches, the intensity-flow [164, 168] and amplitude-flow models [165–167], with the latter being better in both empirical success rate and convergence property. In particular, the amplitude-flow models have been proven to converge linearly to the true solution under $O(n)$ Gaussian measurements for a signal with dimension [167].

4.2.2 The modified RAF algorithm

Herein, the cutting-edge reweighted amplitude flow (RAF) algorithm [167] is adopted for the TM retrieval problem. Solving Eq. (4.3) can be recast as an optimization problem

$$\min_{\mathbf{a}_i \in \mathbb{C}^N} L(\mathbf{a}_i) = \|\|\mathbf{X}^H \mathbf{a}_i - \mathbf{y}_i\|_2\|^2, \quad (4.4)$$

where $\|\cdot\|_2$ denotes the L2 norm of a vector, and $L(\mathbf{a}_i)$ is an amplitude-based least square error (LSE) loss function. While most nonconvex algorithms contain two stages, *i.e.*, spectral initialization and gradient descent, RAF applies reweighting techniques in both stages that accelerates the signal recovery significantly. Considering Eq. (4.4), the signal, *i.e.*, one row of TM \mathbf{a} (the row index i is omitted for genericity) is first estimated with the weighted maximum correlation initialization. A subset of the row vectors in the probe matrix ($\mathbf{X}^H = [\mathbf{x}_1^H; \dots; \mathbf{x}_p^H]$) that correspond to the $|S|$ (subset $S \subset \{1, \dots, P\}$) largest entries in the measurements $\mathbf{y} = \{y_j\}_{1 \leq j \leq P}$ are selected, which are called direction vectors, as they are more correlated to the true signal. The signal can be estimated by maximizing its correlation to the direction vectors $\{\mathbf{x}_j^H | j \in S\}$ such that

$$\max_{\|\mathbf{a}\|=1} \frac{1}{|S|} \sum_{j \in S} |\langle \mathbf{x}_j^H, \mathbf{a} \rangle|^2 = \max_{\|\mathbf{a}\|=1} \mathbf{a}^H \left(\frac{1}{|S|} \sum_{j \in S} \mathbf{x}_j \mathbf{x}_j^H \right) \mathbf{a}. \quad (4.5)$$

By weighting more to the selected \mathbf{x}_j^H vectors corresponding to larger y_j values with the weights $y_j^\alpha, \forall j \in S$ (exponent $\alpha = 0.5$, by default), the solution $\tilde{\mathbf{a}}^0$ of Eq. (4.5) is the unit-norm principle eigenvector of the Hermitian matrix

$$\mathbf{H} = \frac{1}{|S|} \sum_{j \in S} y_j^\alpha \mathbf{x}_j \mathbf{x}_j^H = \frac{1}{|S|} \mathbf{X} \cdot \text{diag}(\tilde{y}_1^\alpha, \tilde{y}_2^\alpha, \dots, \tilde{y}_p^\alpha) \cdot \mathbf{X}^H, \quad (4.6)$$

where $\tilde{y}_j^\alpha = \begin{cases} y_j^\alpha, j \in S \\ 0, \text{otherwise} \end{cases}$. $\tilde{\mathbf{a}}^0$ is then scaled to obtain the signal initial guess $\mathbf{a}^0 = \sqrt{\sum_{j=1}^P y_j^2 / P} \cdot \tilde{\mathbf{a}}^0$.

The initialized signal \mathbf{a}^0 is further refined by reweighted "gradient-like" iterations. The gradient of the LSE loss in Eq. (4.4) with respect to \mathbf{a} is

$$\nabla L(\mathbf{a}) = \frac{1}{P} \cdot \mathbf{X} \left(\mathbf{X}^H \mathbf{a} - \mathbf{y} \circ \frac{\mathbf{X}^H \mathbf{a}}{|\mathbf{X}^H \mathbf{a}|} \right), \quad (4.7)$$

where \circ denotes element-wise multiplication. Let t be the iteration index, then the gradient descent is described by

$$\mathbf{a}^{t+1} = \mathbf{a}^t - \mu \cdot \nabla L(\mathbf{a}^t), \quad (4.8)$$

where μ is the step size, and $t = 0, 1, \dots$. One can reweight the gradients in Eq. (4.7) that have larger values of $|\mathbf{X}^H \mathbf{a}^t| \oslash \mathbf{y}$ (\oslash represents element-wise division), which are deemed more reliable in directing to the true signal. The adaptive weight vector is

$$\mathbf{w} = |\mathbf{X}^H \mathbf{a}^t| \oslash (|\mathbf{X}^H \mathbf{a}^t| + \beta \mathbf{y}), \quad (4.9)$$

where β is a pre-selected parameter. The above descriptions show the reweighted gradient flow for TM retrieval.

On the use of RAF for TM retrieval, we have tried to speed up the computation or convergence process. Efforts include employing an adaptive step size for gradient descent or other gradient acceleration strategies, such as nonlinear conjugate gradient (NCG) [169] and Limited memory-BFGS (L-BFGS) [170] methods. However, the improvements are limited compared to the original scheme of using a fixed step size and steepest descent (SD) method, with the results given in section 4.3.1.

Inspired by Ref. [24], herein we modify the original RAF algorithm, by dividing the gradient-descent process into two steps. In Step 1, the normalized measurement vector \mathbf{y} is replaced with its square \mathbf{y}^2 for gradient computation, which enlarges the gradient and functions as the artificial heat data. In Step 2, still the amplitude measurement \mathbf{y} is used. Step 1 contains the first 2/3 total iterations, which is set empirically (the rationale is referred to section 4.3.2). The modified algorithm is simple yet surprisingly effective, which is named RAF 2-1 and shown to reduce the measurement error to a much lower level than the original RAF (see section 4.3.3).

The pseudocode of retrieving one row of TM \mathbf{a} with RAF 2-1 is summarized in Algorithm 1. The time complexity for spectral initialization and gradient iteration in Algorithm 1 is $O(N|S|)$ (with the adaptation of power method or Lanczos algorithm) [167] and $O(NP)$ (usually $P \geq N$) respectively, so that the total time complexity is (at least) $O(N^2)$ for retrieving a TM row. Luckily, multiple rows of TM can easily be retrieved in a parallel way.

Algorithm 1: RAF 2-1 for retrieving a TM row $\mathbf{a} \in \mathbb{C}^N$

- 1: **Input:** Data $\mathbf{y} \in \mathbb{R}_+^P$ with $\{y_j\}_{1 \leq j \leq P}$, $\mathbf{X} \in \mathbb{C}^{N \times P}$; number of iterations T ; step size $\mu = 3$ and weighting parameter $\beta = 5$; subset cardinality $|S| = \lfloor 3P/13 \rfloor$, and exponent $\alpha = 0.5$.
- 2: **Construct** S to include indices associated with the $|S|$ largest entries among $\{y_j\}_{1 \leq j \leq P}$.
- 3: **Initialization:** Let $\mathbf{a}^0 := \sqrt{\sum_j y_j^2/P} \cdot \tilde{\mathbf{a}}^0$ where $\tilde{\mathbf{a}}^0$ is the unit-norm principle eigenvector of the Hermitian matrix

$$\mathbf{H} := \frac{1}{|S|} \mathbf{X} \cdot \text{diag}(\tilde{y}_1^\alpha, \tilde{y}_2^\alpha, \dots, \tilde{y}_P^\alpha) \cdot \mathbf{X}^H,$$

$$\text{where } \tilde{y}_j^\alpha := \begin{cases} y_j^\alpha, & j \in S \\ 0, & \text{otherwise} \end{cases}.$$

- 4: **Gradient-descent loop**
Step 1: for $t = 0$ to $\lfloor \frac{2}{3}T \rfloor - 1$ do

$$\mathbf{a}^{t+1} = \mathbf{a}^t - \frac{\mu}{P} \cdot \mathbf{X} \left[\mathbf{w} \circ (\mathbf{X}^H \mathbf{a}^t - \mathbf{y} \circ \frac{\mathbf{X}^H \mathbf{a}^t}{|\mathbf{X}^H \mathbf{a}^t|}) \right]$$

- Step 2: for $t = \lfloor \frac{2}{3}T \rfloor$ to $T - 1$ do

$$\mathbf{a}^{t+1} = \mathbf{a}^t - \frac{\mu}{P} \cdot \mathbf{X} \left[\mathbf{w} \circ (\mathbf{X}^H \mathbf{a}^t - \mathbf{y} \circ \frac{\mathbf{X}^H \mathbf{a}^t}{|\mathbf{X}^H \mathbf{a}^t|}) \right]$$

$$\text{where } \mathbf{w} := |\mathbf{X}^H \mathbf{a}^t| \oslash (|\mathbf{X}^H \mathbf{a}^t| + \beta \mathbf{y}).$$

- 5: **Output:** \mathbf{a} .
-

4.3 Results

4.3.1 Accelerated gradient descent for RAF

As an early effort, the ways to accelerate the convergence of RAF have been studied, by using an adaptive step size or more advanced gradient descent solvers. First, we adopted the Barzilai-Borwein method for non-monotonic backtracking line-search of step size, which was compared with the fixed one ($\mu = 3$). As seen in Figure 4.1 a, the measurement error of using adaptive μ drops slightly more rapidly than that of fixed μ within the first 20 s of running time, while the latter eventually declines to a lower level. This shows the adaptive step size method is not very

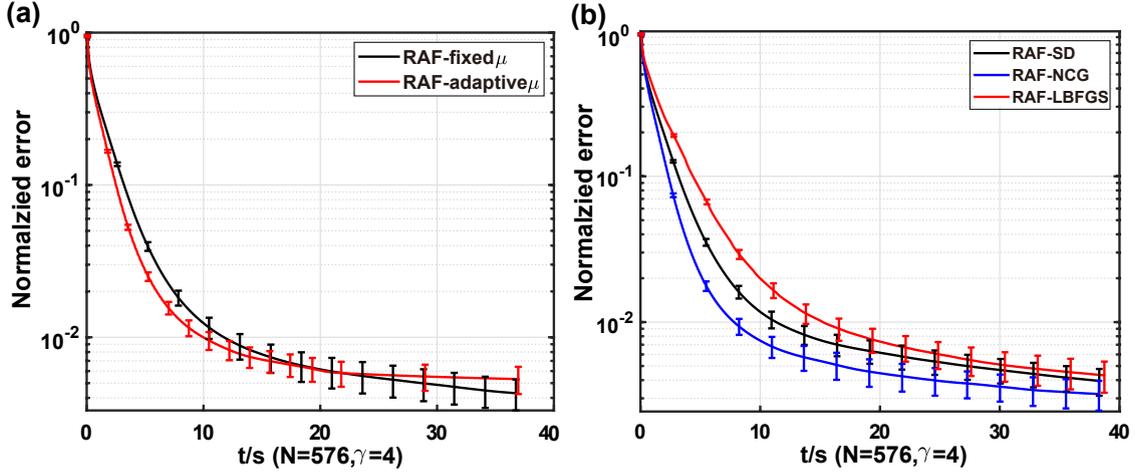


Figure 4.1: (a). Normalized curves of error as a function of running time for RAF with a fixed step size (μ) or an adaptive one. (b) Normalized curves of error as a function of running time for RAF with steep descent (SD), non-linear conjugate gradient (NCG), or L-BFGS.

effective, although it could be better with parameter finetuning. As for the gradient descent solver, besides the regular steep descent using the negative first derivative (*i.e.*, the gradient) as the descent direction, acceleration methods such as NCG and L-BFGS were employed for comparison. Since NCG and especially L-BFGS entail more computations per iteration than SD, for fair comparison, the curves of error as a function of running time (instead of iterations) for different optimization methods were compared, as shown in Figure 4.1b. We can see that NCG has the fastest convergence with the same running time. The reason that L-BFGS method is even slower in the declining trend of error is attributed to the far more seconds per iteration it requires. In fact, the average number of iterations taken by SD, NCG, and L-BFGS are 671, 656, and 411, respectively. The convergence for L-BFGS could be the fastest if compared from the perspective of error declining versus iterations.

4.3.2 The best two-step iteration ratio for RAF 2-1

There are two steps in the gradient iteration process regarding the proposed RAF 2-1. In order to determine the number of iterations in Steps 1 and 2 (with the total number fixed) for the best performance, numerical experiments were conducted to compare the focusing efficiency achieved by RAF 2-1 under a series of iteration ratios. Moreover, since GGS 2-1 also involved the two-step gradient descent, it inspires this work and used for performance comparison. Therefore, the best ratio of iteration of GGS should also be determined. Figure 4.2 gives the results of both RAF 2-1 and GGS 2-1, with a total iteration of 300. It can be observed for RAF 2-1, there are minor differences of focusing efficiency among different iteration ratios, while

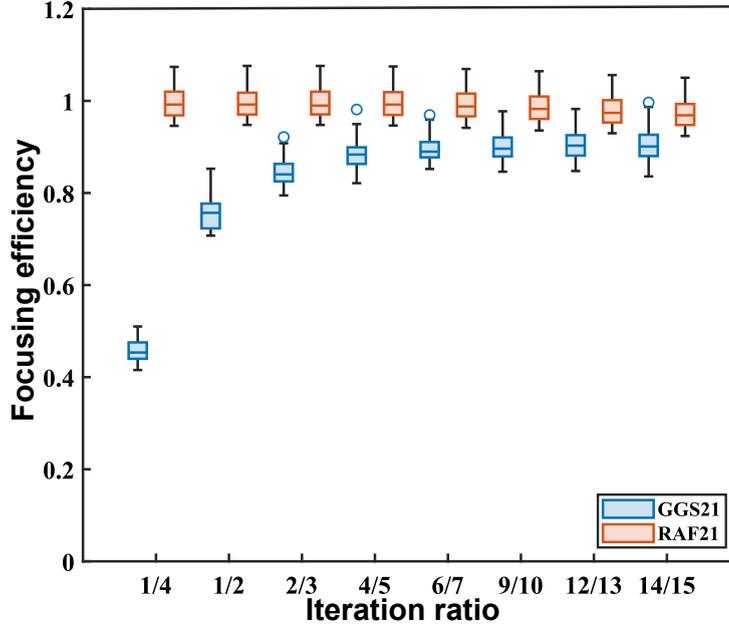


Figure 4.2: (a). Normalized focusing PBR achieved by GGS 2-1 and RAF 2-1 under a series of iteration ratios during their two-step gradient iterations, respectively.

the one at $2/3$ is chosen as the best iteration ratio due to a slight advantage. As for GGS 2-1, the best focusing efficiency is around an iteration ratio of $9/10$, which is also consistent with the original research that adopted 287 and 34 iterations in Steps 1 and 2 for GGS 2-1, respectively.

4.3.3 Numerical evaluation of RAF and RAF 2-1

As mentioned in the Methods section, the modified version, RAF 2-1, is more effective for TM retrieval. To evaluate how the performance of RAF 2-1 is better than the original RAF, a numerical experiment in a noiseless condition was performed for retrieving the TM that corresponds to 400 output modes. The curves of the averaged errors after normalization are presented in Fig. 4.3. The measurement error for the i^{th} row of TM is defined as

$$\text{error}_i = \left\| \left\| \mathbf{X}^H \hat{\mathbf{a}}_i - \mathbf{y}_i \right\|_2 \right\|_2, i = 1, \dots, 400 \quad (4.10)$$

where $\hat{\mathbf{a}}_i$ is the estimate of the i^{th} row of TM and other notations are with the same meaning as in the Methods section. It can be observed that the error of RAF 2-1 can finally decline to a level of as low as 10^{-4} , much lower than that of RAF. This indicates a more accurate result of TM retrieval by RAF 2-1.

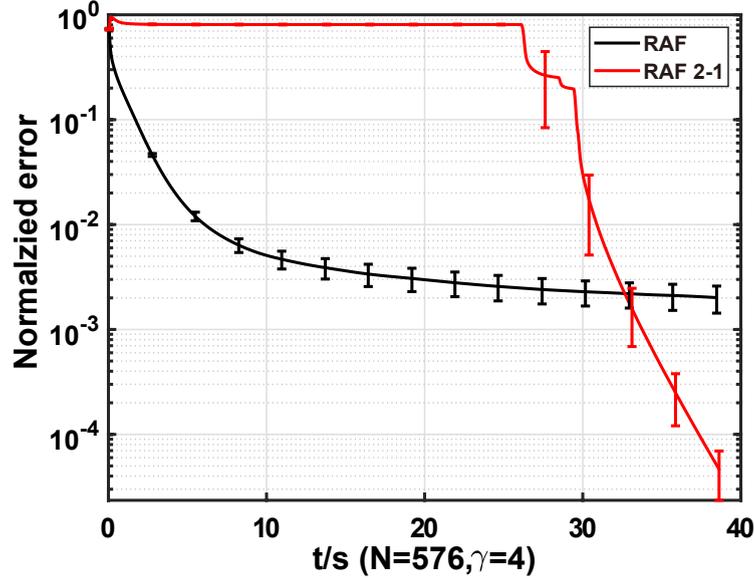


Figure 4.3: Normalized curves of error as a function of running time for RAF and RAF 2-1 when $N = 576$ and $\gamma = 4$. Note the error bars denote the standard deviations of 30 repeated tests.

4.3.4 Numerical comparisons between RAF 2-1 and representative algorithms

Numerical evaluations were conducted at first to assess the efficiency of the proposed RAF 2-1, with comparisons with several representative TM retrieval algorithms including the pioneering prVBEM and the recent GGS 2-1 that outperformed previous ones. Note that unless otherwise specified, all the following simulations were conducted using a PC with an Intel Xeon CPU (3.50 GHz, 6 cores) and 64 GB RAM in the environment of MATLAB 2022a. For each algorithm, the performance is evaluated by the efficiency of focusing with the retrieved TM, which is indicated PBR. This is performed by taking the conjugate of one row of TM to construct the phase mask for focusing light onto the corresponding position. The TM was modelled using the speckle field produced by random phase mask with zero-padding in the Fourier domain, whose elements obeyed a circular Gaussian distribution. According to Ref. [8], the theoretical PBR of focusing is linearly proportional to the input size N , as given by

$$\eta = \frac{\pi}{4}(N - 1) + 1. \quad (4.11)$$

The focusing efficiency a certain TM retrieval algorithm can achieve is typically determined by the iterations (or the running time) it has taken and the sampling ratio ($\gamma = P/N$).

Fig. 4.4a shows the schematic diagram of TM retrieval for a MMF. We first examined the focusing performance of different TM retrieval algorithms under a range of running times when the sampling ratio was fixed to be $\gamma = 4$. The input phase mask had a size of 24×24 . A total of 20×20 foci were produced sequentially that corresponded to 400 rows of TM to be retrieved, with the mean PBRs of the foci as the focusing PBR. The focusing PBR is further normalized by the theoretical value to obtain the focusing efficiency. Fig. 4.4b presents the average focusing efficiency achieved by different methods with a running time ranging from 1 to 60 s based on 30 repeated tests. It can be seen RAF 2-1 reaches the optimum efficiency after running for ~ 8 s, while GGS 2-1 requires quite longer running time (~ 40 s) and prVBEM cannot fully approach the optimum within 60 s. This indicates our nonconvex method is superior in algorithm convergence given the same condition. Fig. 4.4c additionally shows the number of iterations versus running times, which reveals the seconds per iteration for prVBEM, GGS 2-1, and RAF 2-1 are roughly 5:1:1 in such a case. Hence, the nonconvex approach is at least as highly efficient as GGS 2-1 in computation time, and both are much better than prVBEM.

The influence of sampling ratio was also explored for TM retrieval algorithms, by fixing the running time to be 20 s when $N = 576$. As seen in Fig. 4.4d, the focusing efficiency is close to 0 for all algorithms when $\gamma = 1$, which increases dramatically starting from $\gamma = 2$ for RAF 2-1 and GGS 2-1. Higher focusing efficiency can be achieved with a larger γ for a TM retrieval algorithm, since more measurements taken per degree of freedom in the signal allows for more accurate recovery. Obviously, our RAF 2-1 outperforms its competing peers as it averagely realizes more than 95% efficiency when $\gamma = 3$ and 100 % when $\gamma = 4$. By comparison, GGS 2-1 requires a sampling ratio of 5, while prVBEM requires 7 for the optimal performance under the same conditions.

Since measurements are inevitably contaminated with noise in practice, a good noise-robustness is preferred for a TM retrieval algorithm. Thus, the algorithm performances were also evaluated under a range of SNR levels using simulated noisy measurements. In the simulation, a range of signal-to-noise (SNR) levels using simulated noisy measurements. In the simulation, a multiplicative noise is added to the output field intensity $\mathbf{I} \in \mathbb{R}_+^P$. The SNR is defined as

$$SNR = \langle \mathbf{I} \rangle / \sqrt{\langle (\boldsymbol{\varepsilon} - \langle \boldsymbol{\varepsilon} \rangle)^2 \rangle}, \quad (4.12)$$

where $\boldsymbol{\varepsilon} = \mathbf{I}_{\text{noisy}} - \mathbf{I}$, which denotes the noise vector, $\mathbf{I}_{\text{noisy}}$ is the noisy measurement vector, and $\langle \cdot \rangle$ takes the average for the elements inside. For the focusing results of multiple foci,

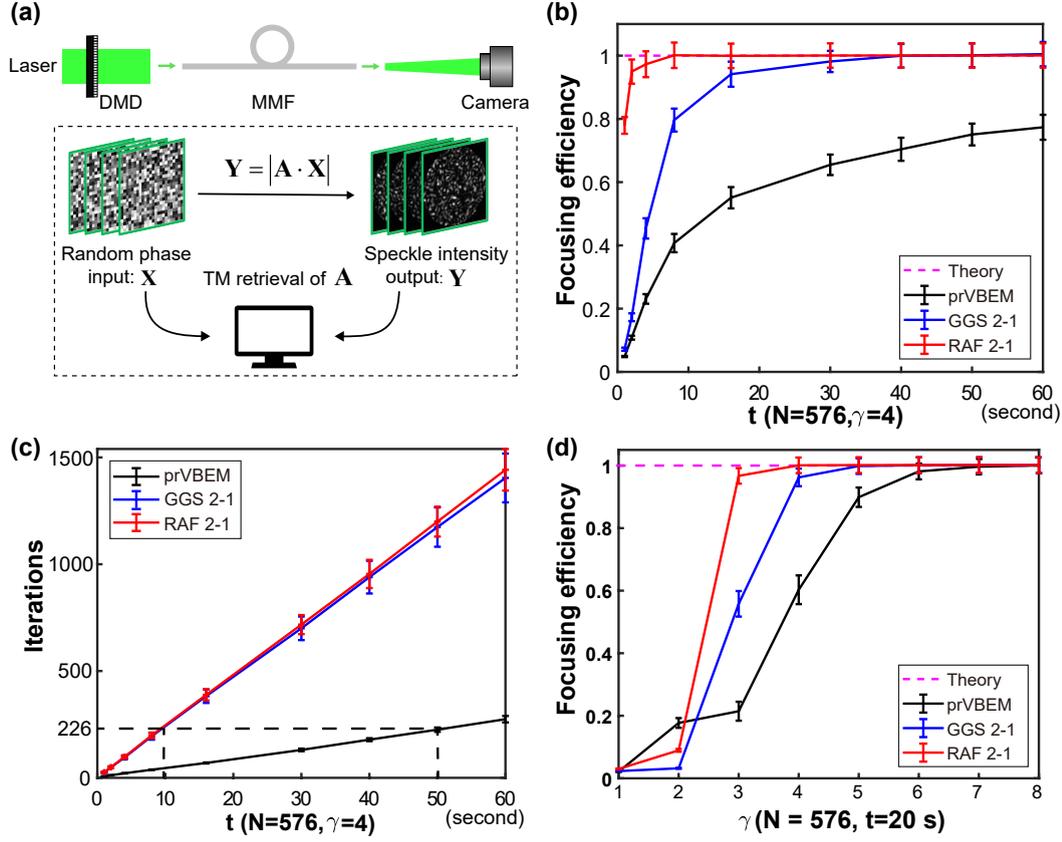


Figure 4.4: Theoretical comparisons of TM retrieval performances of prVBEM, GGS 2-1 and our RAF 2-1. (a) Schematic diagram of TM retrieval for a MMF. (b) Focusing efficiency achieved by different algorithms under a range of running times when $N = 576$ and $\gamma = 4$. (c) The iterations taken by different algorithms versus running times for the case of (b). (d) Focusing efficiency achieved by different algorithms under a range of γ when using $N = 576$ and a running time of 20 seconds. Note the error bars denote the standard deviations of 30 repeated tests.

uniformity is an important metric to measure the focus quality. The focusing uniformity (μ) is given by

$$u = 1 - \sqrt{\frac{\langle (I_k - \langle I_k \rangle)^2 \rangle}{\langle I_k \rangle}}, k \in K \quad (4.13)$$

where K is a set of the indexes of foci. With parameter settings that $N = 1024$, $\gamma = 5$ and a running time of 50 s, the results of focusing efficiency and uniformity of 400 foci produced using various TM retrieval algorithms are given in Fig 4.5 (a-b). Note the original RAF was also included to highlight the improved anti-noise capability by our modification.

It is found that a maximum improvement of 15.5% and 32.4% in terms of focusing efficiency can be realized by RAF 2-1 over RAF and GGS 2-1 respectively when the SNR is relatively low (e.g., no more than 3.1). Such a performance advantage of RAF 2-1 over other algorithms weaken

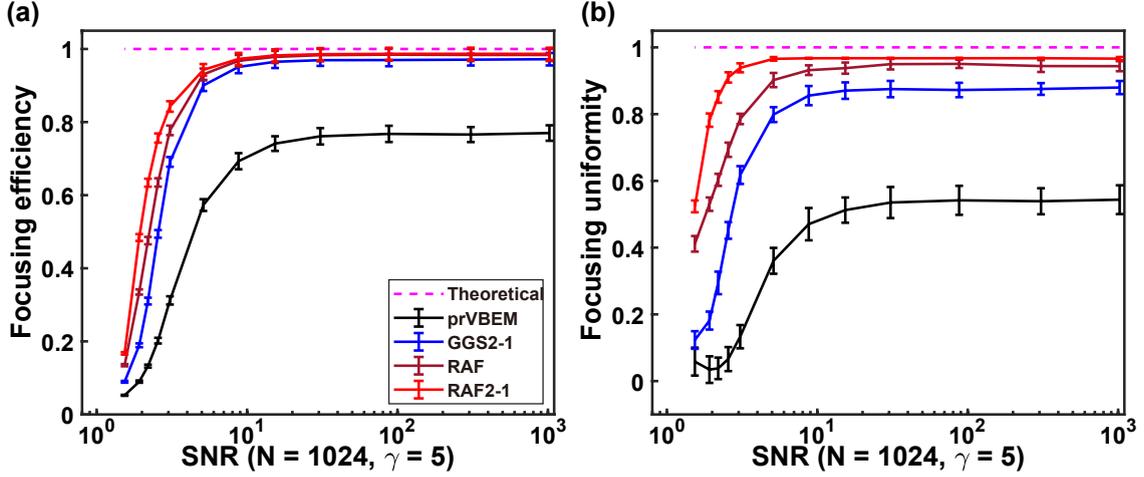


Figure 4.5: Simulated (a) focusing efficiency and (b) focusing uniformity achieved by prVBEM, GGS 2-1, RAF, and RAF 2-1 respectively under different SNR levels when using $N = 1024$, $\gamma = 5$, and a running time of 50 s. Note the error bars denote the standard deviations of 30 repeated tests.

as the SNR increases, and the focusing efficiency is almost the same when the SNR is about 15. This explains why RAF was excluded in the previous noiseless tests. Besides, an obvious improvement of focusing uniformity is achieved by RAF 2-1, which is at best 25.2% and 60.1% higher than RAF and GGS 2-1 respectively when the SNR is 1.92. The advantage of RAF 2-1 over RAF becomes minor when the SNR is large, while it still sees $\sim 9\%$ improvement than GGS 2-1. Overall, algorithm performances in both focusing efficiency and uniformity follows the order: RAF 2-1 > RAF > GGS 2-1 > prVBEM. The difference between GGS 2-1 and RAF is relatively small, whereas prVBEM falls behind GGS 2-1 considerably.

4.3.5 Experimental results

The experimental configuration for retrieving the TM of an MMF is illustrated in Figure 4.6. A beam from a 532 nm continuous-wave laser (EXLSR-532-300-CDRH, Spectra Physics, USA) was expanded by a $40\times$ objective lens and collimated by a lens (L1). The linearly polarized beam was then modulated into circularly polarized by a quarter-wave plate ($\lambda/4$) before impinging onto a digital micromirror device (DMD, V-9501, VIALUX, Germany). Based on the Lee hologram technique, the DMD, combined with a $4f$ system composed of L2, iris, and L3, allowed for phase modulation at a high-speed pattern rate of up to 23 KHz, rendering fast data acquisition for TM calibration. The phase-encoded and shrunk beam was subsequently coupled into an MMF of 0.22 NA and diameter of $105\ \mu\text{m}$ (SUH105, Xinrui, China) by a collimator

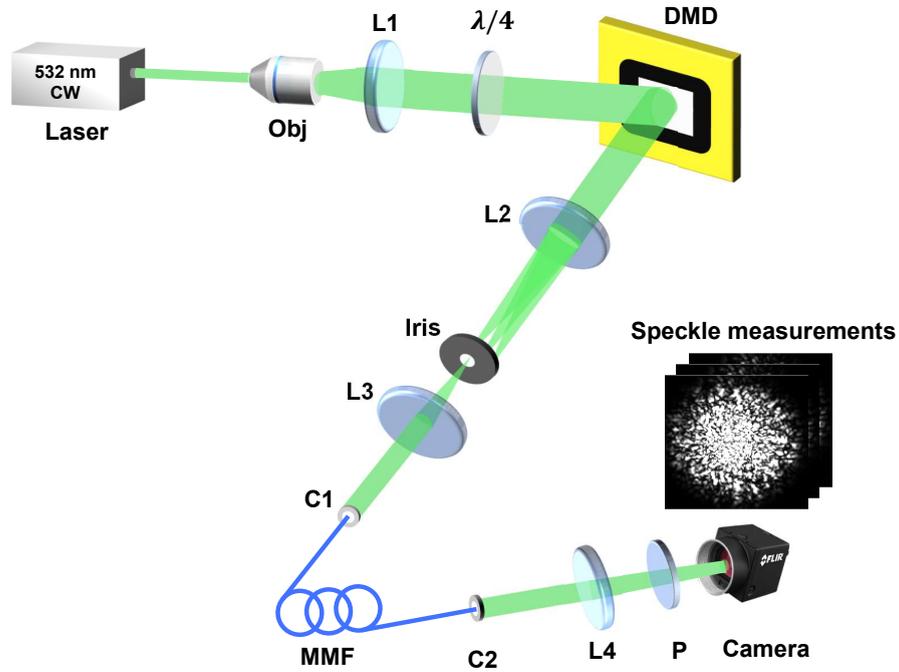


Figure 4.6: Experimental configuration for retrieving the TM of an MMF with speckle intensity measurements. C, collimator; DMD, digital mirror device; L, lens; P, polarizer; MMF, multimode fiber; Obj, objective lens. $\lambda/4$, quarter-wave plate.

(C1). The transmitted light was imaged with a collimator (C2), too. The beam then passed through a lens (L4) for convergence and a polarizer before being captured by a CMOS camera (BFS-U3-04S2M, FLIR, USA). For TM retrieval, there was a sequence of speckle intensity measurements under the input modulations of random phase.

In the experiment, the performances of using different TM retrieval algorithms to control light delivery despite scattering were compared from the aspects of single-spot and multi-spot focusing through MMF. For single-spot focusing, a total of 20×20 foci were generated sequentially on the working plane of the MMF, which meant 400 rows of TM were to be retrieved. The sampling ratio was set to be 5 for all algorithms to ensure the quality of TM retrieval, given that the acquired speckle data suffered from noises such as shot noise, dark current noise, and readout-noise. Fig. 4.7a presents the histogram distribution of the normalized PBRs of 400 foci achieved with different algorithms. Since the experimentally acquired TM of the MMF also obeyed the circular Gaussian distribution, it could be reasonable to use Eq. 4.11 for normalizing the focus PBRs generated by the MMF and calculating the focusing efficiency. It can be seen that the average focusing efficiency (denoted by the crosses in the boxplots) are 16.45%, 26.01%, 37.46%, and 55.17% for prVBEM, GGS 2-1, RAF, and RAF 2-1 respectively. Surprisingly,

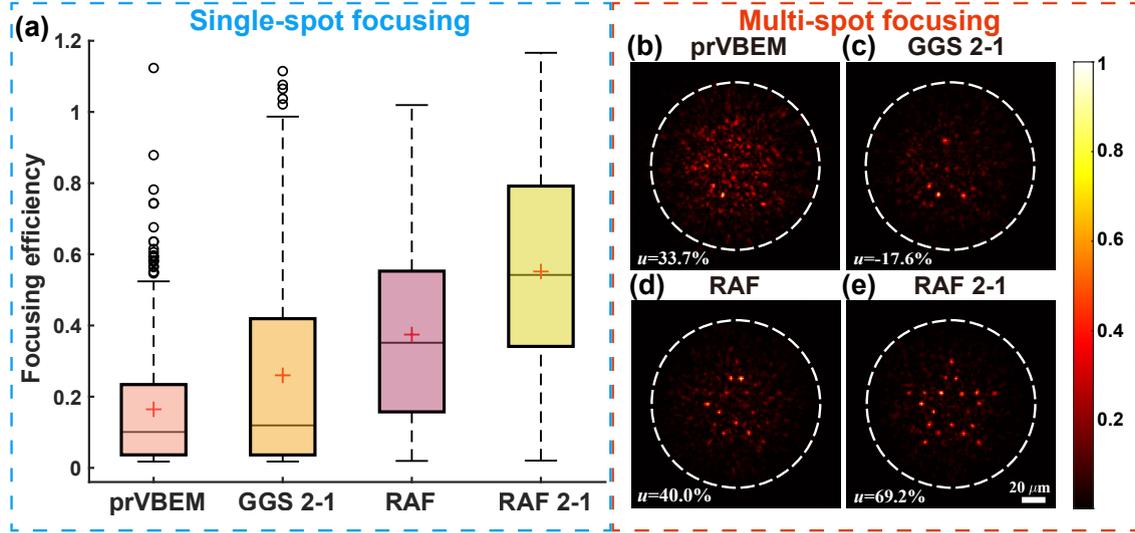


Figure 4.7: Comparison of light focusing results through MMF with the TMs retrieved by different algorithms. (a) The histograms of normalized PBR of 20×20 foci and (b) the results of multi-spot focusing (pentagram) in the output field of MMF, both obtained by prVBEM, GGS 2-1, RAF, and RAF 2-1 respectively with $N = 1024$ and $\gamma = 5$. Note the crosses in (a) represent the mean values, the white dashed circles in (b) show the fiber region. The scale bar in (b-e) was 20 μm .

RAF 2-1 achieved a considerably higher efficiency than RAF, *i.e.*, 17.72%, in comparison to the simulation results in Figure 4.5a. The reasons may include the noise type difference in the simulation and experiment, the larger TM dimension in the experiment (more challenging for recovery so that the gap is more obvious), *etc.*. According to the boxplots of Figure 4.7a, quite a few foci approached or even surpassed the theoretical PBR for GGS 2-1, RAF, and especially RAF 2-1. However, their overall focusing efficiencies of 400 different foci on the working plane of MMF still saw a distance from the ideal level, using the retrieved TMs when $\gamma = 5$.

In addition to single-spot focusing, multi-spot focusing experiment was also conducted under the same condition. This was achieved by superposing multiple phase-conjugate rows of the retrieved TM to construct a phase mask. The results of light focusing onto a pentagram pattern composed of 20 spots by different algorithms are shown in Fig. 4.7b. The focusing uniformity were $u = 33.7\%$, -17.6% , 40.0% , and 69.2% , respectively for the four algorithms. It can be observed that only RAF 2-1 produced a high-quality pentagram pattern by focusing light on all the pre-selected positions, thanks to its superior performance of TM retrieval.

The accuracy of TM retrieval by our RAF 2-1 was further compared with the off-axis holography, *i.e.*, the golden standard for the measurement of TM. To do so, we scanned across the whole fiber region on the working plane of the MMF, so that a map of focusing PBR could

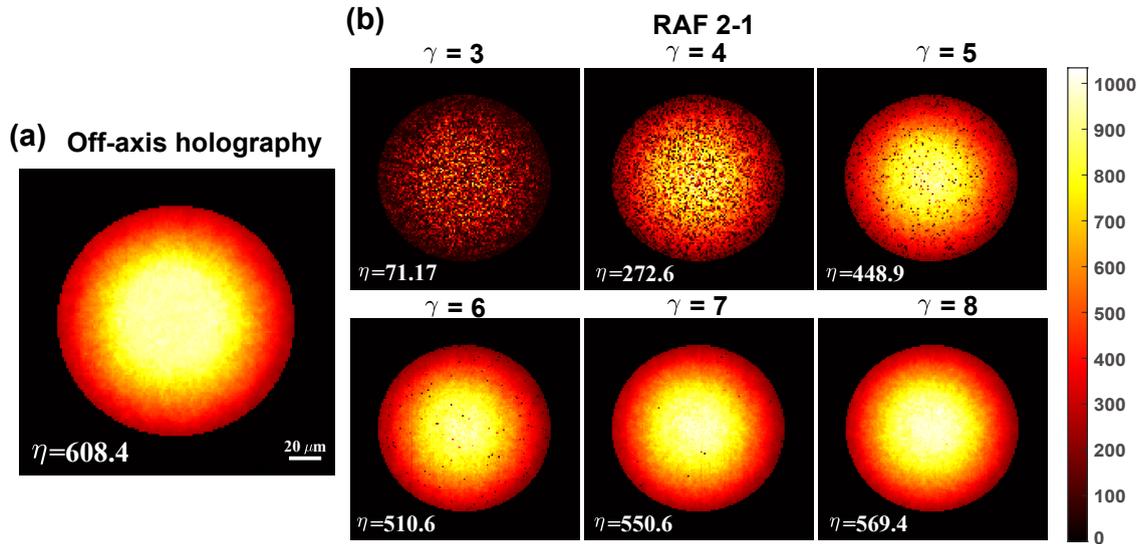


Figure 4.8: Comparison of the PBR maps of focusing on the working plane of the MMF using the TMs (a) measured by off-axis holography and (b) retrieved by RAF 2-1 under a range of γ with $N = 1024$. The scale bar in (a-b) was $20 \mu\text{m}$.

be synthesized, which was used to fully evaluate the accuracy of TM. The fiber region was determined by identifying the largest connected region in the binarized image taken when all pixels on the DMD were turned on. Using circular fitting of the fiber region, the center and radius of the fiber region were obtained, which were then used to create a binary mask of the fiber region. In the experiment, there were 8685 pixels inside the circular fiber region of the 140×140 output field, which correspond to 8685 rows of TM to be retrieved. Fig. 4.8 summarizes the results of focusing PBR maps with the TM measured by off-axis holography and the TMs retrieved by RAF 2-1 under a range of γ from 3 to 8. The mean PBR by the golden standard method is 608.4. Compared with the theoretical PBR (*i.e.*, $\eta = 804$), it is reasonable given that the focusing quality degraded in the fiber fringe area due to the inhomogeneous mode excitation inside the MMF. As for RAF 2-1, there are many dark spots in the PBR map synthesized under small γ , indicating poor accuracy of the corresponding rows of TM being retrieved. With a larger γ , the PBR map becomes more homogeneous with less dark spots, which means an overall improvement of the TM accuracy. Notably, when $\gamma = 8$, the PBR map by RAF 2-1 is comparable to that of the holography, with a mean PBR of 569.4 reaching $\sim 93.6\%$ efficiency of the golden standard experimentally. In practice, the choice of the sampling ratio should strike a balance between the computing costs and the expected focusing performance with the recovered TM. Compared to off-axis holography, the key advantage of RAF 2-1 is

the elimination of reference beam and interferometry, which is compatible with more applications. Additionally, with parallel operation and GPU implementation, the TM retrieval process by RAF 2-1 was fast enough. For example, under $\gamma = 8$, retrieving an 8685×1024 TM by RAF 2-1 averagely took 42.3 s, when using a computer with an Intel Xeon CPU E5-1650 v3 @3.50 GHz, a NVIDIA RTX2080 Ti GPU, and 128 GB RAM.

Using the retrieved TM by RAF 2-1, one can further reconstruct an input object from intensity-only speckle measurement with one more phase retrieval. The reconstruction result relies heavily on the quality of the recovered TM, which acts as the measurement matrix. Fig. 4.9a shows the results of reconstructing a 32×32 phase object of a Chinese character "光" (meaning "light"), by taking 100 iterations with the TM of MMF retrieved by RAF 2-1 when γ increased from 3 to 8. Pearson correlation coefficient (PCC) is used to quantify the fidelity between the reconstructed image I_R and the ground truth I_G , which is given by

$$\text{PCC} = \frac{\langle (I_R - \langle I_R \rangle) (I_G - \langle I_G \rangle) \rangle}{\sqrt{\langle (I_R - \langle I_R \rangle)^2 \rangle \langle (I_G - \langle I_G \rangle)^2 \rangle}}. \quad (4.14)$$

The curves of PCC under different cases of γ are also provided in Fig 4.9b. The upsurges of PCCs at 67th iteration confirm that the signal recovery is significantly boosted after the gradient heating in the first 2/3 iterations. The final PCCs are: 0.02, 0.16, 0.65, 0.74, 0.78, 0.80 for $\gamma = 3, 4, 5, 6, 7, 8$ respectively. As can be observed, the reconstructed image could be recognized starting from $\gamma = 5$ and attains the best quality when $\gamma = 8$. Nonetheless, there is still speckle background noise among the images (Fig. 4.9a), which is common in the reported empirical speckle imaging results with TM method [136, 154], mainly because of the imperfect fidelity of the recovered TM. The reconstruction quality can be further improved with a larger γ to overcome the influence of noise. To summarize, image transmission through the MMF is demonstrated with the proposed nonconvex approach, which further validates the effectiveness of the TM retrieved.

4.4 Discussion

There have been various phase retrieval algorithms used for solving the TM retrieval problem, as introduced earlier. RAF, as one of the best in the nonconvex family, has been reported previously [171] to be highly competitive for image restoration from speckle measurement. To the

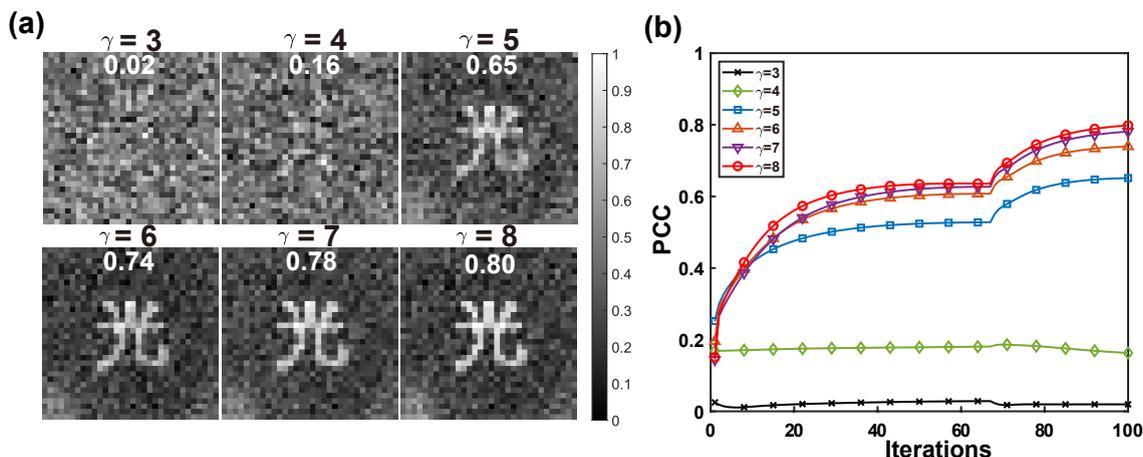


Figure 4.9: Comparison of image transmission results through MMF using the retrieved TMs by RAF 2-1 under a range of γ with $N = 1024$. (a) Normalized reconstructed images for an object of Chinese character "光", with the values of PCC to the object labeled (b) The progression curves of PCC during the iterative reconstruction.

best of our knowledge, we first adopted it for non-holographic calibration of TM [157]. More importantly, our modified version, RAF 2-1, with an additional step of gradient heating, has shown remarkable improvement in the robustness against noise and the TM retrieval accuracy in both simulations and experiments. The numerical evaluation of RAF and RAF 2-1 can be further seen in Appendix B. Besides the above modification, we resort to speeding up the convergence of RAF for TM retrieval. However, the improvements are not very impressive.

There are also several limitations in the study. In the experimental setup, the MMF output field was relayed by a collimator instead of an objective lens. Consequently, the working plane of the MMF was immovable, which had a certain distance (about tens of micrometers) away from the fiber end. That said, the setup was sufficient for retrieving a reliable TM and focusing on the working plane for demonstration. An objective lens is needed only for measuring the TMs corresponding to different working planes. In addition, since there is phase ambiguity for the formulated LSE objective function in Eq. (4.4), a phase offset exists for each row of the retrieved TM. However, it does not affect the intensity of the generated 2D foci. Further phase correction [146] is indispensable when it comes to MMF 3D volumetric focusing and imaging.

It has been reported that TM retrieval can be implemented by fast Fourier transform (FFT) [146], which greatly accelerates the computations. Basically, the principle is by designing the probing matrix using a DFT matrix, the matrix-vector multiplication (MvM) between the probing matrix and the TM vector is equivalent to DFT of the vector. Hence, TM retrieval with FFT is also applicable to this study. Mathematically, the probing matrix is designed to

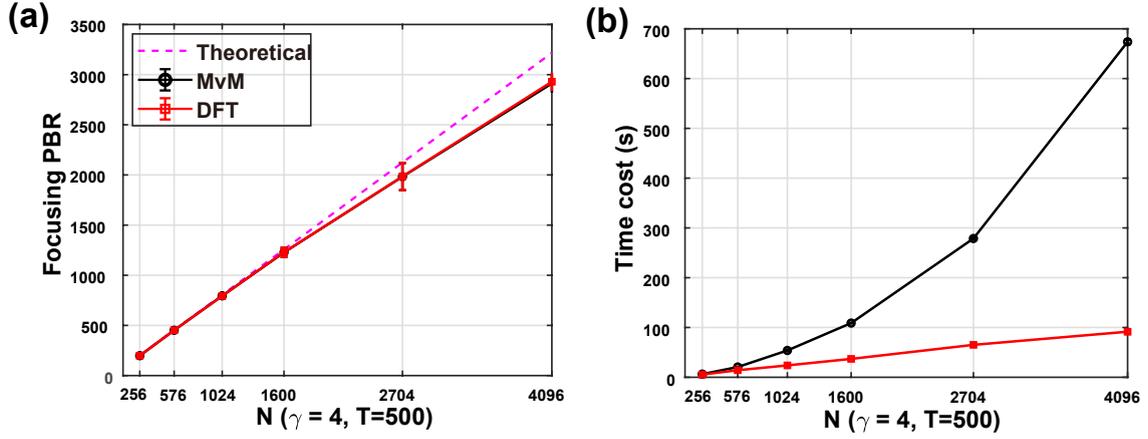


Figure 4.10: Simulations of the retrieval of TM corresponding to 20×20 foci under a range of input size (N) with a sampling ratio of 4 and an iteration number (T) of 500, implemented by using MvM and DFT, respectively. (a) Focusing PBR versus N . (b) Time cost versus N . Note the error bars denote the standard deviations of 5 repeated tests.

be $\mathbf{X} = [\mathbf{X}_1 \dots \mathbf{X}_i \dots \mathbf{X}_\gamma]$ where $\mathbf{X}_i = \text{diag}(e^{j\phi_i^1}, e^{j\phi_i^2}, \dots, e^{j\phi_i^N}) \cdot \mathbf{W}$ with $i = 1, \dots, \gamma$ and \mathbf{W} is the N -by- N DFT matrix. As such, considering two typical MvM forms:

- \mathbf{X} multiplied by a vector \mathbf{z} ($\mathbf{z} \in \mathbb{C}^N$). Let $\mathbf{z} = [\mathbf{z}_1; \dots; \mathbf{z}_i; \dots; \mathbf{z}_\gamma]$ with $\mathbf{z}_i \in \mathbb{C}^N$. Then we have

$$\mathbf{X} \cdot \mathbf{z} = \sum_{i=1}^{\gamma} \mathbf{X}_i \mathbf{z}_i = \sum_{i=1}^{\gamma} \left(e^{j\phi_i^1}, e^{j\phi_i^2}, \dots, e^{j\phi_i^N} \right)^T \circ \text{fft}(\mathbf{z}_i), \quad (4.15)$$

where $\text{fft}(\cdot)$ denotes FFT operation.

- \mathbf{X}^H multiplied by a vector \mathbf{z} ($\mathbf{z} \in \mathbb{C}^N$). We have $\mathbf{X}^H \cdot \mathbf{z} = [\mathbf{X}_1^H \mathbf{z}; \dots; \mathbf{X}_\gamma^H \mathbf{z}]$, which reads

$$\begin{bmatrix} N \cdot \text{ifft} \left(\left(e^{-j\phi_1^1}, e^{-j\phi_1^2}, \dots, e^{-j\phi_1^N} \right)^T \circ \mathbf{z} \right) \\ \vdots \\ N \cdot \text{ifft} \left(\left(e^{-j\phi_\gamma^1}, e^{-j\phi_\gamma^2}, \dots, e^{-j\phi_\gamma^N} \right)^T \circ \mathbf{z} \right) \end{bmatrix}. \quad (4.16)$$

Compared to MvM, the FFT implementation reduces the time complexity from $O(MN^2)$ to $O(MN \log N)$ for retrieving one row of TM. Besides, it also circumvents the need of storing the probe matrix in memory. By designing the probing matrix with DFT matrix, we have compared the TM retrieval results with MvM and FFT implementation respectively, as given in Figure 4.10. It can be seen from Figure 4.10b that the acceleration by FFT is more obvious with a larger input size N , and a maximum of $\sim 7 \times$ speed-up is achieved at $N = 4096$. However, it is also noted that the retrieved TMs cannot always reach the ideal focusing efficiency at

$\gamma = 4$, especially when N is very large (see Figure 4.10a). This is ascribed to a lack of enough modulation diversity for TM recovery, and a larger sampling ratio may alleviate it.

4.5 Conclusion

In conclusion, we have proposed a modified nonconvex approach, RAF 2-1, for retrieving the TM of MMF based on speckle intensity measurements. Theoretically, RAF 2-1 can achieve optimum focusing efficiency with less running time or sampling ratio than the previously reported TM retrieval methods. The experimental results of light control through a MMF confirm a comparable performance of RAF 2-1 to the golden standard holography method for TM measurement. RAF 2-1 is also computationally efficient that took averagely 42.3 s to recover a 8685×1024 TM ($\gamma = 8$) on a regular computer under parallel operation and GPU implementation. Endowed with the advantages of optimum efficiency, fast execution, and a reference-less setup, RAF 2-1 allows for broad applications in MMF-based imaging, manipulation, and treatment *etc.*

Chapter 5

Pilot Study on Multimode Fiber-based Fluorescence Imaging with Transmission- and Reflection-mode Calibration Methods

In Part III of this thesis, with the ability for manipulating the light output of an MMF, we are capable of conducting MMF-based endoscopic imaging. For Chapter 5, we mainly explore MMF-based fluorescence imaging, with special discussion on the online calibration method of MMF. Briefly, we adopt two schemes. In Scheme 1, MMF-based point scanning fluorescence imaging is investigated with conventional transmission-mode calibration method. We have confirmed the imaging performance on the samples of microspheres and mouse brain slice. In Scheme 2, a fluorescence-based intensity TM for MMF calibration in the reflection mode is proposed based on the unmixing of epi-detected fluorescence speckle. Fluorescent targets could be further reconstructed or localized based on the above unmixing results. In principle, Scheme 2 offers a feasible solution for the re-calibration of MMF in the absence of distal fiber access during fluorescence imaging and has been verified by simulation. However, it achieved limited success in preliminary experiments and hence requires further validation.

5.1 Introduction

Optical imaging in biomedicine holds the advantages of safety, no ionizing radiation, and sensitivity to tissue abnormalities. However, biological tissues are opaque to light and arouse strong optical scattering due to the highly inhomogeneous distribution of refractive index, which fundamentally poses a limit to the depth of high-resolution optical imaging [1, 4]. Although great efforts have been taken by using multiphoton fluorescence microscopy [172–175] that utilizes longer excitation wavelength to reduce scattering and (or) adaptive optics technology [175] to overcome scattering, the reported maximum imaging depth so far is merely 1-2 mm beneath mouse brain cortex [172, 174].

To meet the needs of high-resolution deep-tissue diagnosis and image-guided minimally invasive surgery, medical endoscopes, which invasively introduce a light source into tissues and directly observe tissue subsurface, have been developed. There are early-stage white light rigid endoscope and modern electronic endoscope based on a wide-field image sensor, which has been the working horse in clinical applications. However, electronic endoscopes have limited spatial resolution (10-100 μm) and poor sensitivity, with large probe size. Hence, CLE [30, 31], such as pCLE and eCLE, has been developed since 2000s. However, the large diameters of CLE probes still see limitations in imaging sensitive tissue regions such as deep brain.

In the past decade, thanks to the invention of optical WFS technology, the field has witnessed the enormous development of single MMF-based endomicroscopy [34–36], which is minimally invasive as the size of fiber core is typically 50 \sim 200 μm . Although the light output from the MMF is naturally diffusive, with WFS the light field can be controlled for selective excitation of the optical modes in MMF to form diffraction-limited focus at the distal side, available for laser-scanning imaging. Various imaging modalities have been realized through MMF, such as fluorescence [37], PA [38], reflectance [20], nonlinear [39] imaging and so on. Here, we focus on fluorescence imaging, which has excellent molecular specificity and gains most applications in deep-tissue optics.

The major progress of MMF-based fluorescent imaging is summarized in 5.1, and the key evaluation metrics for MMF imaging include spatial resolution, FOV, imaging speed, and application impact, etc. We can further extract several key technical points and development trends in this field: (1) calibration of MMF, including coaxial and off-axis holography; (2) light modulator, for which DMD is dominant as it allows much faster modulation and imaging speed;

Table 5.1: Major progress of MMF fluorescence endomicroscopy.

Reference @ year	Imaging samples	Modulator + TM measurement	Resolution (μm)	Field of view (μm)	Imaging size (pixels) @ speed
Ref.[176] @ 2012	Fluorescent beads	LC-SLM + on-axis holography	1.6	$\sim 100 \times 100$	60×60 @ 0.005 Hz
Ref.[177] @ 2017	Fluorescent beads, monkey brain slice	DMD + on-axis holography	~ 2	80×80	/ @ 0.05 Hz
Ref.[37] @ 2018	Hippocampus tissue, <i>in-vivo</i> mouse brain	DMD + on-axis holography	~ 2.1	100×100	/ @ 7 ~ 15 Hz
Ref.[178] @ 2018	<i>In-vivo</i> mouse brain dorsal striatum	DMD + off-axis holography	~ 1.35	50×50	120×120 @ 0.42 Hz
Ref.[140] @ 2018	<i>In-vivo</i> mouse brain cortex and hippocampus	DMD + off-axis holography	1.18	50×50	7000 @ 3.5 Hz
Ref.[40] @ 2023	Throughout whole mouse brain	DMD + off-axis holography	0.8	$100 \times 6000 \times 22.5$	/ @ 0.15 ~ 1000 Hz
Ref.[179] @ 2023	Invitro samples, <i>In-vivo</i> gastrointestinal tissue	DMD + off-axis holography	0.25	$110 \times 110 \times 200$	96×96 @ 2.5 Hz

(3) more advanced optical fiber technology and imaging methods, such as side-viewing MMF that enables larger FOV [40], STABLE that allows subcellular resolution *etc.*[179]; (4) imaging applications towards clinical practice, which evolves from *in-vitro* samples to *in-vivo* mouse brain and gastrointestinal tract. Despite that, MMF imaging faces several major challenges, such as limited imaging FOV and contrast, and more importantly, the extreme sensitivity to perturbations like fiber deformation and temperature change, which necessitates a fiber recalibration. There have been many works for online MMF calibration in the absence of access to distal fiber side, such as using the reflective signals based on a distal reflector [180–182], a visual beacon [183] or guide-star [184], or updating the TM based on a pre-calibrated TM library [179], as well as data-driven methods [185, 186]. However, these implementations are either technically challenging or limited in application scope. For online calibration of MMF in the scenario of fluorescent imaging, there is still a lack of feasible scheme.

Based on the previous research on wavefront optimization and TM measurement for MMF, in this chapter, we further study MMF-based fluorescence imaging and especially explore online calibration method for MMF by fully exploiting the epi-detected fluorescence speckles. We first built an MMF fluorescence point-scanning imaging system with the regular transmission-mode

calibration module (Scheme 1). Typically, a lateral resolution of $1.85\ \mu\text{m}$ and axial resolution of $35.93\ \mu\text{m}$ were realized in the experiment. Endoscopic imaging on $2\ \mu\text{m}$ fluorescent microspheres and mouse brain slice verified the reliable performance of our system. In Scheme 2, the single-pixel fluorescence detection module was replaced with a wide-field module. We constructed a fluorescence-based real-valued intensity TM (RVITM) without access to the fiber end for MMF calibration in reflection mode, which was accomplished by unmixing the acquired fluorescence speckles. Fluorescent targets could be further reconstructed or localized based on the above unmixing results. The proposed Scheme 2 is feasible in principle and has been verified by simulation, but with limited success in preliminary experimental results, which needs further study.

5.2 MMF fluorescence imaging methods

5.2.1 Experimental setup and techniques

The integrated MMF-based fluorescent imaging setup is illustrated in Figure 5.1, which is compatible for both Schemes 1 and 2. A laser beam is emitted from a 532-nm continuous wave laser (EXLSR-532-300-CDRH, Spectra Physics, USA) and split into a reference beam and a signal beam by a PBS. The signal beam is spatially filtered and expanded before passing through a $\lambda/4$ plate, which turns the linearly polarized beam into a circularly polarized one as it can be well-maintained in propagating through the MMF. Then, the beam is incident on a DMD (V-9501, VIALUX, Germany). Through the Lee hologram technique and a 4f system (L2, iris and L3), phase-only modulation is applied. The modulated beam is then reflected by a long-pass dichroic mirror (DM, DMLP550, Thorlabs, USA) and coupled by a $20\times$ objective (RMS20X, Olympus, Japan) into an MMF (SUH105, 0.22 NA, core diameter $105\ \mu\text{m}$, Xinrui, China).

Scheme 1 employs a transmission-mode pre-calibration module for off-axis interferometry of TM, as off-axis TM is the golden standard and is also permissible within our experimental setup due to the long coherence length of the CW laser. It is noted that the developed non-interferometric TM retrieval in Chapter 4 is preferred when interferometry is challenging (*e.g.*, using a pulsed laser) but less practicable than the off-axis TM for fluorescence imaging. During the phase of pre-calibration, the transmitted speckled beam is collected by obj3 that interferes with a reference beam by the BS for off-axis TM measurement. During the phase of imaging, the pre-calibration module is removed, and slide samples will be placed at the distal tip of

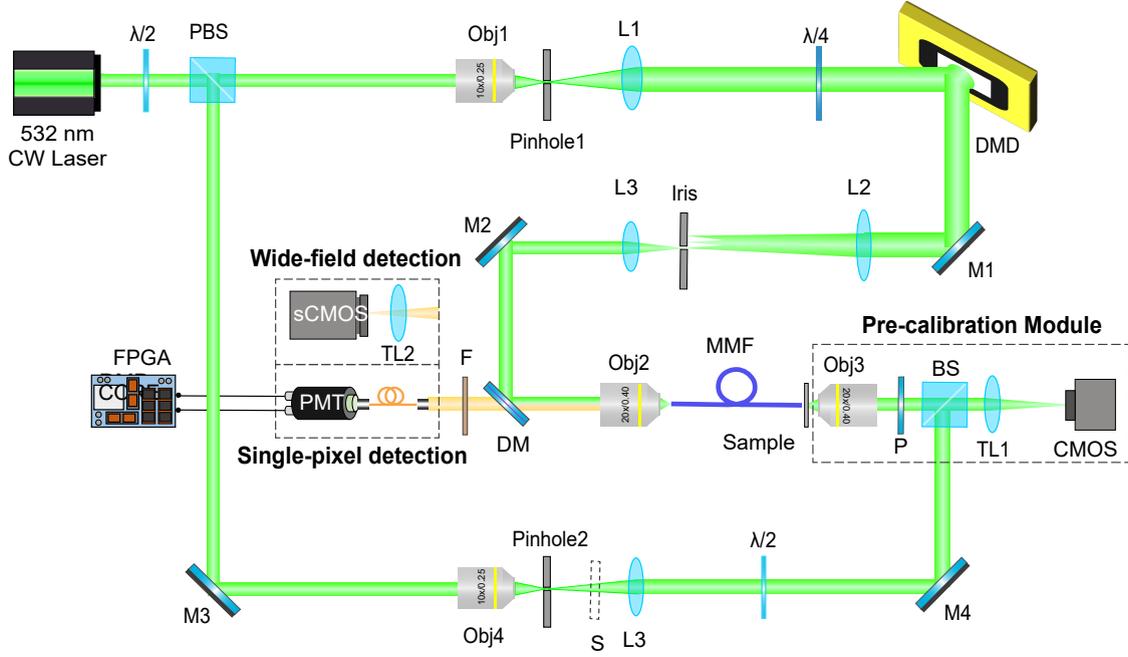


Figure 5.1: Experimental configuration of MMF fluorescence imaging with two schemes. Scheme 1, point-scanning imaging with transmission-mode calibration method and a single-pixel fluorescence detection module; Scheme 2, object reconstruction with reflection-mode calibration method by exploiting a wide-field fluorescence detection module. CMOS, complementary metal-oxide-semiconductor camera; DMD, digital mirror device; DM, dichroic mirror; F, filter; FPGA, field-programmable gated array; L, lens; M, mirror; MMF, multimode fiber; BS, beam splitter; Obj, objective; P, polarizer; PBS, polarizing beam splitter; PMT, photomultiplier tube; S, shutter; sCMOS, scientific CMOS; TL, tube lens; $\lambda/2$, half-wave plate; $\lambda/4$, quarter-wave plate.

the fiber. The fluorescence emission can be backscattered through the same fiber and passes through the DMD and a set of filters including a 533-nm notch filter (NF533-17, Thorlabs, USA) and a long-pass filter (HLPH 550, Hengyang, China) before being detected by a photomultiplier tube (PMT, CH345, Beijing Hamamatsu, China). For Scheme 2, the above single-pixel detection module is replaced with a sCMOS (pco edge 5.5 CLHS, Excelitas, USA) for the wide-field epidetection of fluorescence speckle, and the pre-calibration module is not needed.

The Lee hologram technique [127, 128] is illustrated in Figure 5.2, which uses a binary computed generated hologram (CGH) loaded on the DMD for phase-only or even complex-amplitude modulation of the incident beam. Suppose a complex-amplitude field is $A(x, y) \exp(j\varphi(x, y))$, the binary CGH $t(x, y)$ reads

$$t(x, y) = 0.5 + 0.5 \operatorname{sgn}[\cos(2\pi(\alpha x + \beta y) - \varphi(x, y)) - \cos(\arcsin A(x, y))], \quad (5.1)$$

where α , β is the carrier frequency in the x , y direction respectively. There are a series of

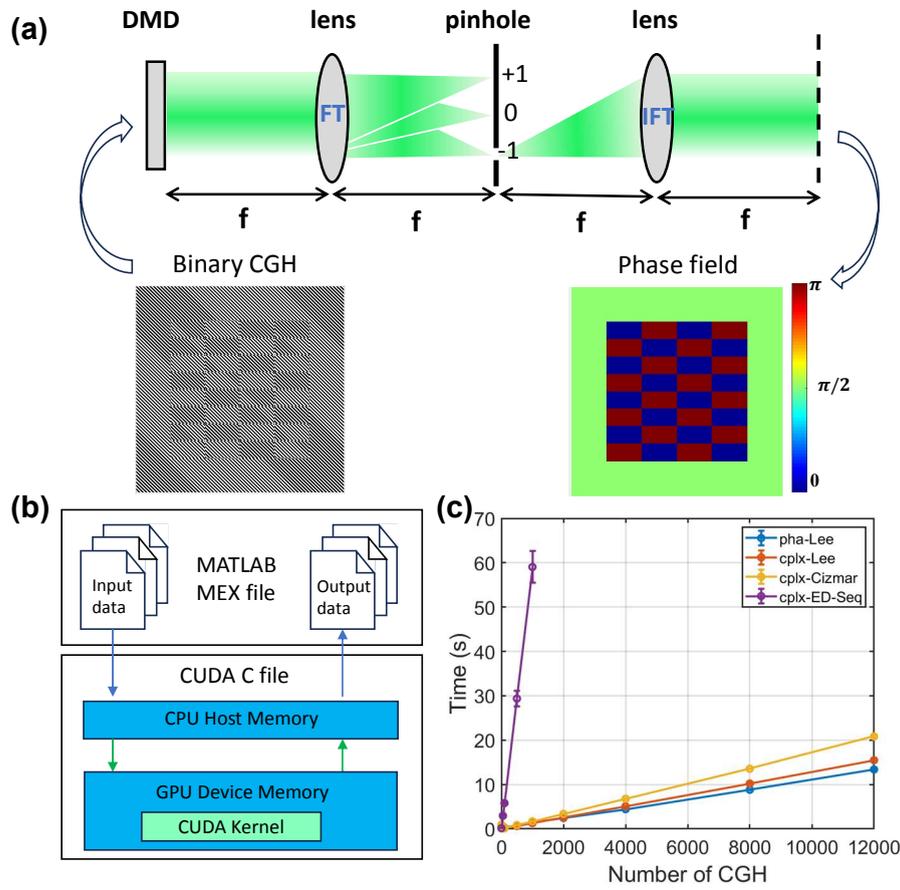


Figure 5.2: Schematic illustration on using DMD for phase (complex-amplitude) modulation and fast CGH generation. (a) DMD and a $4f$ spatial filtering system can produce a desired phase (complex-amplitude) field encoded by a binary CGH. (b) Data flow of generating CGHs rapidly through MEX-CUDA programming. (c) Speed curves of generating CGHs for several different MEX functions, where the error bars denote 5 repeated tests.

spectrum orders of the CGH in the spectrum plane, in which the -1st order is selected for encoding the desired field. For phase-only modulation, set $A(x,y) = 1$. Usually, phase-only modulation is more popular as it has both good modulation efficiency and energy efficiency. For complex-amplitude modulation, the original Lee method suffers problems like low energy efficiency (as a very small number of pixels of DMD are on) and also introduces boundary artifacts due to the direct amplitude threshold in Eq. (5.1). Hence, there are several improved versions for complex-field modulation with higher energy efficiency and better quality, such as Prof. T. Čižmár's method [187] and the error diffusion method introduced by Hu et al [188]. Since fast generation of CGH is highly desired in performing point-scanning imaging once the TM is available, we use MEX-CUDA programming (see Figure 5.2b) to produce the modulation CGHs, which usually spends slightly more than 10 s for producing 10k CGHs, as shown in Figure 5.2c.

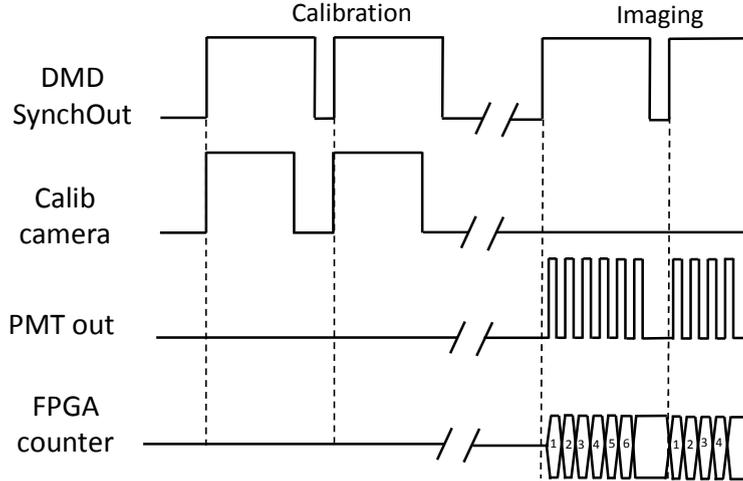


Figure 5.3: Timing diagram of optical modulation and imaging devices in MMF fluorescence imaging during calibration and imaging phases.

Since there are many devices required for the calibration of MMF and fluorescent imaging, suitable triggering and synchronization are very important for coordinating the hardware to work in order. In the experiment, we set the DMD to work in the 'master' mode, which produces a sync out signal that is used for triggering other devices. As depicted in Figure 5.3, in the calibration phase, an image is acquired by the CMOS camera each time the DMD refreshes a pattern; in the imaging phase, the digital PMT outputs TTL signals as the DMD generating focus excites fluorescence, which is processed by a FPGA board (MZ7X MZ7XA, Milianke, China) for photon counting.

5.2.2 Fluorescence-based TM recovery

In Scheme 2, inspired by Refs. [189, 190], we fully exploit the epi-detected fluorescence speckle for the online calibration of MMF in reflection mode by constructing a fluorescence-based RVITM. The principle is illustrated in Figure 5.4 with conceptual simulation results. The fluorescence speckle collected by the sCMOS results from the fluorescence emitted by the targets at the distal end of MMF, which is then backscattered through the MMF for epi-detection. Each fluorescent target could produce a unique speckle pattern on the sCMOS, which is referred as a "fingerprint" (FP). The epi-detected fluorescence speckle $I_p^{\text{fluro}}(x, y)$ is actually a linear superposition of the FP $w_k(x, y)$ from each individual target k , weighted by the target intensity $h_p(k)$ at the p th measurement:

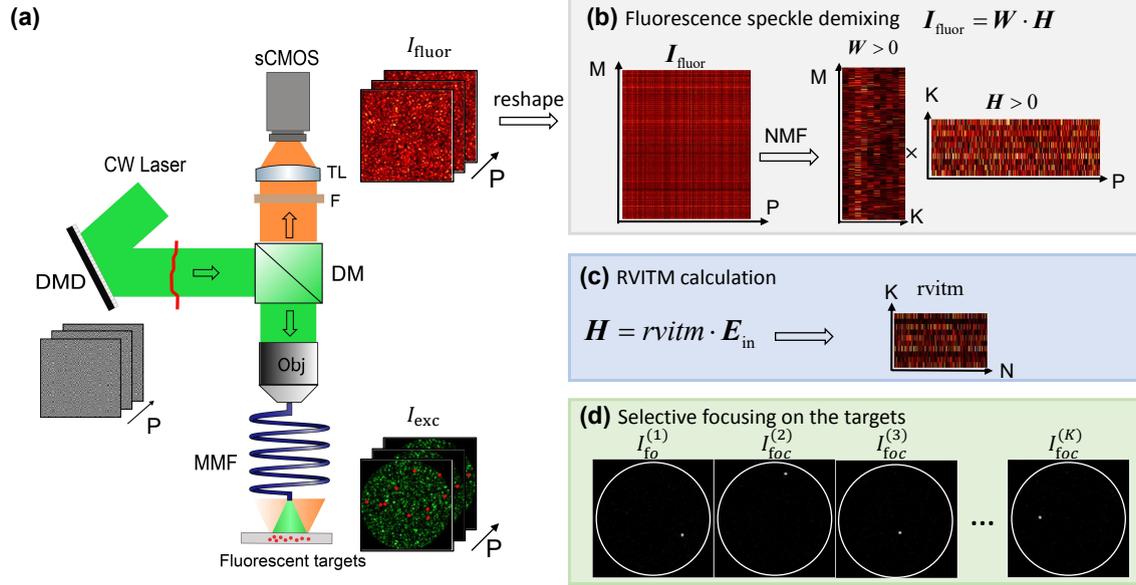


Figure 5.4: Schematic of reflection-mode calibration of MMF and the proposed fluorescence-based RVITM. (a) Illustration of the epi-detection of fluorescence speckle I_{fluor} backscattered through an MMF. (b) Illustration of fluorescence speckle demixing via non-negative matrix factorization. (c) Fluorescence-based RVITM can be calculated with the recovered target intensities under a series of measurements. (d) Selective focusing on different targets can be achieved with the RVITM.

$$I_p^{fluoro}(x, y) = \sum_{k=1}^K w_k(x, y) h_p(k), p = 1, \dots, P, \quad (5.2)$$

where K and P is the total number of targets and measurements, respectively. As such, by unmixing the acquired fluorescence speckle matrix I_{fluor} , it is possible to retrieve a matrix of FPs W and a matrix of target intensities H at different measurements.

With the information of target intensity known, we propose to construct a fluorescence-based RVITM, which relates the n th input $E_p^{in}(n)$ to the k th target intensity $h_p(k)$:

$$h_p(k) = \sum_{n=1}^N rvitm(k, n) \cdot E_p^{in}(n), k = 1, \dots, K, \quad (5.3)$$

where N is the total number of input modes. The principle of RVITM is as follows [149]: two binary modulation basis based on Hadamard matrix H_{had} are constructed as: $H_1 = (\mathbf{1} + H_{had})/2$, $H_2 = (\mathbf{1} - H_{had})/2$, by modulating the input light with H_1 and H_2 , corresponding intensity outputs I_p^k ($k = 1, \dots, K; p = 1, \dots, P$) are obtained,

$$\begin{bmatrix} I_1^1 & \cdots & I_P^1 \\ \vdots & \ddots & \vdots \\ I_1^K & \cdots & I_P^K \end{bmatrix} = r_{vitm} \cdot [\mathbf{H}_1, \mathbf{H}_2]. \quad (5.4)$$

The RVITM can be calculated by

$$r_{vitm} = \frac{1}{2N} \cdot \begin{bmatrix} 2I_1^1 - I_1^1 & \cdots & 2I_P^1 - I_1^1 \\ \vdots & \ddots & \vdots \\ 2I_1^K - I_1^1 & \cdots & 2I_P^K - I_1^1 \end{bmatrix} \cdot [\mathbf{H}_{\text{had}}, -\mathbf{H}_{\text{had}}]. \quad (5.5)$$

The RVITM method is straightforward in implementation and requires only $2N$ measurements. Once it is available, one can selectively focus on the targets through the MMF.

5.2.3 Fluorescent object localization and reconstruction

Due to the natural property of fluorescence speckle demixing, we can only recover the target intensities, while the information of target positions remains unknown. Hence, the above fluorescence-based RVITM is not a complete but a partial TM, and even if we can focus on each individual target, we still do not know its location. So, it is necessary to retrieve the relative distribution of targets. Essentially, the problem is, with only the fluorescence FPs collected from each blinking target at the distal side of MMF, can we localize or reconstruct the targets?

If there is an optical translational "memory effect", the targets close to each other would create highly correlated FPs. Here, we adopt the method from Ref. [190] for the reconstruction of fluorescent object. The first step is pairwise deconvolution between the FPs w_k and w_i ($i, k = 1, \dots, K$) with a total-variation (TV) regularization term:

$$\underset{o_{i,k}}{\operatorname{argmin}} \frac{\mu}{2} \|w_i - o_{i,k} * w_k\|_2^2 + \|o_{i,k}\|_{TV}. \quad (5.6)$$

The idea is by regarding w_k as the system point-spread-function (PSF) and w_i as the image, then the deconvolution will infer an object $o_{i,k}$ that depicts the relative position between target i and k . One can obtain the sub-image of object O_k centered around target k by adding $o_{i,k}$ ($i = 1, \dots, K$) such that $O_k = \sum_{i=1}^K o_{i,k}$. Eventually, all the sub-images can be merged for global reconstruction according to the relative position to target 1, $O_k^{\text{globl}} = \sum_{k=1}^K O_k(\mathbf{r} - \mathbf{r}_{k,1})$.

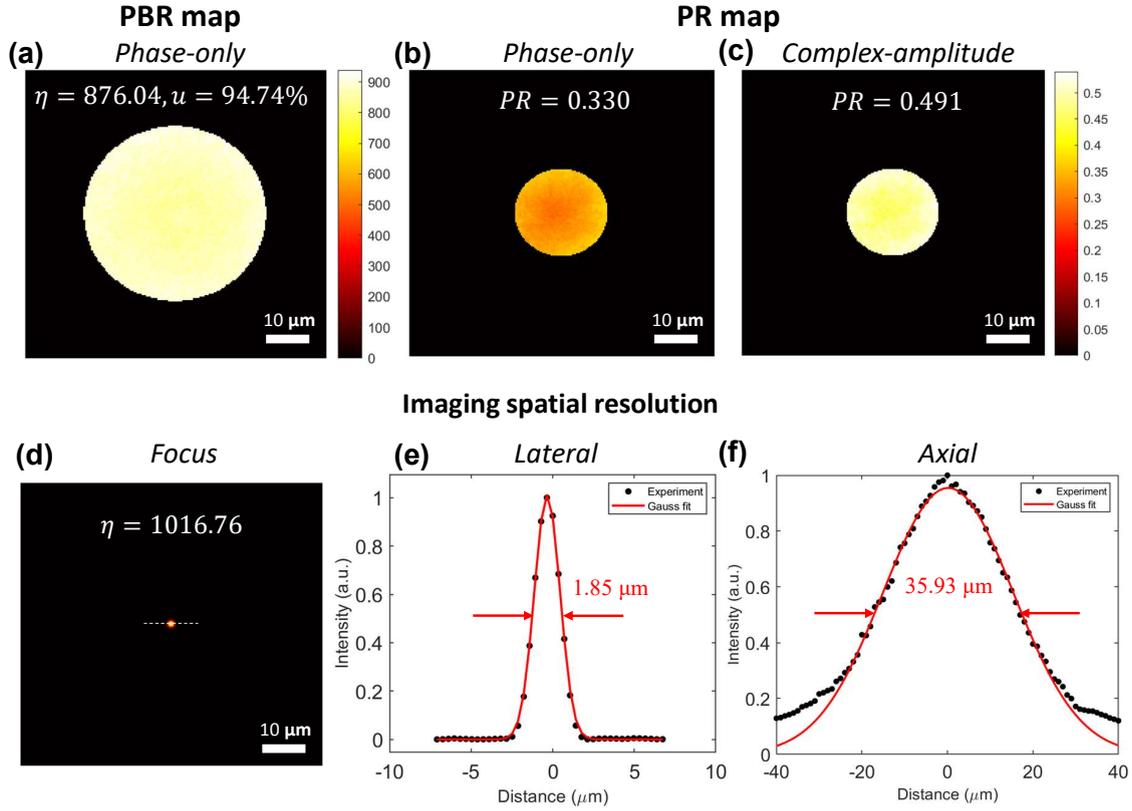


Figure 5.5: Performance characterization for MMF focusing, at a working distance of $150 \mu\text{m}$ from the distal fiber tip. (a) Map of peak-to-background (PBR) under phase-only modulation. Map of power ratio (PR) under (b) phase-only and (c) complex-amplitude modulation, respectively, both calculated via HDR imaging. (d) A focus spot in which a white dashed line denotes the intensity profile plotted in (e), which has a FWHM of $\sim 1.85 \mu\text{m}$ used for quantifying the lateral resolution. (f) The axial resolution of $\sim 35.93 \mu\text{m}$, measured by the FWHM of the foci intensity profile along axial direction.

5.3 Results

5.3.1 Scheme 1: MMF focusing performance characterization

The focusing performance of MMF was first characterized before point-scanning imaging, with the results given in Figure 5.5. Two indicators were employed to show the focusing quality, the PBR and power ratio (PR). The latter denotes the light energy within the focusing region (7×7 square) over the total output light energy. We used a number of $N = 64 \times 64$ modulation units and realized an average PBR of ~ 876 on the working plane of MMF under phase-only modulation (Figure 5.5a). We also compared the focusing efficiency with complex-amplitude modulation. As shown in Figure 5.5 (b-c), the average PR was improved from 0.330 to 0.491, showing a $\sim 49\%$ increase. Note that the calculation of PR shall be conducted by HDR imaging, as the 12-bit CMOS camera we used could not accurately measure the peak and background

at a single exposure setting. We used four sets of exposure time, 5 μs , 21 μs , 198 μs , and 1999 μs , and calculated the HDR results according to the maximum pixel values after exposure compensation. A PR value of at least 30% can be used for high-quality point-scanning imaging. In addition, the spatial resolution of MMF imaging was also quantified, as shown in Figure 5.5 (d-f). The lateral resolution was measured by the FWHM of the intensity profile of a focus spot, and a typical value was 1.85 μm . The axial resolution was 35.93 μm , determined by the FWHM of the curves of focus intensities in the axial direction of MMF.

5.3.2 Scheme 1: MMF point scanning fluorescence imaging

We first imaged on the sample of 2- μm fluorescence microspheres (Nile red, 532/580, Zhichuan Tech, China) to verify the MMF point-scanning imaging performance. As shown in Figure 5.6 (a-b), the fluorescence image matches very well with bright-field image by the transmission camera. Taking a closer look at the ROI represented by a white dashed square, it can be found the two adjacent beads can be almost resolved, as the closest distance of the two beads was larger than the FWHM of a focus spot. The intensity profile of a single focus spot also corresponds well to the lateral resolution we have previously characterized. The imaging FOV depends on the working plane of MMF, which increases with a larger distance from the fiber tip. In our case, the effective FOV is about $120 \times 120 \text{ m}^2$ at a distance of 150 μm from the fiber tip. The imaging speed usually depends on the number of image pixels, the refreshing rate of the DMD, and digital I/O processing rate of the FPGA. We achieved 0.445 frame per second (fps) for producing an image of 150×150 pixels by setting the DMD to run at 10 kHz. Due to the limited board memory of DMD in storing the CGH array, we stitched four adjacent images formed one-by-one into a larger one.

In the experiment, when conducting spot-scanning imaging for the sample sandwiched by a coverslip and a slide, the generated foci by MMF may suffer some aberrations when passing through the coverslip due to a mismatch of refractive index between air and the glass. Hence, we improved the practice of MMF calibration by measuring the TM through the blank area of the slide (Figure 5.7b). In this way, the optical field transformation when propagating through the MMF, air and the coverslip are all included for the TM measurement. As shown in Figure 5.7 (d-f), The new calibration practice gained a better refractive index match and reduced aberration during focusing. Hence, fluorescence imaging on mouse brain slice with less background noise and higher contrast could be achieved.

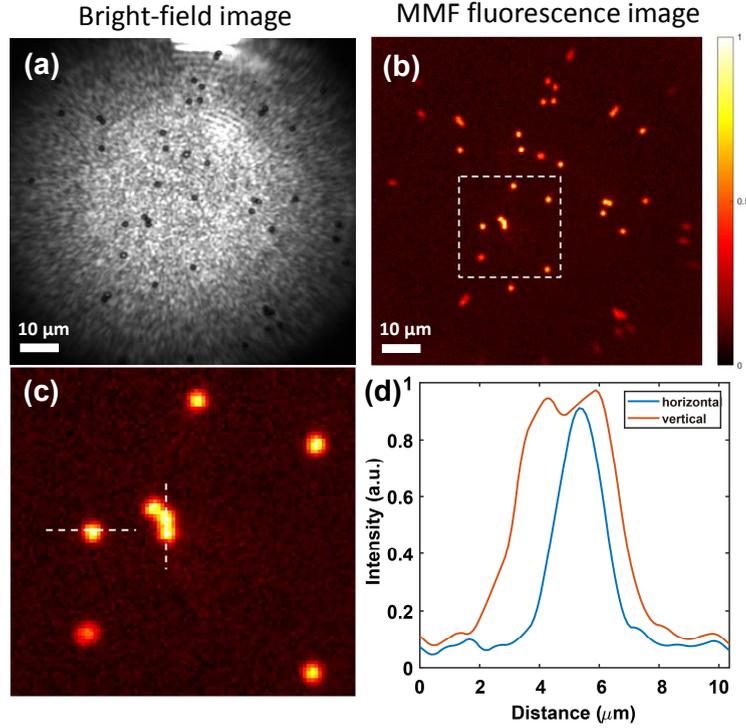


Figure 5.6: MMF point-scanning imaging of 2 μm fluorescence microspheres. (a) Bright-field image of the particles under MMF speckle illumination, used for reference. (b) Fluorescence image of the particles within the same FOV as (a). (c) Zoomed-in view of the region denoted by a white dashed square in (b). (d) Curves of intensity profiles for the single particle (horizontal) and the two adjacent particles (vertical) shown in (c).

5.3.3 Scheme 2: NMF algorithm evaluation and rank estimation

Non-negative matrix factorization (NMF) has been employed for unmixing the fluorescence speckle [189, 190] into a FP matrix $\mathbf{W} \in \mathbb{R}_+^{M \times K}$ and a target intensity matrix $\mathbf{H} \in \mathbb{R}_+^{K \times P}$, which can be formulated as the following minimization problem:

$$\min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \|\mathbf{I}_{\text{fluro}} - \mathbf{W} \cdot \mathbf{H}\|_F, \quad (5.7)$$

where $\|\cdot\|_F$ is the Frobenius norm. In order to achieve the best effects of demixing, we have evaluated the performances of representative Frobenius norm based NMF algorithms using an open-source NMF library [191]. Basically, three types of algorithms, including multiplicate updates (MU), projected gradient descent (PGD) and alternative least squares (ALS), were used for comparison. The root-mean-square residual (RMSR) described by $\delta = \frac{\|\mathbf{I}_{\text{fluro}} - \mathbf{W} \cdot \mathbf{H}\|_F}{\sqrt{(M \cdot P)}}$, and the focusing PBR with the constructed RVITM based on NMF results, were both employed as the evaluation metrics. As seen in Figure 5.8 (b-c), ACC HALS converges to the lowest RMSR level and reaches the highest focusing PBR averagely with the least running time. Also note that

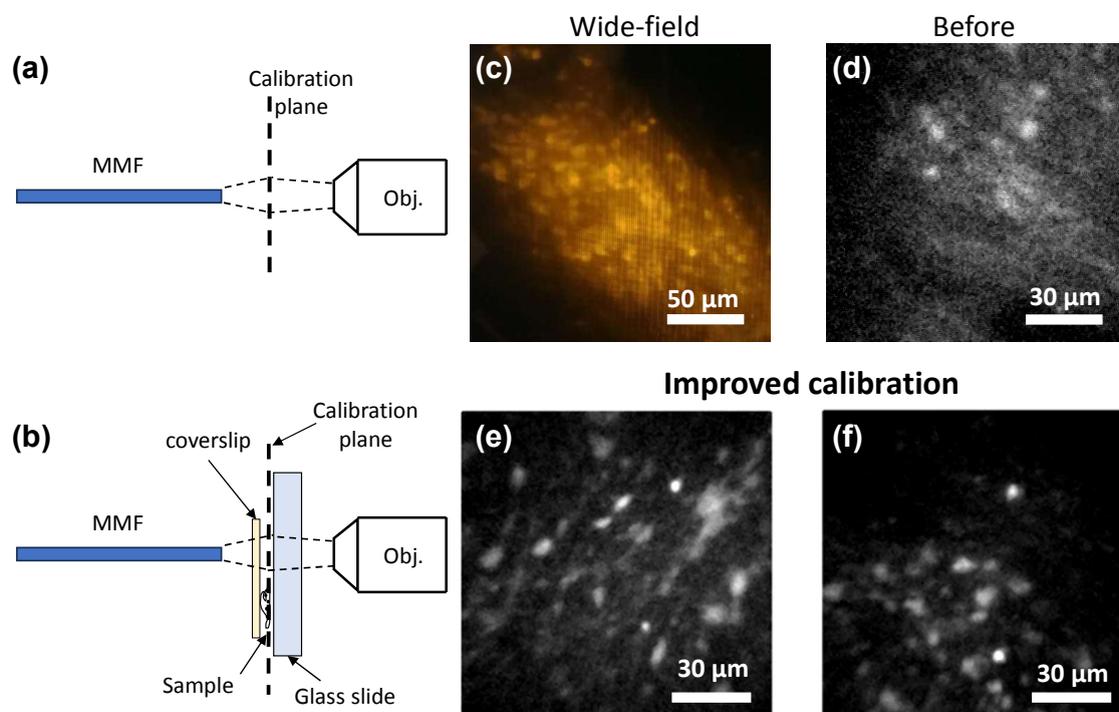


Figure 5.7: Improved practice for MMF calibration and the imaging results on a fluorescent labeled mouse brain slice before and after that. (a) Previous calibration for MMF in air. (b) Current calibration for both fiber and slide (in blank area). (c) A sample of brain slice imaged with a wide-field fluorescence microscope. (d) Previous imaging result. (e-f) Improved imaging results with the new calibration practice.

the results of ALS are not plotted for the sake of clarity as it tends to produce unstable results sometimes. Under a fixed running time of 60 s, we also compare the mean distribution of PBRs of different targets over 5 tests, and ACC HALS again achieves the best performance.

After confirming ACC HALS as the optimum NMF algorithm, we also need to estimate the rank of NMF, which shall correspond to the number of individual targets K of the sample. However, it remains unknown and cannot be used as a Priori knowledge. Here, we show that the best rank for NMF can be collectively determined by comparing the levels of RMSR and PBR under a series of rank estimations. As an example, in Figure 5.9, for the case of 12 targets, only at the rank value of 12 that the lowest level of RMSR and the highest level of focusing PBR are achieved over 5 repeated tests. The above strategy provides clear guidance on the rank estimation for NMF in real experiments.

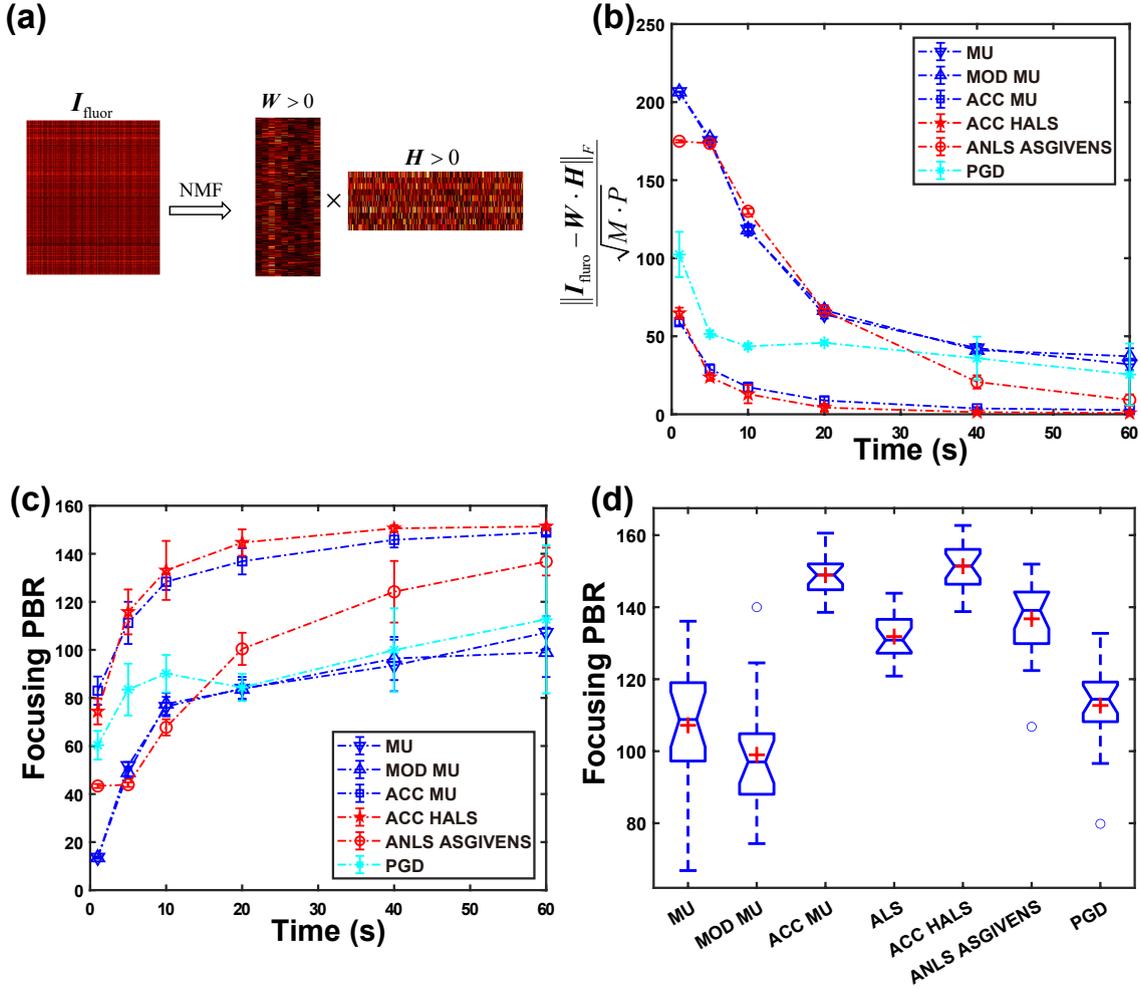


Figure 5.8: Algorithm evaluations for NMF with a rank value of 20. (a) Illustration of fluorescence speckle demixing based on NMF. (b) Root-mean-square residual versus running time for representative NMF algorithms. (c) Average focusing PBR achieved by the RVITMs constructed with representative NMF algorithms versus running time. (d) Boxplot of focusing PBR achieved by the RVITMs constructed with different NMF algorithms at a running time of 60 s. Note the results in (b-d) were all repeated for 5 tests.

5.3.4 Scheme 2: simulations of object reconstruction

Under the assumption of memory effect, we simulated the reconstruction of object at the distal end of MMF based on the retrieved fluorescence FPs by NMF. A simple humanoid fluorescent pattern made of 61 discrete targets is used as the ground truth (Figure 5.10d). After NMF, the columns of matrix \mathbf{W} correspond to the FPs (Figure 5.10a). Let the first FP (#1) be the PSF and the other FPs be the images, a total of 61 deconvolution results can be obtained. The recovered objects are pre-processed first (lowpass filtering and thresholding) before being summed to form a sub-image of the object O_1 , as shown in Figure 5.10b. In this way, the other sub-images can be obtained sequentially. Eventually, all the sub-images are merged into a global

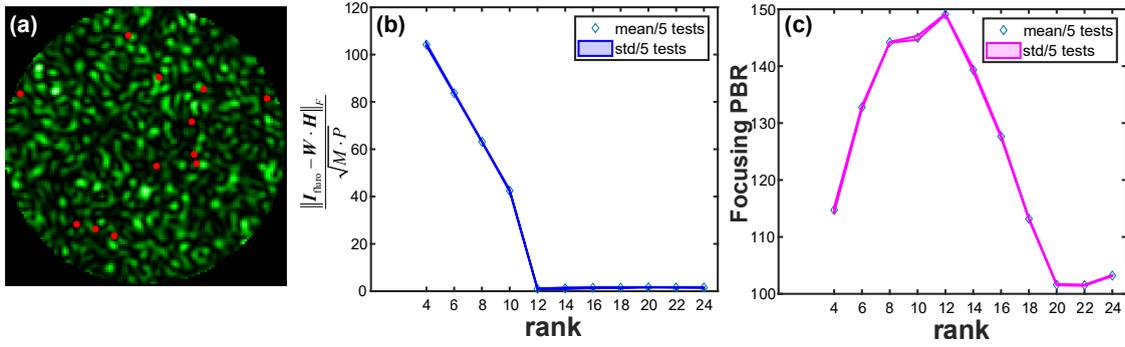


Figure 5.9: Simulations of rank estimation for NMF. (a) Ground truth of 12 fluorescence targets. (b) Root-mean-square residual of NMF versus different values of rank estimation. (c) Focusing PBR achieved by the RVITMs constructed with NMF versus different ranks. Note results in (b-c) were repeated for 5 tests.

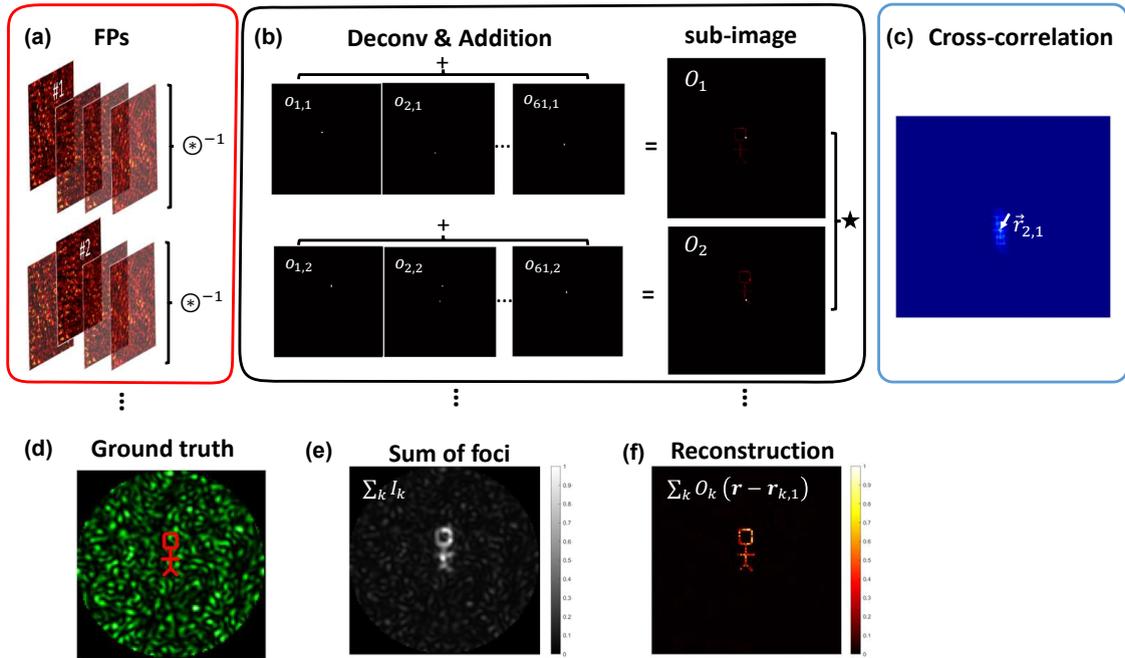


Figure 5.10: Simulations of fluorescence object reconstruction based on fingerprints. (a) Fingerprints of each individual target. (b) Deconvolution over FPs gives relative positions among the targets, which can be summed to form a sub-image centered around a target. (c) The cross-correlation map of the sub-images reveals the relative position of each target with respect to target 1, used for merging all the sub-images for (f) final reconstruction. (d) Ground truth. (e) The sum of foci achieved by the fluorescence-based RVITM. (f) Reconstruction.

reconstruction (Figure 5.10f) according to their relative positions with respect to the first sub-image, which is determined by the maximum position in a cross-correlation map (Figure 5.10c). The reconstruction correlates very well with the ground truth, showing the effectiveness of the above method based on the pairwise deconvolution of fingerprints. Besides that, the sum of foci formed through the MMF in Figure 5.10e also confirms that the proposed fluorescence-based RVITM allows controllable focusing of all the targets.

5.3.5 Scheme 2: preliminary experimental results

Preliminary experiments were also taken to verify the reflection-mode calibration method for MMF using epi-detected fluorescence speckle. Figure 5.11a shows the wide-field fluorescence detection module at the proximal side of MMF, in which the set of imaging filters ensures the maximum transmission of fluorescence while filtering out the reflected excitation light as much as possible. In the experiment, we used fluorescent sample of 2- μm particles, at a distance of 200 μm from the fiber tip, as shown in Figure 5.11b acquired under white light illumination. Due to the small fiber core size and limited NA, the collected fluorescence by the MMF was very weak, and we had to increase the exposure time of sCMOS to be 1s for capturing a bright speckle pattern (Figure 5.11c) with enough structures. The center square region was cropped as the input data, which was high-pass filtered (Figure 5.11d) for removing the background fluorescence envelope and detection noise. The filtered data (Figure 5.11e) was used as the input data for fluorescence speckle unmixing based on NMF.

Experimentally, the number of modulation units was set to be 16×16 , and the acquisition time of fluorescence speckles was $\sim 614\text{s}$. We used ACC HALS algorithm for conducting NMF with a rank estimation of 40. With the NMF results, both the FPs that correspond to 40 discrete targets and the intensities of 40 targets under different measurements were recovered, as shown in Figure 5.11(f-g). It can be seen the FPs are darker but with higher contrast than a typical fluorescence speckle in Figure 5.11e, as the latter is a weighted supposition of all FPs and could be averaged out to some degree. The intensity profiles for each target show temporal fluctuations, corresponding to the random speckle illuminations by the binary Hadamard basis. However, with the constructed RVITM, focusing through MMF was failed in the experiment, with only speckle patterns (Figure 5.11h), as monitored by the calibration camera.

There are several reasons that may account for the failure. First, the long-pass filter may not be optimum for transmitting the 580-nm fluorescence emitted by the Nile red microspheres. The low transmission could lead to a longer exposure time of sCMOS for an accumulation of enough fluorescence flux. Second, the small fiber core (105 μm) and limited NA (0.22) reduced the efficiency of fluorescence collection, which also required a longer exposure time of sCMOS. Third, for the sake of reducing the calibration time, a small number of modulation units (16×16) was employed, which can also result in poor focusing ability (theoretically, $\eta = 82.2$ under binary phase modulation). Even so, it still took more than 10 minutes for the fluorescence data

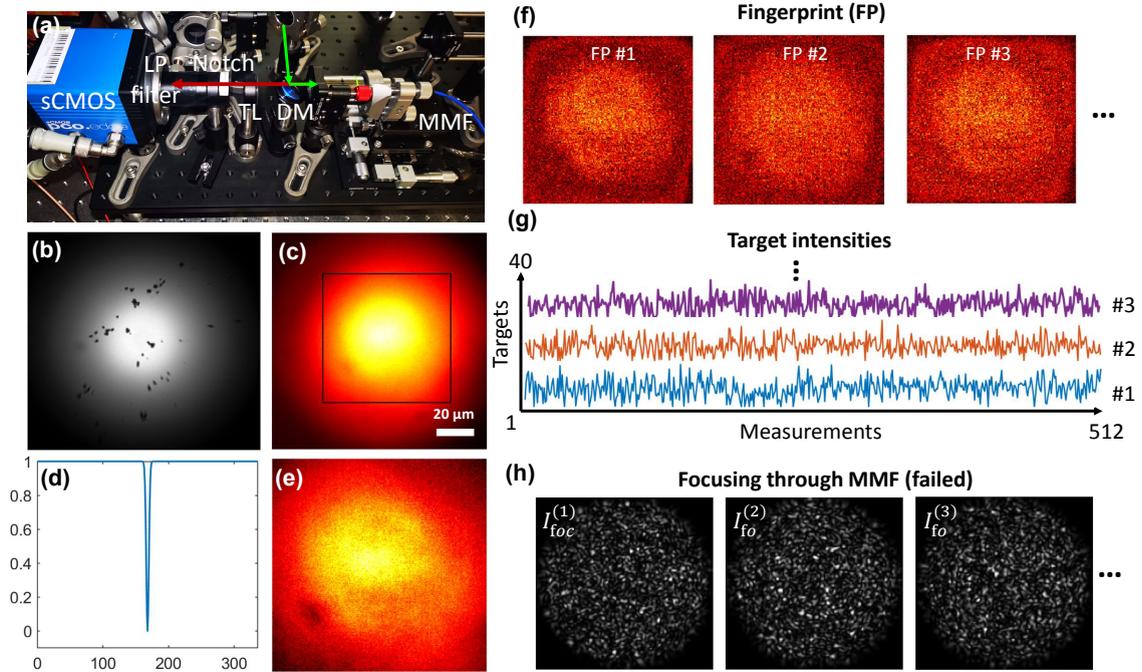


Figure 5.11: Preliminary experimental results on fluorescence-based RVITM. (a) Module of fluorescence filtering and epi-detection with an sCMOS camera. (b) 2- μm fluorescence particles placed at the distal fiber end. (c) Fluorescence speckle captured at an exposure time of 1 s, in which the area denoted by a black solid box is cropped for input data. (d) Profile of high-pass filter for fluorescence data filtering. (e) Filtered fluorescence speckle. (f) FPs of each target obtained by NMF. (g) Recovered target intensities under a series of measurements. (h) Focusing on the targets (failed) with the constructed RVITM.

acquisition, which might cause decorrelations of the MMF, leading to an invalid measurement of TM and failed focusing on the targets.

5.4 Discussion

For MMF-based fluorescence imaging via conventional transmission-mode calibration method (Scheme 1), we confirmed the feasibility of point-scanning imaging of MMF and improved the practice of calibration to obtain fluorescence images with less background noise and better contrast. However, these are still very basic in the field, with modest results achieved in terms of the imaging resolution, quality, speed and application impacts. In the next step, we will replace the digital PMT and FPGA counter with an analog PMT with better sensitivity and a more advanced DAQ card with a higher sampling rate (hundreds of megabits per second). As such, the refreshing rate of the DMD can be set up to 23 kHz, allowing for faster imaging speed and also better fluorescent imaging quality. The spatial resolution of the MMF-based imaging can be improved by fiber facet micromachining technology, such as coating the distal facet with

a scattering layer to improve the NA of the MMF, or using the fluorescence emission difference method for super-resolution imaging [179]. For application, volumetric fluorescence imaging will be conducted on in-vitro samples or small animals *in vivo* in the future, by constructing a free-space propagation operator to complement the measured TM [192] to shift foci axially. Besides, finding an impactful application through technological breakthroughs will be of great interest. For instance, developing MMF-based polarized fluorescence imaging is promising for detecting the structures of biological molecules *in vivo* based on the fluorescent dipole orientations.

For the proposed fluorescence-based RVITM as a reflection-mode calibration method for MMF, there are also several comments on that. First, we apply the method from Ref. [189] to focusing through MMF by measuring an RVITM in the reflection mode. Compared to a complex TM, the RVITM significantly reduces the number of measurements (*i.e.*, $P = 2N$) that is necessary for the calibration of MMF, as it is vulnerable to any perturbations. According to simulations, one can easily achieve more than 90% of the theoretical PBR with the constructed RVITM for focusing. Unfortunately, it failed in focusing through MMF in our preliminary experiments. With the reason analyzed in Section 5.3.5, we will update the fluorescence emission filter for the best transmission efficiency and use an MMF with larger NA, which allows less exposure time of sCMOS and a larger number of modulation units for better focusing. The effective calibration of MMF in the reflection-mode will be feasible after some improvements. Second, we have demonstrated the fluorescent object reconstruction based on FPs with the assumption of translational memory effect, which, however, does not exist in regular MMF with circular core, but exist in square-core MMF [193]. Hence, new methods of reconstructing or localizing the targets that do not rely on the correlation of FPs will be explored. For example, we have been studying a scheme of fluorescence-based reflection matrix, which is possible to localize the targets with singular value decomposition (SVD). On the other hand, FP-based reconstruction is applicable to square-core MMF or regular multicore fiber (MCF) for fluorescence imaging, as MCF shows strong translational memory effect and grows to be an important solution for high-resolution endoscopic imaging with wavefront shaping techniques.

5.5 Conclusion

In this chapter, we have conducted a pilot study on MMF-based fluorescence imaging, including point-scanning imaging with conventional transmission-mode calibration (Scheme 1), as well

as focusing and imaging through MMF based on reflection-mode calibration (Scheme 2). For Scheme 1, the focusing and point-scanning imaging performance of MMF have been confirmed, and we improve the practice of MMF calibration for better imaging contrast. For Scheme 2, numerical simulations show the feasibility of the proposed fluorescence-based RVITM for online calibration of MMF, which, however, has yet been verified in preliminary experiment. Further improvements on experimental design and fluorescence object reconstruction method in the absence of translational memory effect are needed in the next phase of study.

Chapter 6

Summary and Future Work

6.1 Summary and contributions

This thesis presents a comprehensive study on deep-tissue optical imaging and patterned light delivery by leveraging computational approaches to enhance the performances of deep-tissue photoacoustic imaging and MMF-based focusing and fluorescence imaging. The research work encompasses the design and implementation of advanced computational methods to address the relevant challenges, such as resolution enhancement of PAM under low SNR, precise light delivery in the existence of scattering, and high-efficiency retrieval of TM from only intensity measurements, *etc.*. The major contributions of each chapter in the thesis can be summarized as follows.

Chapter 2 (Part I) introduces a novel deep learning image transformation method to improve the resolution of blurry images acquired by AR-PAM while maintaining its penetration depth. An improved GAN was trained on co-registered OR- and AR-PAM images from in-vivo mouse ear and generalized to other unseen image data, demonstrating superior performances of image deblurring and resolution enhancement in contrast to blind deconvolution. The method significantly improved a mouse deep-brain image and retained high imaging resolution at tissue depths beyond the light diffusion limit, promising high-resolution monitoring of hemodynamics through the compact skulls of living small animals.

Chapter 3 (Part II) presents two efficient wavefront optimization methods for controlling patterned light delivery through MMF. Specifically, a naturally gradient ascent-based wavefront

shaping strategy optimized with a vector cosine similarity measure was proposed for high-fidelity pattern projection. The proposed scheme showed fast searching speed and robust performance in noisy and perturbed environments. It might be attractive for deep-tissue applications requiring precise optical delivery, such as photon therapy and optogenetics. Apart from focusing, a TM-based alternating projection phase optimization method was also developed for fast and arbitrary glare suppression in the output of an MMF. Two speckle-domain constraints, ER and HIO, were adopted for comparative study. The HIO constraint, in particular, achieved rapid convergence and significant suppression performance, with potential applications for speckle-related imaging or sensing in complex environments.

Chapter 4 (Part II) develops a non-interferometric TM retrieval method with optimum efficiency via nonconvex optimization, *i.e.*, RAF 2-1, for the retrieval of the TM of an MMF. Compared to existing TM retrieval algorithms, RAF 2-1 needs less running time or sampling ratio to reach theoretical focusing efficiency. We validated the performance of RAF 2-1 experimentally, which approached the golden standard off-axis holography method. With optimum efficiency, fast execution, and a reference-less setup, it holds promise for deep-tissue imaging, manipulation, and data communication with the usage of MMFs.

Chapter 5 (Part III) explores MMF-based fluorescence imaging by exploiting transmission-mode calibration (Scheme 1) and the proposed reflection-mode calibration (Scheme 2). Scheme 1 confirmed the MMF's point-scanning-imaging capability and improved a MMF calibration practice for imaging specimen sandwiched by slides. Scheme 2 introduced a fluorescence-based RVITM method for online calibration of MMF during fluorescence imaging, which was feasible in numerical simulations but required further experimental validation and optimization.

In summary, different parts of the thesis cohesively advance the field of deep-tissue optical imaging and focusing by integrating computational methods with optical technologies to address the inherent contradiction between imaging resolution and penetration depth. Part I utilizes deep learning image transformation for postprocessing blurry yet "seeing-deeper" AR-PAM images, which virtually alleviates the influence of optical scattering and realizes high-resolution deep-penetrating PAM. Then, it comes to the imaging technology of MMF-based endoscopy that physically bypass optical scattering in an ultra-minimal invasive manner. Part II

develops high-efficiency wavefront optimization and TM retrieval methods for controlling patterned light delivery through MMF. With such light control capabilities, Part III explores MMF-based fluorescence endomicroscopy by using transmission-mode and the proposed reflection-mode calibration methods. Our schemes demonstrate endoscopic fluorescence imaging at a lateral resolution of $\sim 1.85 \mu\text{m}$, while still needing further experimental validation and optimization. Among all the research parts, it should be noted that Part I is not isolated from the rest, since the wavefront modulation and MMF imaging methods developed in Part II and III may also be applied to physically enhance in-tissue optical focusing and AR-PAM imaging quality. Besides, the deep learning method used in Part I may also play a role in wavefront optimization [194] and TM reconstruction [186]. Overall, these studies have collectively enhanced the understanding and pushed the boundaries of deep tissue PA imaging and MMF-based endoscopic imaging. Our work may pave the way for high-resolution optical applications in biomedicine, including imaging, manipulation, precise phototherapy and optogenetics *etc.*

6.2 Future work

Many deep-tissue optical imaging approaches have been developed in the past decades, aiming to tackle the challenge of superficial imaging depth for optical imaging, such as NIR-II imaging, diffuse optical tomography, multiphoton microscopy *etc.*, as mentioned in Section 1.3. These imaging technologies usually utilize NIR laser source to reduce light scattering and have gained wide applications in biomedical diagnosis and therapy. Compared with the proposed network-enhanced PAM imaging and MMF-based fluorescence endomicroscopy, we emphasize that each method has its own advantage and usage scenario. For example, PAM is especially useful for label-free imaging of hemodynamics and blood oxygen saturation with very high resolution, and MMF-based imaging breaks the limit of penetration depth with the least invasiveness, which holds promise for deep-brain neuroimaging and optical biopsy *in vivo*. Importantly, our work has pushed the boundaries of deep-tissue optical imaging in terms of improved SNR and resolution in AR-PAM and signal stability in MMF-based fluorescence endomicroscopy. The PSNR for AR-PAM imaging on mouse ears has improved averagely from 16.77 to 20.02 dB, while the CNR for hair phantom imaged at a depth of $1700 \mu\text{m}$ has increased from 10.6 to 17.1 dB. The lateral resolution of AR-PAM also sees a significant enhancement from ~ 54.0 to $\sim 5.1 \mu\text{m}$ with our image transformation network, making it comparable to OR-PAM.

For MMF-based endoscopy, the sensitivity to external disturbance—common in dynamic tissues with *in-vivo* physiological motions—has always been a major challenge and not been solved perfectly. The thesis proposes a fluorescence-based RVITM scheme for MMF online calibration from reflective signals, which may help retain the imaging robustness of MMF during endoscopic fluorescence imaging. Moreover, our deep-tissue optical imaging technologies including network-enhanced AR-PAM and single MMF-based fluorescence endoscopy, have potential for low-cost application in preclinical or clinical scenarios. The cost can be reduced by using a relatively inexpensive LED source for signal excitation in AR-PAM and employing a single standard MMF as the probe, compared to the conventional pCLE and eCLE systems that require sophisticated imaging probes. Last but not least, it is emphasized that the imaging technologies involved in the thesis are safe for biomedical applications. The pulse energy for OR- and AR-PAM was ~ 80 nJ and ~ 200 nJ respectively, far less than the damage threshold of biological tissues. The single MMF-based endoscopy has ultrafine probe size (~ 100 μm). With the use of a biocompatible indwelling catheter, it will be more clinically practicable for long-term monitoring.

For the network-enhanced deep-tissue PAM, it takes ~ 3.75 s to process a large image sized 2000×2000 when running on a PC with an Intel Xeon CPU, a NVIDIA 3070 GPU and 64 GB RAM. With more powerful computation resources and further network model compression, we believe our technology may allow real-time imaging. Also, it will be more impactful to combine the computational method with hardware, such as optimizing and deploying the network model into a portable AR-PAM device. Moreover, our deep learning PAM image transformation method may also be combined with other medical imaging techniques such as magnetic resonance imaging or X-ray computed tomography for multi-modality diagnosis of deep tissues. After image infusion, it will provide more comprehensive information including soft tissue and bone structures, hemodynamics etc. Still, further improvements can be made in the following aspects. First, more advanced network architecture like vision transformer [195] can be employed to enhance the imaging transformation results. Then, the training dataset of the GAN model can be expanded to improve its generalization and transformation capabilities. Given the laborious animal experiments for acquiring co-registered living PAM data, utilizing open-source OR-PAM data [196] with digital blurring operations could be a good choice. Further, attention could also be paid to physically improving the SNR and resolution of AR-PAM by applying the wavefront modulation and MMF imaging methods developed in Part II and III for in-tissue optical focusing with high resolution. More recently, acoustic-feedback

wavefront-adapted PAM [197] was proposed that corrected tissue-induced aberrations with the feedback of PA signals to enhance PAM imaging. The speed of wavefront optimization was shown to be enough even for imaging in living specimens thanks to a low-order Zernike polynomials model. Towards this direction, it may lead to high-resolution PAM in deep tissues, for example, through an intact brain.

For MMF-based focusing and imaging based on wavefront optimization or TM method, fibers of different types and lengths have been tested: for long-distance pattern projection through a MMF, a 15-m-long and unfixed fiber was utilized; for the glare suppression experiment, a regular 1-m-long MMF was used instead; for the reflection-mode MMF fluorescence imaging, a short MMF with larger NA would be preferred. We emphasize different fibers have their own usage scenarios, while the WFS methods for controlling light transmission through MMF remain generally applicable. Future directions for controlling MMF light delivery are discussed as follows. Firstly, since the time efficiency of iterative WFS is identified as a major limitation for pattern projection through MMF, efforts will be taken to enhance the system engineering, such as adopting a faster feedback mechanism and accelerating the data transfer and processing by a customized FPGA circuit. These advancements will be crucial for real-time application of precise light delivery in living tissues. Next, we will address the challenges faced by glare suppression through MMF, including the complexity of acquiring TM with an interferometric setup and the discrepancies between experimental and simulated results. Solutions may involve adopting a reference-less TM retrieval scheme and improving the accuracies of TM measurement and DMD-based phase modulation. In addition, practical applications of glare suppression, such as MMF-based customizable speckle optical tweezer, may also be explored. Then, the limitations of the experimental setup and the problem of phase ambiguity among different TM rows in the study of TM retrieval will be addressed. These include measuring the TM at multiple working planes by using a movable objective lens for speckle data acquisition, and conducting phase correction [17, 198] before applying the TM retrieval method to MMF volumetric focusing and imaging. Lastly, deep learning, as used in Part I, may also be applied for TM retrieval by feeding a network model with calibration patterns and speckle data. This could have potential in greatly reducing the computation time of TM recovery, compared to iterative optimization method. However, the size of the TM that a network model can process may be very limited due to the substantial memory burden required during the training process of TM recovery, as noted in ref. [186].

Currently, our study on MMF-based fluorescence imaging sees limitations in imaging quality

(*e.g.*, resolution and contrast), speed and an instability to external perturbation. Effects may be taken to increase the imaging speed by utilizing DMD for fast line scanning and upgrading the fluorescence detection and data acquisition modules. The MMF point-scanning-imaging resolution can be further improved by applying fiber facet micromachining technology to increase the fiber NA or exploring super-resolution imaging methods. Additionally, 3D volumetric imaging and impactful applications such as MMF-based polarized fluorescence imaging will also be pursued, which is important for observing the microscopic structures of biological molecules *in vivo*. Last but not the least, the proposed fluorescence-based RVITM could be promising for online calibration of MMF during fluorescence imaging but requires further experimental validations. Efforts will focus on optimizing the fluorescence emission filter and using a shorter MMF with larger NA to improve the fluorescence collection efficiency. The development of new targets localization or reconstruction method that does not rely on translational memory effect will also be a priority, potentially through the study of fluorescence-based reflection matrix and SVD technique.

In this thesis, we integrate computational methods and optical technologies to propel the fields of deep-tissue optical imaging and focusing. The proposed future work aims to address existing challenges related to imaging quality and robustness, while also exploring potential avenues. Our studies strive to break the inherent limitation on high-resolution optical imaging at depths in tissues by leveraging PA imaging and MMF-based endoscopy technologies. Through these and ongoing efforts, our investigations may unveil new possibilities towards seeing deeper with greater clarity into biological tissues and even living organisms.

Reference

- [1] V. Ntziachristos, “Going deeper than microscopy: the optical imaging frontier in biology,” *Nature methods* **7**(8), 603–614 (2010).
- [2] M. J. Booth, “Adaptive optical microscopy: the ongoing quest for a perfect image,” *Light: Science & Applications* **3**(4), e165–e165 (2014).
- [3] R. Horstmeyer, H. Ruan, and C. Yang, “Guidestar-assisted wavefront-shaping methods for focusing light into biological tissue,” *Nature photonics* **9**(9), 563–571 (2015).
- [4] S. Yoon, M. Kim, M. Jang, *et al.*, “Deep optical imaging within complex scattering media,” *Nature Reviews Physics* **2**(3), 141–158 (2020).
- [5] J. Oh, C. Lee, G. Song, *et al.*, “Review of endomicroscopic imaging with coherent manipulation of light through an ultrathin probe,” *Journal of Optical Microsystems* **3**(1), 011004–011004 (2023).
- [6] N. Ji, “Adaptive optical fluorescence microscopy,” *Nature methods* **14**(4), 374–380 (2017).
- [7] L. V. Wang and J. Yao, “A practical guide to photoacoustic tomography in the life sciences,” *Nature methods* **13**(8), 627–638 (2016).
- [8] I. M. Vellekoop and A. P. Mosk, “Focusing coherent light through opaque strongly scattering media,” *Optics letters* **32**(16), 2309–2311 (2007).
- [9] I. M. Vellekoop, “Feedback-based wavefront shaping,” *Optics express* **23**(9), 12189–12206 (2015).
- [10] P. Lai, L. Wang, J. W. Tay, *et al.*, “Photoacoustically guided wavefront shaping for enhanced optical focusing in scattering media,” *Nature photonics* **9**(2), 126–132 (2015).
- [11] O. Katz, E. Small, Y. Guan, *et al.*, “Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers,” *Optica* **1**(3), 170–174 (2014).

- [12] I. M. Vellekoop and A. Mosk, "Phase control algorithms for focusing light through turbid media," *Optics communications* **281**(11), 3071–3080 (2008).
- [13] D. B. Conkey, A. N. Brown, A. M. Caravaca-Aguirre, *et al.*, "Genetic algorithm optimization for focusing through turbid media in noisy environments," *Optics express* **20**(5), 4840–4849 (2012).
- [14] Y. Zhao, Q. He, S. Li, *et al.*, "Gradient-assisted focusing light through scattering media," *Optics Letters* **46**(7), 1518–1521 (2021).
- [15] S. Cheng, T. Zhong, C. M. Woo, *et al.*, "Long-distance pattern projection through an unfixed multimode fiber with natural evolution strategy-based wavefront shaping," *Optics express* **30**(18), 32565–32576 (2022).
- [16] S. M. Popoff, G. Lerosey, R. Carminati, *et al.*, "Measuring the transmission matrix in optics: An approach to the study and control of light propagation in disordered media," *Physical review letters* **104**(10), 100601 (2010).
- [17] J. Zhong, Z. Wen, Q. Li, *et al.*, "Efficient reference-less transmission matrix retrieval for a multimode fiber using fast fourier transform," *Advanced Photonics Nexus* **2**(5), 056007–056007 (2023).
- [18] T. Čižmár and K. Dholakia, "Shaping the light transmission through a multimode optical fibre: complex transformation analysis and applications in biophotonics," *Optics express* **19**(20), 18871–18884 (2011).
- [19] M. Plöschner, B. Straka, K. Dholakia, *et al.*, "Gpu accelerated toolbox for real-time beam-shaping in multimode fibres," *Optics express* **22**(3), 2933–2947 (2014).
- [20] Y. Choi, C. Yoon, M. Kim, *et al.*, "Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber," *Physical review letters* **109**(20), 203901 (2012).
- [21] M. Hofer and S. Brasselet, "Manipulating the transmission matrix of scattering media for nonlinear imaging beyond the memory effect," *Optics letters* **44**(9), 2137–2140 (2019).
- [22] P. Jákl, M. Šiler, J. Ježek, *et al.*, "Endoscopic imaging using a multimode optical fibre calibrated with multiple internal references," in *Photonics*, **9**(1), 37, MDPI (2022).
- [23] A. Drémeau, A. Liutkus, D. Martina, *et al.*, "Reference-less measurement of the transmission matrix of a highly scattering material using a dmd and phase retrieval techniques," *Optics express* **23**(9), 11898–11911 (2015).

- [24] G. Huang, D. Wu, J. Luo, *et al.*, “Generalizing the Gerchberg–Saxton algorithm for retrieving complex optical transmission matrices,” *Photonics Research* **9**(1), 34–42 (2020).
- [25] S. Cheng, X. Zhang, T. Zhong, *et al.*, “Nonconvex optimization for optimum retrieval of the transmission matrix of a multimode fiber,” *Advanced Photonics Nexus* **2**(6), 066005–066005 (2023).
- [26] M. J. Booth, “Adaptive optics in microscopy,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **365**(1861), 2829–2843 (2007).
- [27] J. Yao and L. V. Wang, “Photoacoustic microscopy,” *Laser & Photonics Reviews* **7**(5), 758–778 (2013).
- [28] C. Xu, W. Zipfel, J. B. Shear, *et al.*, “Multiphoton fluorescence excitation: new spectral windows for biological nonlinear microscopy,” *Proceedings of the National Academy of Sciences* **93**(20), 10763–10768 (1996).
- [29] *Deep Tissue Fluorescence Microscopy*, 61–91. *Advances in Microscopy and Microanalysis*, Cambridge University Press (2019).
- [30] J. M. Jabbour, M. A. Saldua, J. N. Bixler, *et al.*, “Confocal endomicroscopy: instrumentation and medical applications,” *Annals of Biomedical Engineering* **40**, 378–397 (2012).
- [31] S. S. Chauhan, B. K. A. Dayyeh, Y. M. Bhat, *et al.*, “Confocal laser endomicroscopy,” *Gastrointestinal Endoscopy* **80**(6), 928–938 (2014).
- [32] Y. Li, G. Lu, Q. Zhou, *et al.*, “Advances in endoscopic photoacoustic imaging,” in *Photonics*, **8**(7), 281, MDPI (2021).
- [33] M. J. Gora, M. J. Suter, G. J. Tearney, *et al.*, “Endoscopic optical coherence tomography: technologies and clinical applications,” *Biomedical Optics Express* **8**(5), 2405–2444 (2017).
- [34] H. Cao, T. Čižmár, S. Turtaev, *et al.*, “Controlling light propagation in multimode fibers for imaging, spectroscopy, and beyond,” *Advances in Optics and Photonics* **15**(2), 524–612 (2023).
- [35] C. Shengfu, Z. Tianting, W. Chi Man, *et al.*, “Lensless fiber-optic imaging via coherent light modulation and its applications (invited),” *Laser & Optoelectronics Progress* **61**(6), 0618002 (in Chinese) (2024).

- [36] G. Wu, R. Zhu, Y. Lu, *et al.*, “Optical scanning endoscope via a single multimode optical fiber,” *Opto-Electronic Science* **3**(3), 230041–1 (2024).
- [37] S. Ohayon, A. Caravaca-Aguirre, R. Piestun, *et al.*, “Minimally invasive multimode optical fiber microendoscope for deep brain fluorescence imaging,” *Biomedical optics express* **9**(4), 1492–1509 (2018).
- [38] T. Zhao, T. T. Pham, C. Baker, *et al.*, “Ultrathin, high-speed, all-optical photoacoustic endomicroscopy probe for guiding minimally invasive surgery,” *Biomedical optics express* **13**(8), 4414–4428 (2022).
- [39] A. Cifuentes, T. Pikálek, P. Ondráčková, *et al.*, “Polarization-resolved second-harmonic generation imaging through a multimode fiber,” *Optica* **8**(8), 1065–1074 (2021).
- [40] M. Stibůrek, P. Ondráčková, T. Tučková, *et al.*, “110 μm thin endo-microscope for deep-brain in vivo observations of neuronal connectivity, activity and blood flow dynamics,” *Nature Communications* **14**(1), 1897 (2023).
- [41] Z. Yu, H. Li, and P. Lai, “Wavefront shaping and its application to enhance photoacoustic imaging,” *Applied Sciences* **7**(12), 1320 (2017).
- [42] T. Zhao, M. T. Ma, S. Ourselin, *et al.*, “Video-rate dual-modal photoacoustic and fluorescence imaging through a multimode fibre towards forward-viewing endomicroscopy,” *Photoacoustics* **25**, 100323 (2022).
- [43] L. V. Wang, “Multiscale photoacoustic microscopy and computed tomography,” *Nature photonics* **3**(9), 503–509 (2009).
- [44] K. Maslov, H. F. Zhang, S. Hu, *et al.*, “Optical-resolution photoacoustic microscopy for in vivo imaging of single capillaries,” *Optics letters* **33**(9), 929–931 (2008).
- [45] L. Wang, K. Maslov, J. Yao, *et al.*, “Fast voice-coil scanning optical-resolution photoacoustic microscopy,” *Optics letters* **36**(2), 139–141 (2011).
- [46] Y. Zhou, J. Chen, C. Liu, *et al.*, “Single-shot linear dichroism optical-resolution photoacoustic microscopy,” *Photoacoustics* **16**, 100148 (2019).
- [47] Y. Zhou, S. Liang, M. Li, *et al.*, “Optical-resolution photoacoustic microscopy with ultrafast dual-wavelength excitation,” *Journal of biophotonics* **13**(6), e201960229 (2020).

- [48] R. Cao, J. Li, B. Ning, *et al.*, “Functional and oxygen-metabolic photoacoustic microscopy of the awake mouse brain,” *Neuroimage* **150**, 77–87 (2017).
- [49] T. Jin, H. Guo, H. Jiang, *et al.*, “Portable optical resolution photoacoustic microscopy (porpam) for human oral imaging,” *Optics letters* **42**(21), 4434–4437 (2017).
- [50] B. Zabihian, J. Weingast, M. Liu, *et al.*, “In vivo dual-modality photoacoustic and optical coherence tomography imaging of human dermatological pathologies,” *Biomedical optics express* **6**(9), 3163–3178 (2015).
- [51] T. T. Wong, R. Zhang, P. Hai, *et al.*, “Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy,” *Science advances* **3**(5), e1602168 (2017).
- [52] A. Berezhnoi, M. Schwarz, A. Buehler, *et al.*, “Assessing hyperthermia-induced vasodilation in human skin in vivo using optoacoustic mesoscopy,” *Journal of biophotonics* **11**(11), e201700359 (2018).
- [53] Y. Zhou, F. Cao, H. Li, *et al.*, “Photoacoustic imaging of microenvironmental changes in facial cupping therapy,” *Biomedical optics express* **11**(5), 2394–2401 (2020).
- [54] Y. Zhou, C. Liu, X. Huang, *et al.*, “Low-consumption photoacoustic method to measure liquid viscosity,” *Biomedical optics express* **12**(11), 7139–7148 (2021).
- [55] H. Dai, Q. Shen, J. Shao, *et al.*, “Small molecular nir-ii fluorophores for cancer phototheranostics,” *The Innovation* **2**(1) (2021).
- [56] S. Park, C. Lee, J. Kim, *et al.*, “Acoustic resolution photoacoustic microscopy,” *Biomedical Engineering Letters* **4**, 213–222 (2014).
- [57] H. Zhong, T. Duan, H. Lan, *et al.*, “Review of low-cost photoacoustic sensing and imaging based on laser diode and light-emitting diode,” *Sensors* **18**(7), 2264 (2018).
- [58] M. Erfanzadeh and Q. Zhu, “Photoacoustic imaging with low-cost sources; a review,” *Photoacoustics* **14**, 1 (2019).
- [59] X. Li, V. T. Tsang, L. Kang, *et al.*, “High-speed high-resolution laser diode-based photoacoustic microscopy for in vivo microvasculature imaging,” *Visual Computing for Industry, Biomedicine, and Art* **4**, 1–6 (2021).

- [60] J. Shi, T. T. Wong, Y. He, *et al.*, “High-resolution, high-contrast mid-infrared imaging of fresh biological samples with ultraviolet-localized photoacoustic microscopy,” *Nature photonics* **13**(9), 609–615 (2019).
- [61] J. Gröhl, M. Schellenberg, K. Dreher, *et al.*, “Deep learning for biomedical photoacoustic imaging: A review,” *Photoacoustics* **22**, 100241 (2021).
- [62] C. Yang, H. Lan, F. Gao, *et al.*, “Review of deep learning for photoacoustic imaging,” *Photoacoustics* **21**, 100215 (2021).
- [63] A. Hariri, K. Alipour, Y. Mantri, *et al.*, “Deep learning improves contrast in low-fluence photoacoustic imaging,” *Biomedical optics express* **11**(6), 3360–3373 (2020).
- [64] R. Manwar, X. Li, S. Mahmoodkalayeh, *et al.*, “Deep learning protocol for improved photoacoustic brain imaging,” *Journal of biophotonics* **13**(10), e202000212 (2020).
- [65] A. Sharma and M. Pramanik, “Convolutional neural network for resolution enhancement and noise reduction in acoustic resolution photoacoustic microscopy,” *Biomedical Optics Express* **11**(12), 6826–6839 (2020).
- [66] A. DiSpirito, D. Li, T. Vu, *et al.*, “Reconstructing undersampled photoacoustic microscopy images using deep learning,” *IEEE transactions on medical imaging* **40**(2), 562–570 (2020).
- [67] H. Zhao, Z. Ke, F. Yang, *et al.*, “Deep learning enables superior photoacoustic imaging at ultralow laser dosages,” *Advanced science* **8**(3), 2003097 (2021).
- [68] A. Hauptmann, F. Lucka, M. Betcke, *et al.*, “Model-based learning for accelerated, limited-view 3-d photoacoustic tomography,” *IEEE transactions on medical imaging* **37**(6), 1382–1393 (2018).
- [69] N. Davoudi, X. L. Deán-Ben, and D. Razansky, “Deep learning photoacoustic tomography with sparse data,” *Nature Machine Intelligence* **1**(10), 453–460 (2019).
- [70] S. Guan, A. A. Khan, S. Sikdar, *et al.*, “Fully dense unet for 2-d sparse photoacoustic tomography artifact removal,” *IEEE journal of biomedical and health informatics* **24**(2), 568–576 (2019).
- [71] T. Vu, M. Li, H. Humayun, *et al.*, “A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer,” *Experimental Biology and Medicine* **245**(7), 597–605 (2020).

- [72] C. Dong, C. C. Loy, K. He, *et al.*, “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence* **38**(2), 295–307 (2015).
- [73] C. Ledig, L. Theis, F. Huszár, *et al.*, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690 (2017).
- [74] B. Lim, S. Son, H. Kim, *et al.*, “Enhanced deep residual networks for single image super-resolution,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 136–144 (2017).
- [75] X. Wang, K. Yu, S. Wu, *et al.*, “Esrgan: Enhanced super-resolution generative adversarial networks,” in *Proceedings of the European conference on computer vision (ECCV) workshops*, 0–0 (2018).
- [76] Y. Rivenson, Z. Göröcs, H. Günaydin, *et al.*, “Deep learning microscopy,” *Optica* **4**(11), 1437–1443 (2017).
- [77] H. Zhao, Z. Ke, N. Chen, *et al.*, “A new deep learning method for image deblurring in optical microscopic systems,” *Journal of biophotonics* **13**(3), e201960147 (2020).
- [78] S. Cheng, H. Li, Y. Luo, *et al.*, “Artificial intelligence-assisted light control and computational imaging through scattering media,” *Journal of innovative optical health sciences* **12**(04), 1930006 (2019).
- [79] Y. Rivenson, H. Wang, Z. Wei, *et al.*, “Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning,” *Nature biomedical engineering* **3**(6), 466–477 (2019).
- [80] H. Wang, Y. Rivenson, Y. Jin, *et al.*, “Deep learning enables cross-modality super-resolution in fluorescence microscopy,” *Nature methods* **16**(1), 103–110 (2019).
- [81] I. Gulrajani, F. Ahmed, M. Arjovsky, *et al.*, “Improved training of wasserstein gans,” *Advances in neural information processing systems* **30** (2017).
- [82] W. Xing, L. Wang, K. Maslov, *et al.*, “Integrated optical-and acoustic-resolution photoacoustic microscopy based on an optical fiber bundle,” *Optics letters* **38**(1), 52–54 (2012).

- [83] A. Dosovitskiy, J. T. Springenberg, M. Riedmiller, *et al.*, “Discriminative unsupervised feature learning with convolutional neural networks,” *Advances in neural information processing systems* **27** (2014).
- [84] E. Castro, J. S. Cardoso, and J. C. Pereira, “Elastic deformations for data augmentation in breast cancer mass detection,” in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 230–234, IEEE (2018).
- [85] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* **18**, 234–241, Springer (2015).
- [86] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, *et al.*, “Albumentations: fast and flexible image augmentations,” *Information* **11**(2), 125 (2020).
- [87] I. Goodfellow, “Nips 2016 tutorial: Generative adversarial networks,” *arXiv preprint arXiv:1701.00160* (2016).
- [88] M. Arjovsky and L. Bottou, “Towards principled methods for training generative adversarial networks,” *arXiv preprint arXiv:1701.04862* (2017).
- [89] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *International conference on machine learning*, 214–223, PMLR (2017).
- [90] K. He, X. Zhang, S. Ren, *et al.*, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
- [91] Y. Wu and K. He, “Group normalization,” in *Proceedings of the European conference on computer vision (ECCV)*, 3–19 (2018).
- [92] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Instance normalization: The missing ingredient for fast stylization,” *arXiv preprint arXiv:1607.08022* (2016).
- [93] T. Nguyen, Y. Xue, Y. Li, *et al.*, “Deep learning approach for fourier ptychography microscopy,” *Optics express* **26**(20), 26470–26484 (2018).
- [94] K. He, X. Zhang, S. Ren, *et al.*, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 1026–1034 (2015).

- [95] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” *arXiv preprint arXiv:1711.05101* (2017).
- [96] P. Ruiz, X. Zhou, J. Mateos, *et al.*, “Variational bayesian blind image deconvolution: A review,” *Digital Signal Processing* **47**, 116–127 (2015).
- [97] T. E. Bishop, S. D. Babacan, B. Amizic, *et al.*, “Blind image deconvolution: problem formulation and existing approaches,” in *Blind Image Deconvolution*, 21–62, CRC press (2017).
- [98] M. R. Chowdhury, J. Qin, and Y. Lou, “Non-blind and blind deconvolution under poisson noise using fractional-order total variation,” *Journal of Mathematical Imaging and Vision* **62**(9), 1238–1255 (2020).
- [99] Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing* **13**(4), 600–612 (2004).
- [100] M. Fauver, E. J. Seibel, J. R. Rahn, *et al.*, “Three-dimensional imaging of single isolated cell nuclei using optical projection tomography,” *Optics express* **13**(11), 4210–4223 (2005).
- [101] J. Kim, G. Kim, L. Li, *et al.*, “Deep learning acceleration of multiscale superresolution localization photoacoustic imaging,” *Light: science & applications* **11**(1), 131 (2022).
- [102] T. Chaigne, J. Gateau, O. Katz, *et al.*, “Light focusing and two-dimensional imaging through scattering media using the photoacoustic transmission matrix with an ultrasound array,” *Optics letters* **39**(9), 2664–2667 (2014).
- [103] T. Zhao, S. Ourselin, T. Vercauteren, *et al.*, “High-speed photoacoustic-guided wavefront shaping for focusing light in scattering media,” *Optics Letters* **46**(5), 1165–1168 (2021).
- [104] Z. Zhang, H. Jin, Z. Zheng, *et al.*, “Photoacoustic microscopy imaging from acoustic resolution to optical resolution enhancement with deep learning,” in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5, IEEE (2021).
- [105] I. M. Vellekoop and A. P. Mosk, “Universal optimal transmission of light through disordered materials,” *Physical review letters* **101**(12), 120601 (2008).
- [106] S. M. Popoff, G. Lerosey, M. Fink, *et al.*, “Controlling light through optical disordered media: transmission matrix approach,” *New Journal of Physics* **13**(12), 123021 (2011).
- [107] D. Loterie, S. Farahi, I. Papadopoulos, *et al.*, “Digital confocal microscopy through a multimode fiber,” *Optics express* **23**(18), 23845–23858 (2015).

- [108] M. Plöschner and T. Čižmár, “Compact multimode fiber beam-shaping system based on gpu accelerated digital holography,” *Optics letters* **40**(2), 197–200 (2015).
- [109] D. B. Conkey and R. Piestun, “Color image projection through a strongly scattering wall,” *Optics express* **20**(25), 27312–27318 (2012).
- [110] L. Wan, Z. Chen, H. Huang, *et al.*, “Focusing light into desired patterns through turbid media by feedback-based wavefront shaping,” *Applied Physics B* **122**, 1–7 (2016).
- [111] Q. Feng, F. Yang, X. Xu, *et al.*, “Multi-objective optimization genetic algorithm for multi-point light focusing in wavefront shaping,” *Optics Express* **27**(25), 36459–36473 (2019).
- [112] L. Zhan, D. Liu, A. Yang, *et al.*, “Fast compensation for arbitrary focusing through scattering media,” *Applied Optics* **58**(10), 2624–2629 (2019).
- [113] H. Li, X. Wu, G. Liu, *et al.*, “Generation of focal patterns with uniform intensity distribution from speckle by hadamard-genetic algorithm,” *IEEE Photonics Journal* **13**(3), 1–8 (2021).
- [114] Z. Yang, L. Fang, X. Zhang, *et al.*, “Controlling a scattered field output of light passing through turbid medium using an improved ant colony optimization algorithm,” *Optics and Lasers in Engineering* **144**, 106646 (2021).
- [115] A. Daniel, L. Liberman, and Y. Silberberg, “Wavefront shaping for glare reduction,” *Optica* **3**(10), 1104–1106 (2016).
- [116] E. H. Zhou, A. Shibukawa, J. Brake, *et al.*, “Glare suppression by coherence gated negation,” *Optica* **3**(10), 1107–1113 (2016).
- [117] D. Huang, E. A. Swanson, C. P. Lin, *et al.*, “Optical coherence tomography,” *science* **254**(5035), 1178–1181 (1991).
- [118] G. Volpe, L. Kurz, A. Callegari, *et al.*, “Speckle optical tweezers: micromanipulation with random light fields,” *Optics express* **22**(15), 18159–18167 (2014).
- [119] E. Mudry, K. Belkebir, J. Girard, *et al.*, “Structured illumination microscopy using unknown speckle patterns,” *Nature Photonics* **6**(5), 312–315 (2012).
- [120] J. Luo, Z. Wu, D. Wu, *et al.*, “Efficient glare suppression with hadamard-encoding-algorithm-based wavefront shaping,” *Optics Letters* **44**(16), 4067–4070 (2019).

- [121] H. Zhang, B. Zhang, K. Liu, *et al.*, “Large-scale, high-contrast glare suppression with low-transmittance eigenchannels of aperture-target transmission matrices,” *Optics Letters* **46**(7), 1498–1501 (2021).
- [122] D. Wierstra, T. Schaul, T. Glasmachers, *et al.*, “Natural evolution strategies,” *The Journal of Machine Learning Research* **15**(1), 949–980 (2014).
- [123] J. Yang, Q. He, L. Liu, *et al.*, “Anti-scattering light focusing by fast wavefront shaping based on multi-pixel encoded digital-micromirror device,” *Light: Science & Applications* **10**(1), 149 (2021).
- [124] R. W. Gerchberg, “A practical algorithm for the determination of plane from image and diffraction pictures,” *Optik* **35**(2), 237–246 (1972).
- [125] J. R. Fienup, “Phase retrieval algorithms: a comparison,” *Applied optics* **21**(15), 2758–2769 (1982).
- [126] J. R. Fienup, “Reconstruction of an object from the modulus of its fourier transform,” *Optics letters* **3**(1), 27–29 (1978).
- [127] D. B. Conkey, A. M. Caravaca-Aguirre, and R. Piestun, “High-speed scattering medium characterization with application to focusing light through turbid media,” *Optics express* **20**(2), 1733–1740 (2012).
- [128] W.-H. Lee, “Binary computer-generated holograms,” *Applied Optics* **18**(21), 3661–3669 (1979).
- [129] J.-H. Park, J. Park, K. Lee, *et al.*, “Disordered optics: exploiting multiple light scattering and wavefront shaping for nonconventional optical elements,” *Advanced Materials* **32**(35), 1903457 (2020).
- [130] S. Gigan, “Imaging and computing with disorder,” *Nature Physics* **18**(9), 980–985 (2022).
- [131] S. Rotter and S. Gigan, “Light fields in complex media: Mesoscopic scattering meets wave control,” *Reviews of Modern Physics* **89**(1), 015005 (2017).
- [132] S. Gigan, O. Katz, H. B. De Aguiar, *et al.*, “Roadmap on wavefront shaping and deep imaging in complex media,” *Journal of Physics: Photonics* **4**(4), 042501 (2022).
- [133] Z. Yu, H. Li, T. Zhong, *et al.*, “Wavefront shaping: a versatile tool to conquer multiple scattering in multidisciplinary fields,” *The Innovation* **3**(5) (2022).

- [134] C. M. Woo, Q. Zhao, T. Zhong, *et al.*, “Optimal efficiency of focusing diffused light through scattering media with iterative wavefront shaping,” *APL Photonics* **7**, 046109 (2022).
- [135] M. Kim, W. Choi, Y. Choi, *et al.*, “Transmission matrix of a scattering medium and its applications in biophotonics,” *Optics express* **23**(10), 12648–12668 (2015).
- [136] S. Popoff, G. Lerosey, M. Fink, *et al.*, “Image transmission through an opaque material,” *Nature communications* **1**(1), 81 (2010).
- [137] H. Li, Z. Yu, Q. Zhao, *et al.*, “Learning-based super-resolution interpolation for subnyquist sampled laser speckles,” *Photonics Research* **11**(4), 631–642 (2023).
- [138] S. Cheng, T. Zhong, C. M. Woo, *et al.*, “Alternating projection-based phase optimization for arbitrary glare suppression through multimode fiber,” *Optics and Lasers in Engineering* **161**, 107368 (2023).
- [139] T. Čižmár and K. Dholakia, “Exploiting multimode waveguides for pure fibre-based imaging,” *Nature communications* **3**(1), 1027 (2012).
- [140] S. Turtaev, I. T. Leite, T. Altwegg-Boussac, *et al.*, “High-fidelity multimode fibre-based endoscopy for deep brain in vivo imaging,” *Light: Science & Applications* **7**(1), 92 (2018).
- [141] D. Stellinga, D. B. Phillips, S. P. Mekhail, *et al.*, “Time-of-flight 3d imaging through multimode optical fibers,” *Science* **374**(6573), 1395–1399 (2021).
- [142] I. T. Leite, S. Turtaev, X. Jiang, *et al.*, “Three-dimensional holographic optical manipulation through a high-numerical-aperture soft-glass multimode fibre,” *Nature Photonics* **12**(1), 33–39 (2018).
- [143] T. Zhong, Z. Qiu, Y. Wu, *et al.*, “Optically selective neuron stimulation with a wavefront shaping-empowered multimode fiber,” *Advanced Photonics Research* **3**(3), 2100231 (2022).
- [144] A. Liutkus, D. Martina, S. Popoff, *et al.*, “Imaging with nature: Compressive imaging using a multiply scattering medium,” *Scientific reports* **4**(1), 5552 (2014).
- [145] L. Gong, Q. Zhao, H. Zhang, *et al.*, “Optical orbital-angular-momentum-multiplexed data transmission under high scattering,” *Light: Science & Applications* **8**(1), 27 (2019).

- [146] J. Zhong, Z. Wen, Q. Li, *et al.*, “Efficient reference-less transmission matrix retrieval for a multimode fiber using fast fourier transform,” *Advanced Photonics Nexus* **2**(5), 056007–056007 (2023).
- [147] Y. Choi, T. D. Yang, C. Fang-Yen, *et al.*, “Overcoming the diffraction limit using multiple light scattering in a highly disordered medium,” *Physical review letters* **107**(2), 023902 (2011).
- [148] H. Yu, K. Lee, and Y. Park, “Ultrahigh enhancement of light focusing through disordered media controlled by mega-pixel modes,” *Optics express* **25**(7), 8036–8047 (2017).
- [149] T. Zhao, S. Ourselin, T. Vercauteren, *et al.*, “Seeing through multimode fibers with real-valued intensity transmission matrices,” *Optics express* **28**(14), 20978–20991 (2020).
- [150] X. Tao, D. Bodington, M. Reinig, *et al.*, “High-speed scanning interferometric focusing by fast measurement of binary transmission matrix for channel demixing,” *Optics express* **23**(11), 14168–14187 (2015).
- [151] I. Yamaguchi, “Phase-shifting digital holography: Principles and applications,” *Digital holography and three-dimensional display: principles and applications*, 145–171 (2006).
- [152] E. CuChe, P. Marquet, and C. Depeursinge, “Spatial filtering for zero-order and twin-image elimination in digital off-axis holography,” *Applied optics* **39**(23), 4070–4075 (2000).
- [153] B. Rajaei, E. W. Tramel, S. Gigan, *et al.*, “Intensity-only optical compressive imaging using a multiply scattering material and a double phase retrieval approach,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4054–4058, IEEE (2016).
- [154] C. A. Metzler, M. K. Sharma, S. Nagesh, *et al.*, “Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset,” in *2017 IEEE International Conference on Computational Photography (ICCP)*, 1–16, IEEE (2017).
- [155] M. N’ Gom, T. B. Norris, E. Michielssen, *et al.*, “Mode control in a multimode fiber through acquiring its transmission matrix from a reference-less optical system,” *Optics letters* **43**(3), 419–422 (2018).
- [156] G. Huang, D. Wu, J. Luo, *et al.*, “Retrieving the optical transmission matrix of a multimode fiber using the extended kalman filter,” *Optics express* **28**(7), 9487–9500 (2020).

- [157] S. Cheng, T. Zhong, and P. Lai, “Non-convex optimization for retrieving the complex transmission matrix of a multimode fiber,” in *TENCON 2022-2022 IEEE Region 10 Conference (TENCON)*, 1–5, IEEE (2022).
- [158] D. Ancora, L. Dominici, A. Gianfrate, *et al.*, “Speckle spatial correlations aiding optical transmission matrix retrieval: the smoothed gerchberg–saxton single-iteration algorithm,” *Photonics Research* **10**(10), 2349–2358 (2022).
- [159] S. Marchesini, “Invited article: A unified evaluation of iterative projection algorithms for phase retrieval,” *Review of scientific instruments* **78**(1), 011301 (2007).
- [160] I. Waldspurger, A. d’Aspremont, and S. Mallat, “Phase recovery, maxcut and complex semidefinite programming,” *Mathematical Programming* **149**, 47–81 (2015).
- [161] P. Schniter and S. Rangan, “Compressive phase retrieval via generalized approximate message passing,” *IEEE Transactions on Signal Processing* **63**(4), 1043–1055 (2014).
- [162] S. Rangan, P. Schniter, and A. K. Fletcher, “Vector approximate message passing,” *IEEE Transactions on Information Theory* **65**(10), 6664–6684 (2019).
- [163] P. Netrapalli, P. Jain, and S. Sanghavi, “Phase retrieval using alternating minimization,” in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS’13*, 2796–2804, Curran Associates Inc., (Red Hook, NY, USA) (2013).
- [164] E. J. Candes, X. Li, and M. Soltanolkotabi, “Phase retrieval via wirtinger flow: Theory and algorithms,” *IEEE Transactions on Information Theory* **61**(4), 1985–2007 (2015).
- [165] G. Wang, G. B. Giannakis, and Y. C. Eldar, “Solving systems of random quadratic equations via truncated amplitude flow,” *IEEE Transactions on Information Theory* **64**(2), 773–794 (2017).
- [166] H. Zhang, Y. Zhou, Y. Liang, *et al.*, “A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms,” *Journal of Machine Learning Research* **18** (2017).
- [167] G. Wang, G. B. Giannakis, Y. Saad, *et al.*, “Phase retrieval via reweighted amplitude flow,” *IEEE Transactions on Signal Processing* **66**(11), 2818–2833 (2018).
- [168] Y. Chen and E. Candes, “Solving random quadratic systems of equations is nearly as easy as solving linear systems,” in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, *et al.*, Eds., **28**, Curran Associates, Inc. (2015).

- [169] T. Goldstein, C. Studer, and R. Baraniuk, “A field guide to forward-backward splitting with a fasta implementation,” *arXiv preprint arXiv:1411.3406* (2014).
- [170] L. Ji and Z. Tie, “On gradient descent algorithm for generalized phase retrieval problem,” in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, 320–325, IEEE (2016).
- [171] R. Chandra, T. Goldstein, and C. Studer, “Phasepack: A phase retrieval library,” in *2019 13th International conference on Sampling Theory and Applications (SampTA)*, 1–5, IEEE (2019).
- [172] D. G. Ouzounov, T. Wang, M. Wang, *et al.*, “In vivo three-photon imaging of activity of gcamp6-labeled neurons deep in intact mouse brain,” *Nature methods* **14**(4), 388–390 (2017).
- [173] W. Zong, R. Wu, M. Li, *et al.*, “Fast high-resolution miniature two-photon microscopy for brain imaging in freely behaving mice,” *Nature methods* **14**(7), 713–719 (2017).
- [174] C. Zhao, S. Chen, L. Zhang, *et al.*, “Miniature three-photon microscopy maximized for scattered fluorescence collection,” *Nature Methods* **20**(4), 617–622 (2023).
- [175] J. Tang, R. N. Germain, and M. Cui, “Superpenetration optical microscopy by iterative multiphoton adaptive compensation technique,” *Proceedings of the National Academy of Sciences* **109**(22), 8434–8439 (2012).
- [176] S. Bianchi and R. Di Leonardo, “A multi-mode fiber probe for holographic micromanipulation and microscopy,” *Lab on a Chip* **12**(3), 635–639 (2012).
- [177] A. M. Caravaca-Aguirre and R. Piestun, “Single multimode fiber endoscope,” *Optics express* **25**(3), 1656–1665 (2017).
- [178] S. A. Vasquez-Lopez, R. Turcotte, V. Koren, *et al.*, “Subcellular spatial resolution achieved for deep-brain imaging in vivo using a minimally invasive multimode fiber,” *Light: science & applications* **7**(1), 110 (2018).
- [179] Z. Wen, Z. Dong, Q. Deng, *et al.*, “Single multimode fibre for in vivo light-field-encoded endoscopic imaging,” *Nature Photonics* **17**(8), 679–687 (2023).
- [180] R. Y. Gu, R. N. Mahalati, and J. M. Kahn, “Design of flexible multi-mode fiber endoscope,” *Optics express* **23**(21), 26905–26918 (2015).

- [181] G. S. Gordon, M. Gataric, A. G. C. Ramos, *et al.*, “Characterizing optical fiber transmission matrices using metasurface reflector stacks for lensless imaging without distal access,” *Physical review X* **9**(4), 041050 (2019).
- [182] H. Chen, N. K. Fontaine, R. Ryf, *et al.*, “Remote spatio-temporal focusing over multimode fiber enabled by single-ended channel estimation,” *IEEE Journal of Selected Topics in Quantum Electronics* **26**(4), 1–9 (2020).
- [183] S. Farahi, D. Ziegler, I. N. Papadopoulos, *et al.*, “Dynamic bending compensation while focusing through a multimode fiber,” *Optics express* **21**(19), 22504–22514 (2013).
- [184] S. Li, S. A. Horsley, T. Tyc, *et al.*, “Memory effect assisted imaging through multimode optical fibres,” *Nature Communications* **12**(1), 3751 (2021).
- [185] P. Fan, Y. Wang, M. Ruddlesden, *et al.*, “Deep learning enabled scalable calibration of a dynamically deformed multimode fiber,” *Advanced Photonics Research* **3**(10), 2100304 (2022).
- [186] Y. Zheng, T. Wright, Z. Wen, *et al.*, “Single-ended recovery of optical fiber transmission matrices using neural networks,” *Communications Physics* **6**(1), 306 (2023).
- [187] A. D. Gomes, S. Turtaev, Y. Du, *et al.*, “Near perfect focusing through multimode fibres,” *Optics Express* **30**(7), 10645–10663 (2022).
- [188] X. Hu, Q. Zhao, P. Yu, *et al.*, “Dynamic shaping of orbital-angular-momentum beams for information encoding,” *Optics express* **26**(2), 1796–1808 (2018).
- [189] A. Boniface, J. Dong, and S. Gigan, “Non-invasive focusing and imaging in scattering media with a fluorescence-based transmission matrix,” *Nature communications* **11**(1), 6154 (2020).
- [190] L. Zhu, F. Soldevila, C. Moretti, *et al.*, “Large field-of-view non-invasive imaging through scattering layers using fluctuating random illumination,” *Nature communications* **13**(1), 1447 (2022).
- [191] H. Kasai, “Nmflibrary: Non-negative matrix factorization library.” Accessed: 2021-08-17.
- [192] M. Plöschner, T. Tyc, and T. Čižmár, “Seeing through chaos in multimode fibres,” *Nature photonics* **9**(8), 529–535 (2015).

-
- [193] A. M. Caravaca-Aguirre, A. Carron, S. Mezil, *et al.*, “Optical memory effect in square multimode fibers,” *Optics Letters* **46**(19), 4924–4927 (2021).
- [194] Y. Luo, S. Yan, H. Li, *et al.*, “Towards smart optical focusing: deep learning-empowered dynamic wavefront shaping through nonstationary scattering media,” *Photonics Research* **9**(8), B262–B278 (2021).
- [195] S. Khan, M. Naseer, M. Hayat, *et al.*, “Transformers in vision: A survey,” *ACM computing surveys (CSUR)* **54**(10s), 1–41 (2022).
- [196] A. III, “Duke pam datase.” Accessed: 2021-08-17.
- [197] Y. Shen, J. Ma, C. Hou, *et al.*, “Acoustic-feedback wavefront-adapted photoacoustic microscopy,” *Optica* **11**(2), 214–221 (2024).
- [198] V. Tran, T. Wang, N. P. Nazirkar, *et al.*, “On the exploration of structured light transmission through a multimode fiber in a reference-less system,” *APL Photonics* **8**(12) (2023).