Title
Single-Shot 3D Microscopy: Optics and Algorithms Co-Design

Author
Liu, Fanglin Linda

Publication Date
2022

Peer reviewed|Thesis/dissertation

Single-Shot 3D Microscopy: Optics and Algorithms Co-Design

by

Fanglin Linda Liu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Laura Waller, Chair
Associate Professor Ren Ng
Associate Professor Hillel Adesnik

Spring 2022

Single-Shot 3D Microscopy: Optics and Algorithms Co-Design

Abstract

Single-Shot 3D Microscopy: Optics and Algorithms Co-Design

by

Fanglin Linda Liu

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Associate Professor Laura Waller, Chair

Computational imaging involves simultaneously designing optical hardware and reconstruction software. Such a co-design framework brings together the best of both worlds for an imaging system. The goal is to develop a high-speed, high-resolution, and large field-of-view microscope that can detect 3D fluorescence signals from single image acquisition. To achieve this goal, I propose a new method called Fourier DiffuserScope, a single-shot 3D fluorescent microscope that uses a phase mask (i.e., a diffuser with random microlenses) in the Fourier plane to encode 3D information, then computationally reconstructs the volume by solving a sparsity-constrained inverse problem.

In this dissertation, I will discuss the design principles of the Fourier DiffuserScope from three perspectives: first-principles optics, compressed sensing theory, and physics-based machine learning. First, in the heuristic design, the phase mask consists of randomly placed microlenses with varying focal lengths; the random positions provide a larger field-of-view compared to a conventional microlens array, and the diverse focal lengths improve the axial depth range. I then build an experimental system that achieves $< 3\,\mu m$ lateral and $4\,\mu m$ axial resolution over a $1000 \times 1000 \times 280\,\mu m^3$ volume. Lastly, we use a differentiable forward model of Fourier DiffuserScope in conjunction with a differentiable reconstruction algorithm to jointly optimize both the phase mask surface profile and the reconstruction parameters. We validate our method in 2D and 3D single-shot imaging, where the optimized diffuser demonstrates improved reconstruction quality compared to previous heuristic designs.

To my parents and Meng.

Contents

List of Figures

List of Tables

Acknowledgments

This thesis is a fruitful journey supported by many people around me. First and foremost, I want to express my most sincere gratitude to my Ph.D. advisor, Prof. Laura Waller. Laura encouraged me in an email to learn optics when I was a sophomore student in China, and that's the beginning of everything. I then came to Berkeley for my Ph.D. — the best choice I've ever made. Throughout my years in graduate school, what I've learned from Laura is beyond technical skills and more about how to become a better person. Laura taught me to be an independent researcher, by giving me the freedom to explore research ideas, introducing me to collaborators, and supporting me to lead my own project since the beginning of my graduate school. Laura taught me to become a good communicator, through inspiring and clear presentations, as well as structural and concise writings; since research should be made accessible to a broader community. Laura inspired me to be more confident. Following Laura's suggestions, I attended conferences, taught a big class with more than 1000 students, mentored undergrads, and applied for internships. After all the training, being confident is now one of my biggest achievements in graduate school. Laura prioritizes her students' well-being, and helps people whenever she can. Throughout these years, I am constantly impressed by Laura's insights, diligence, and kindness. There are still a lot more I can learn from Laura, and Laura will always be a role model for me.

In addition to my academic advisors, I am also indebted to several mentors in both WallerLab and industry. Grace Kuo and Nick Antipa led me into the DiffuserCam project and guided me through the diffuser microscope project. Shwetadwip Chowdhury taught me optical alignment skills in the lab. Li-hao Yeh was my go-to person for math problems, optical simulations, and literally any project blockers that I encountered. I did two internships during my Ph.D. and was lucky enough to have supportive industrial mentors. Tianbo Sun and Jie Sun at LightIC taught me about Lidar, connected me with industrial manufacturers, and introduced to me how early-stage start-up works. Ariel Gomez Diaz hosted me during my internship at Microsoft Research. I learned from Ariel polarization microscopes, lab alignment skills, and most importantly, teamwork skills and communication skills in a big organization. Additional thanks go to my thesis and qualifying exam committee, Prof. Ren Ng, Prof. Hillel Adesnik, and Prof. Saul Kato, for their enlightening input in guiding my explorations.

Having lovely group members made my journey delightful. I always appreciate how diversified my labmates are — almost half of the Ph.D. students and postdocs are women. Grace Kuo's afternoon chit-chat contained the philosophy of both science and life. Regina Eckert was usually the one who took on responsibilities, and she is never hesitant to speak out for other people. Kristina Monakhova had infinite enthusiasm for research and non-research projects, and I got cheered up talking to her. Emma Alexander and Yi Xue taught me about faculty jobs applications. Tiffany Chien, Neerja Aggarwal, and Leyla Kabuli brought joy to the lab during experiments. Besides the women supporting group, the rest half of the group is as awesome. Kryollos Yanny and I both worked on diffuser microscopes, and we shared achievements and struggles together. Eric Markley is the best collaborator I can

ever ask for. He is talented but modest, passionate but calm. Working with Eric is a great bonus to my Ph.D. journey. Zack Phillips, Michael Chen, Herbert Liu, and Nicolas Pegard were very helpful in getting me started in computational imaging during my first year. Zack Phillips, Michael Chen, and Grace Kuo also shared their experiences and suggestions during my industrial job search. I learned state-of-the-art machine learning ideas from Michael Kellman and Emrah Bostan. My cubicle mates, David Ren and Ruiming Cao created a cheerful vibe in the office. Gautam Gunjala and Stuart Sherwin are my same-year cohort, and we discussed classes, quals, and nowaday the nitty-gritty of graduation.

My participation in student groups is another significant part of my graduate school experience. I joined the Chinese Graduate and Postdoctoral Scholars Association (CGPSA) in 2016 and became the copresident in 2018. Many thanks to the other copresident Jialiang Gu, and the VPs of that year, Suifang Mao, Qingjun Wang, Peter Wang, Fangzhou Zhao, Houjun Tang, Yujiao Cheng, and Boying Gong. Together, we held tens of events, including the Newcomer Welcome Party, Lunar New Year's Dumpling Making Party, Spring Break at Angel Island, etc. These events connected Chinese international students, bringing us a sense of belonging away from our hometowns. I also want to thank the Women in Computer Science and Engineering (WiCSE) Group. I truly miss those casual conversations at weekly WiCSE lunch gatherings. In particular, I met Alice Ye at WiCSE, who gave me huge encouragement to help me get over those hard times, and later sent me heartwarming sweets to celebrate my achievements.

Huge thanks to my parents, Ying Pu and Chengfu Liu. They encouraged my curiosity about science and bought me light bulbs and batteries to support my lighthouse project in kindergarten. When I tried to repair electronic gadgets but ended up breaking them apart, my parents still sent me smiles. Even though they live in Mianyang, Sichuan, China, which is 10,890 kilometers away from Berkeley with an 8-hour time difference, they always listen to my grumble, celebrate my achievements, and give me unconditional love and steadfast support.

Finally and most importantly, it feels like a gift from heaven to meet Meng, my husband, at Berkeley. Meng and I are not only significant others, but we enjoy being friends chatting about non-sense, classmates learning new skills, and coworkers discussing technical projects. No matter how challenging the Ph.D. journey was, Meng's humor and reliability light our path forward, and the same for our life in the future.

Chapter 1

Introduction

1.1 Single-Shot 3D Imaging

Volumetric fluorescence imaging with video-rate capture at cellular resolution is essential for understanding dynamic 3D biological systems. To simultaneously monitor large numbers of cells over time, imaging systems must record samples at tens-of-volumes per second. Scanning-based volumetric microscopy techniques (e.g. wide-field, two-photon, confocal, light-sheet) rely on physically-moving components and sequentially acquiring images, forcing a trade-off between temporal resolution and three-dimensional (3D) volume size. Furthermore, spatio-temporal information can be ambiguous since the depth planes are not captured at the same time. In comparison, single-shot 3D imaging techniques form an entire 3D object from a single acquisition, a promising strategy for achieving high speed imaging over a large volume.

Single-shot 3D imaging with a 2D sensor is possible by using a hardware encoding procedure followed by a computational decoding procedure. Light field microscopy (LFM) [1–8] is one popular implementation of this, where a microlens array (MLA) is inserted in front of the microscope's image sensor to simultaneously capture 2D spatial and 2D angular information. The resulting 4D light field can be used for digital refocusing, perspective synthesis, or 3D reconstruction. However, using a 2D sensor to sample a 4D light field requires trading off angular and spatial sampling, resulting in poor resolution. This is particularly undesirable in microscopy, where resolution is the key performance metric.

Besides variations of LFM, there are other methods for single-shot 3D fluorescence microscopy.

Multifocal microscopy methods simultaneously capture multiple in-focus images at different depths. This can be done by using beamsplitters and multiple cameras conjugate to different depth planes [9]; however, the resulting system is expensive and bulky. To acquire multiple depths with a single sensor, a distorted phase grating can be inserted in the pupil plane to project different axial layers onto different sub-images on the sensor [10–12]. A similar result can be achieved with superimposed Fresnel lenses [13] or a diffractive

metalens [14]. For more than a few depth planes, multiplexed volume holography is a good option due to its low cross-talk [15]. In all these methods, however, the FOV is sacrificed in order to increase the number of depth planes (generally less than 25 [12]), limited by how many sub-images fit on the sensor.

PSF engineering for point localization refers to methods that use a coded mask in Fourier space but with the image captured in image (real) space. This results in a depth-dependant PSF (e.g. astigmatic [16, 17], double-helix [18–20], tetrapod [21], etc.) that, along with inverse algorithms, is well-suited to localize separated point-like molecules [16–19, 21], but ill-posed when the object is continuous [20]. Because the Fourier DiffuserScope places both the phase mask and the sensor near the Fourier plane, there is a much larger PSF in which the energy is distributed into more features, so that the cross-correlation of laterally and axially shifted PSFs is lower than that of engineered PSFs. As a result, the random diffuser is better suited to reconstruct a 3D volume from an undersampled 2D measurement, according to compressed sensing theory [22].

Lensless mask-based imaging, which uses a coded aperture for lens-free 2D [23] or 3D [24] imaging, first emerged in X-ray and gamma-ray systems [25, 26] for 2D imaging in situations where lenses are difficult to implement. Amplitude coded masks are straightforward to design and easy to fabricate, but come with the inherent issue of blocking photons, which leads to a low signal-to-noise ratio (SNR) in the acquisition and noise amplification during reconstruction. Phase masks are more difficult to fabricate but have much better light efficiency [27].

Diffuser-based microscopy describes several different architectures that emerged from the original DiffuserCam [28], which is a lensless phase-mask-based imager that uses a diffuser for encoding 3D information. The Computational Imaging Lab at UC Berkeley have demonstrated 2D [29, 30], 3D [28, 29, 31, 32], and 4D light field imaging [33], flat [29] and miniature microscopy [32]. The original diffuser was an off-the-shelf Gaussian pseudo-random phase mask with 100% fill factor, placed directly in front of the sensor. However, the resulting PSFs had substantial background light, which amplifies noise during deconvolution. Moreover, these previous works are based on heuristic designs, lacking a theoretical and numerical framework to design around relevant target system specifications like resolution, depth of field, and field of view. This prevents such approaches from being broadly useful as optical elements in compound system designs. Here, I use a designed diffuser made of randomly-located microlenses to focus light into high-contrast random multi-focal spots, providing good SNR across a large depth range, while maintaining the randomness of the PSF. The system performance can be controlled and predicted based on the aperture and focal length of the lenslets.

1.2 Computational Imaging System Design

Singe-shot 3D imaging aims to use optical elements to encode 3D information into a 2D measurement, then recovers the 3D information computationally. Methods can be bro-

ken down into three main categories: surface depth recovery, source location recovery, and voxelized intensity recovery.

Surface depth recovery solves a semantic segmentation problem by assigning depth information to each pixel in an image. This problem attracts a lot of attention from the self-driving automobile sector. Techniques such as LiDAR, radar, monocular, and stereo images are widely used to achieve this goal [34–38].

Source location recovery's primary application is to localize sparse fluorescence particles at a high resolution [17, 19, 39]. An estimation problem is solved to determine the location of sub-diffraction-limited resolution point sources and reconstruct a point cloud. There exist many numerical methods to solve the estimation problem ranging from maximum likelihood estimators [40, 41] to deep neural networks [42, 43].

Voxelized intensity recovery methods [5, 28, 29, 32, 44] assign an intensity at each voxel in a 3D volume from a single 2D measurement, offering superior temporal resolution over scanning-based methods. It is important to note that voxelized volume reconstruction requires substantially more computer memory than depth maps and point clouds do. This dissertation falls into this category.

For single-shot 3D imaging, many design methodologies have been used [5, 7, 19, 29, 32, 41, 43–57]. On one end of the continuum are heuristic designs based on first principles [5, 19, 29, 44, 45]. In the middle are merit function driven design; for example, optimizing for a specific imaging task based on a system's point spread function's (PSF) properties such as minimizing the PSF's off-axis auto-correlation and axial cross-coherence [32]. These methods do not optimize based on the final volume produced by the pipeline [32, 40, 41]. Finally, on the far end of the continuum are end-to-end design methods. These methods optimize directly on the output volume through the use a differentiable optical simulator in conjunction with a differentiable reconstruction algorithm, then apply backpropagation to update both the imaging system's and reconstruction algorithm's parameters. By customizing the loss function, one can directly optimize for a given task such as (but not limited to) classification, localization, and reconstruction. The differentiable reconstruction algorithm is typically either a deep network [43, 46–55, 58, 59] or a physics-based unrolled network [56, 57]. Physics-based methods inject more prior information about the system's PSF and have fewer learnable parameters than deep methods, allowing for training on smaller datasets with better generalization properties. This dissertation takes this last approach.

1.3 Dissertation Outline

Computational imaging involves simultaneously designing optical hardware and reconstruction software. Such a co-design framework brings together the best of both worlds for an imaging system. The goal is to develop a high-speed, high-resolution, and large field-of-view microscope that can detect 3D fluorescence signals from single image acquisition. In this dissertation, I propose Fourier DiffuserScope, a single-shot 3D fluorescent microscope that uses a phase mask (diffuser with random microlenses) in the Fourier plane to encode 3D

information, then computationally reconstructs the volume by solving a sparsity-constrained inverse problem or applying a pre-trained neural network.

In this dissertation, I will discuss the design principles of Fourier DiffuserScope from 3 perspectives: first-principles optics, compressed sensing theory, and physics-based machine learning. First, in the heuristic design, the phase mask consists of randomly placed microlenses with varying focal lengths; the random positions provide a larger field-of-view compared to a conventional microlens array, and the diverse focal lengths improve the axial depth range. I then build an experimental system that achieves $< 3\,\mu m$ lateral and $4\,\mu m$ axial resolution over a $1000 \times 1000 \times 280\,\mu m^3$ volume. Lastly, my collaborator and I use a differentiable forward model of Fourier DiffuserScope in conjunction with a differentiable reconstruction algorithm to jointly optimize both the phase mask surface shape and the reconstruction parameters. We validate our method on 2D and 3D single-shot imaging, where our learned diffuser demonstrates improved reconstruction quality compared to previous heuristic designs.

The rest of this thesis is organized as follows. In Chap.2, I start from the working principles of Fourier DiffuserScope, followed by first-principles analysis and experimental results. Chap.3 focuses on the optics design of Fourier DiffuserScope, including the Zemax modeling, wave-optics modeling, and details on phase mask manufacturing. I present the simulation results using the wave-optics model, which successfully predicts the performance of the experimental setup. Chap.4, on the other hand, is centered on the design of algorithms, with each algorithm tailored for a specific imaging condition. I introduce Fourier neural networks for 2D plane reconstruction, iterative optimization for 3D volume reconstruction, and sparse and low-rank decomposition for 4D space-time reconstruction. Chap.3 and Chap.4 only tackle either the hardware part or the software part of the system, while Chap.5 brings these two parts together. We use a differentiable forward model of single-shot 3D microscopy in conjunction with an invertible and differentiable reconstruction algorithm to jointly optimize both the diffuser surface shape and the reconstruction parameters. We validate our method on 2D and 3D single-shot imaging, where our learned diffuser demonstrates improved reconstruction quality compared to previous heuristic designs. Finally, in Chap.6, I conclude this thesis by stating the current challenges and pointing out several directions for future exploration.

Chapter 2

Fourier DiffuserScope: Single-Shot 3D Microscopy with a Diffuser

2.1 Introduction

Calcium imaging is a powerful technique for time-resolved optical monitoring of neural activity *in vivo*. To monitor large numbers of neurons, imaging systems must record 3D volumes at tens-of-frames per second. Many existing 3D microscopes use scanning mechanisms (e.g. two-photon, confocal, light sheet), which must trade off temporal resolution and field-of-view (FoV). Light field microscopy (LFM)[1–8] uses a microlens array in front of the sensor to capture 3D information in a single shot without any moving parts. Unfortunately, the 3D capability of LFM comes with a significant loss of lateral resolution at the focal plane.

The resolution of LFM can be improved without requiring multiple measurements [60] by taking a deconvolution approach to single-shot image reconstruction [6, 61]. In this case, the captured 2D measurement is used to directly solve for the 3D object, without the intermediate step of calculating a 4D light field. The method makes an implicit assumption of no occlusions, which is valid for most fluorescence microscopy applications. Deconvolution LFM can achieve significantly better (nearly diffraction-limited) resolution at some depth planes, but its performance degrades quickly with depth, even with wavefront coding [62]. Besides suffering from non-uniform resolution, deconvolution LFM incurs artifacts at the native focal plane and requires a computationally-intensive spatially-varying deconvolution procedure. These artifacts and the resolution loss can be mitigated by placing the MLA in an off-focus plane [63–65], but the spatial variance and resolution inhomogeneity remain.

To solve some of these problems, an alternative configuration, termed Fourier light field microscopy (FLFM), places the MLA at the Fourier (pupil) plane of the objective, with the sensor one microlens focal length away [66–69]. This effectively splits the 2D sensor into a grid of sub-images, with each microlens imaging the sample from a different perspective angle. FLFM achieves more uniform resolution near the native focal plane and has a spatially-invariant point spread function (PSF) for improved computational efficiency. However, the

fundamental trade-off between spatial and angular sampling remains (unless a camera array is used, greatly increasing cost and complexity [70]). Previous single-sensor implementations require limiting the microscope's field-of-view (FOV) in order to avoid overlap of the sub-images at the sensor [66–69]. The resolution is more homogeneous than LFM, but still degrades quickly with depth.

Fourier DiffuserScope improves on FLFM by replacing the regular MLA with a diffuser consisting of randomly-spaced multi-focal microlenses. The new architecture has several advantages: 1) By using microlenses with multiple focal lengths [71–74], the PSF will have sharp features at a wide range of depth planes, improving the axial depth range and resolution homogeneity. 2) The randomness of the diffuser eliminates periodicity in the PSF and thus removes the ambiguities that required FOV limits in FLFM. Thus, we allow the microlens sub-images to overlap, then use compressed sensing algorithms [75, 76] to reconstruct the 3D volume without trading off volumetric FOV and depth resolution. This 'best of both worlds' scenario is possible only when the sample is sparse in some domain, as is generally true in fluorescence microscopy. The resulting system achieves uniform resolution over a large volume, with imaging speed limited only by signal strength or camera frame rate.

Fourier DiffuserScope is a variant of the previous methods for diffuser-based imaging with different architectures [28–33]. Here, I provide the first theoretical framework for Fourier DiffuserScope design with given performance metrics (e.g. resolution, volumetric FOV), and I directly compare with FLFM. I demonstrate the advantages of both the random and multi-focal properties of the diffuser by comparing directly with FLFM and a random uni-focal design. Finally, I build an experimental system, designed in Zemax OpticStudio, that achieves 2-3 µm lateral and 4 µm axial resolution over a $1000 \times 1000 \times 280\,\mu m^3$ volume. I use the system to record 3D videos of a freely-moving *C. elegans* nematode at 25 fps.

In summary, the contributions of this work are:

- The first cohesive theory to predict system performance (e.g. resolution, volumetric FOV) from the diffuser parameters (e.g. number of lenslets, focal lengths), enabling design of custom imaging systems.
- The first theoretical and numerical comparison between a uni-focal microlens array, a random multi-focal microlens diffuser, and an intermediate configuration with a random uni-focal microlens diffuser, showing that the random multi-focal microlens diffuser is able to provide bigger imaging volume and more uniform resolution.
- Demonstration of these principles in an experimental system, featuring 2-3 µm lateral and 4 µm axial resolution over a $1000 \times 1000 \times 280\,\mu m^3$ volume.

2.2 System Overview

Fourier DiffuserScope consists of a diffuser (a phase mask with randomly-located multi-focal microlenses) in the Fourier plane of a microscope objective, with the sensor placed

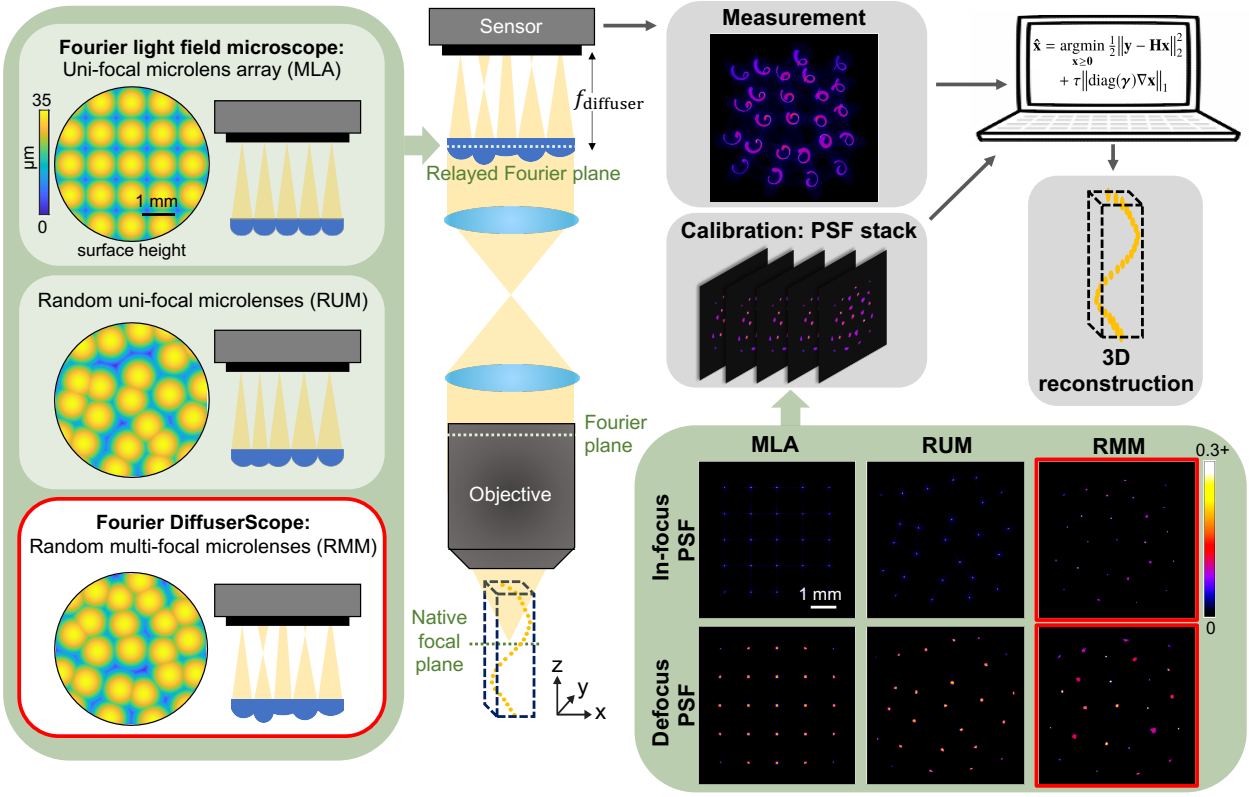


Figure 2.1: System overview for Fourier DiffuserScope.A diffuser or microlens array is placed in the Fourier plane of the objective (relayed by a $4f$ system) and a sensor is placed one microlens focal length after. From a single 2D sensor measurement, together with a previously calibrated point spread function (PSF) stack, 3D objects can be reconstructed by solving a sparsity-constrained inverse problem. Here, I compare three choices of diffuser/microlens array: Fourier light field microscopy(FLFM) with a uni-focal microlens array (MLA), random uni-focal microlenses (RUM) and Fourier DiffuserScope with random multi-focal microlenses (RMM). The in-focus PSFs represent the sensor measured intensity with a point source at the native focal plane of the objective; the defocus PSFs is captured when the point source moves 100 μm towards the objective from the native focal plane. For both uni-focal designs, all the microlenses are in-focus or out-of-focus simultaneously, while the RMM design provides a non-periodic PSF with different spots coming into focus at different planes, enabling 3D reconstructions with full FOV and high resolution across a wider depth range. Note that the PSF images (bottom right) are shown with a gamma correction of 0.4 for better visibility.

after, spaced by the average focal length of the diffuser (Fig. 2.1). Because the actual Fourier plane of the objective is physically inaccessible, we insert a relay system to image its pupil plane onto the diffuser. For each point emitter in the object space, the diffuser will produce a unique multi-spot PSF on the sensor. As compared to Gaussian diffusers or highly-scattered speckle patterns, the diffuser PSF concentrates light onto fewer pixels in order to improve SNR, while also ensuring that different points come into focus at different depths. Because the PSF is distributed and different for each point location within the 3D space, it is possible to reconstruct the whole volume from a single measurement with compressed sensing algorithms.

To model the image formation process, we divide the 3D volume into 2D slices, where each slice corresponds to a single depth plane. Neighboring slices are separated axially by less than half the axial resolution. The experimental system is designed to ensure that the system PSF (the sensor measurement resulting from a single point source) for each depth can be modeled as shift-invariant. Thus, the measurement contribution from each 2D plane is the convolution between the object slice at that depth and the PSF at that depth. The PSFs for different depths have different sizes and different microlenses come into focus, such that each depth has a unique PSF. The final sensor measurement is the sum of the contributions from each 2D layer, assuming that the light from different fluorescent sources is mutually incoherent and there are no occlusions:

$$\mathbf{y} = \sum_z \mathbf{h}_z * \mathbf{x}_z = \mathbf{H}\mathbf{x}. \tag{2.1}$$

Here, $\mathbf{y}$ is the intensity measurement on the sensor, $\mathbf{h}_z$ is the measured PSF at depth z (acquired during calibration), $\mathbf{x}_z$ is the object intensity at depth z, and $*$ represents 2D convolution over the transverse dimensions. Since this is a linear operation, we can write the model in matrix form where vector $\mathbf{x}$ is a vector representing the entire 3D volume and $\mathbf{H}$ is a matrix with columns containing the calibration measurements from every depth. Because the system is shift-invariant at each depth, we can compute $\mathbf{H}\mathbf{x}$ computationally efficiently using FFT-based convolutions. The forward model in Eq. 2.1 defines the data fidelity term of our inverse problem. Because we solve for 3D from a single 2D measurement without reducing the number of lateral pixels in the reconstruction, the inverse problem is under-determined. We solve it by using a compressed sensing algorithm that leverages the multiplexed nature of our measurements and assumes the sample is sparse in some domain. This sparsity-constrained inverse solver can be written as:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\geq\mathbf{0}}{\arg\min}\, \tfrac{1}{2}\left\|\mathbf{y} - \mathbf{H}\mathbf{x}\right\|_2^2 + \tau\left\|\mathrm{diag}(\boldsymbol{\gamma})\nabla\mathbf{x}\right\|_1. \tag{2.2}$$

Here, $\|\cdot\|_2^2$ is the data fidelity term, $\|\cdot\|_1$ is a regularization term that enforces sparsity, and τ is a tuning parameter related to the sparsity level. We use 3D Total Variation (TV) sparsity, with $\nabla = [\nabla_x, \nabla_y, \nabla_z]^\intercal$ being the gradient operator [77]. Since most fluorophores are isotropic in shape, while the resolution of our system is not necessarily isotropic, we add

a weighting vector $\boldsymbol{\gamma} = (\gamma_{xy}, \gamma_{xy}, \gamma_z)$ so that the TV value in the lateral and axial directions are weighted differently.

The system architecture for our Fourier DiffuserScope is essentially the same as FLFM, except that we use random multi-focal microlenses (RMM) instead of a uni-focal MLA. To demonstrate the advantages of the RMM over MLA, we theoretically and numerically compare their performance in Chap. 3. To separate out the effects of randomness and multi-focal, we also compare to random uni-focal microlenses (RUM). In experiment in Chap. 2.4, we demonstrate reconstruction of a 50-layer 3D volume from a single measurement of 2048×2048 pixels with 2-3 µm lateral and 4 µm axial resolution, which boosts the data size from 4.2×10^6 pixels to 2.1×10^8 voxels.

2.3 First-Principles Optical System Analysis

In this section, I will derive the relationship between diffuser design and system performance in terms of lateral resolution, axial resolution, FOV and depth range. The diffuser is characterized by the following parameters: the average microlens pitch p, the number of microlenses on the diffuser N^2 (giving an average of N microlenses in each transverse direction), the minimum focal length f_{min}, the maximum focal length f_{max} and the average focal length f_{ave} of the microlenses. We compare three different phase masks (MLA, RUM and RMM) to be placed in Fourier configuration. All three designs have the same size and number of microlenses, but the locations and focal-length distributions are different. The MLA and RUM microlenses all have a single focal length f_{ave}, whereas the RMM microlenses all have different focal lengths, varying between f_{min} and f_{max}. The minimum and maximum focal lengths are designed to focus at the closest and furthest depth planes within the volume-of-interest. The rest focus at depth planes evenly spaced within that range, which means their focal lengths are dioptrically distributed between f_{min} and f_{max}. The system schematic and parameter definitions are shown in Fig. 2.2 and Table 2.1.

Lateral Resolution

In Fourier configuration, each microlens forms a perspective view of the object. Consider the middle microlens in Fig. 2.2, which collects light from the yellow region, with acceptance angle α, from an in-focus point source (the orange dot in object space) and forms a diffraction-limited spot on the sensor. Other bundles of light from the same point source will reach other microlenses, focusing to separate spots on the sensor. With the MLA, these spots will form a grid, whereas with the RUM or RMM, they will form a random set of points at the sensor. The in-focus lateral resolution is determined by the size of the diffraction-limited spot beneath a single microlens, which is determined by the effective numerical aperture (NA), or the acceptance angle α, of the microlens sub-aperture. Since the the back pupil of the objective is divided into N microlenses in each direction, the effective NA (under

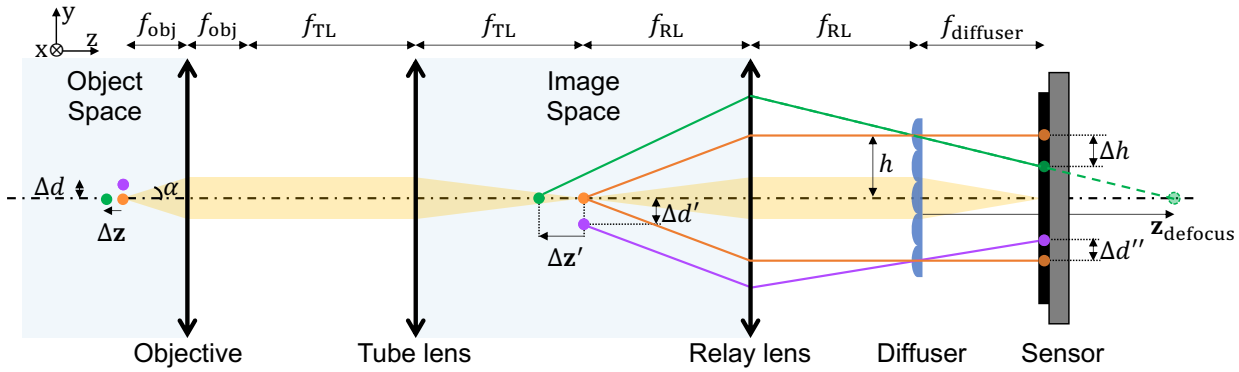


Figure 2.2: First-principles system performance analysis. To analyze lateral resolution, consider the orange and purple point sources, laterally spaced by Δd in object space, with images on the sensor spaced by $\Delta d''$. For axial resolution, consider the orange and green point sources, axially separated by $\Delta \mathbf{z}$ in object space, which map to their images on the sensor spaced by Δh . The axial resolution is determined by the minimum resolvable separation on the sensor. $\Delta \mathbf{z}$ pointing to the left has a negative value.

paraxial approximation) is the objective NA divided by N:

$$\mathrm{NA}_{\mathrm{eff}} = \frac{\mathrm{NA}_{\mathrm{obj}}}{N} \tag{2.3}$$

The diffraction-limited lateral resolution is given by the Rayleigh criterion:

$$R_{\mathrm{lateral}} = \frac{1.22\lambda}{2\mathrm{NA}_{\mathrm{eff}}} = \frac{1.22\lambda N}{2\mathrm{NA}_{\mathrm{obj}}}. \tag{2.4}$$

For the RUM and RMM, the aperture size of each microlens varies, so we determine expected resolution by the average sub-aperture size, which is designed to match the MLA effective NA, in order to compare the two situations fairly.

To achieve the diffraction-limited resolution derived above, the sensor pixel spacing must be small enough to Nyquist sample the pattern after taking into account magnification. To quantify this requirement for our Fourier configuration, consider two point sources laterally separated by Δd (the orange and purple dots in the object space of Fig. 2.2). After the $4f$ system of the objective and the tube lens, their intermediate images will be spaced by $\Delta d' = (f_{\mathrm{TL}}/f_{\mathrm{obj}})\Delta d$. Using similar triangles between the relay lens, the microlens plane and the sensor, the distance between the two chief rays on the sensor is $\Delta d'' = (f_{\mathrm{ave}}/f_{\mathrm{RL}})\Delta d'$. Together, we have $\Delta d'' = M\Delta d$, where M is the lateral magnification rate from the object

λ	optical wavelength	N	(average) number of microlenses
n_r	refractive index of the medium		in one dimension
f_{obj}	focal length of the objective	f_{max}	maximum focal length of RMM
NA_{obj}	numerical aperture of the objective	f_{min}	minimum focal length of RMM
D	diameter of the objective back pupil	f_{ave}	(average) focal length of the
FOV_{obj}	diameter of the objective FOV		(multi-focal) diffuser
f_{TL}	focal length of the tube lens	s	sensor pixel size
f_{RL}	focal length of the relay lens	M	system magnification
p	(average) microlens pitch	NA_{eff}	effective numerical aperture

* FOV: field-of-view. RMM: random multi-focal microlenses.

Table 2.1: Parameter definitions for the optical system.

space to the sensor:

$$M = \frac{f_{\text{TL}}}{f_{\text{obj}}} \frac{f_{\text{ave}}}{f_{\text{RL}}}. \tag{2.5}$$

Thus, the pixel size s satisfies Nyquist sampling when $s \leq MR_{\text{lateral}}/2$.

Because we reconstruct 3D information, we also investigate how lateral resolution changes for objects away from the objective's native focal plane. For MLA and RUM, in which all microlenses have the same focal length, the minimum resolvable spot is determined by the circle of confusion; we define off-focus lateral resolution to be the radius of the circle of confusion, Δc. To derive the off-focus resolution in our setup, we first calculate the defocus distance of the intermediate image for an off-focus point source (the green dot in object space in Fig. 2.2), which is scaled by the objective's magnification: $\Delta \mathbf{z}' = (f_{\text{TL}}/f_{\text{obj}})^2 \Delta \mathbf{z}$. Then, by applying the Newtonian form of the thin lens equation for the relay lens, we calculate the location of the second intermediate image of the green point, relative to the diffuser, after passing through the relay lens: $\mathbf{z}_{\text{defocus}} = -f_{\text{RL}}^2/\Delta \mathbf{z}'$. This serves as the 'object' for the diffuser microlenses and $\mathbf{z}_{\text{defocus}}$ is the 'object distance'. So, the circle of confusion size depends on $\mathbf{z}_{\text{defocus}}$, the diffuser focal length f_{ave} and the size of a single microlens p. The resulting expression describes how the lateral resolution degrades linearly with defocus distance:

$$\Delta c = \frac{p\, f_{\text{ave}}}{2} \frac{1}{|\mathbf{z}_{\text{defocus}}|} = \frac{p\, f_{\text{ave}}}{2} \frac{f_{\text{TL}}^2 |\Delta \mathbf{z}|}{f_{\text{RL}}^2 f_{\text{obj}}^2}. \tag{2.6}$$

The primary advantage of the RMM that we use in Fourier DiffuserScope is that it has multiple focal lengths. Thus, the subset of microlenses that are in focus at each depth will have spots with size matching the in-focus diffraction-limited lateral resolution derived in the previous section. Hence, the lateral resolution does not degrade with depth within the volume-of-interest. When the object moves beyond the designed range, the lateral resolution will degrade linearly with defocus distance. A detailed analysis on the depth range is in a later section within this chapter.

Axial Resolution

We define the axial resolution as the minimum axial distance between two point emitters that can be resolved in the reconstruction. The off-axis microlenses have disparity, such that point sources from different depths are imaged to different lateral locations on the sensor; two points will be resolved if their images are separated by at least the diffraction-limited lateral resolution (after magnification). We analyze the limits for the outermost microlens, which has the largest disparity angle. In Fig. 2.2, the center of the topmost microlens is $h = (N-1)/2 \cdot p$ away from the optical axis. Two point sources with the same lateral location are axially spaced by $\Delta\mathbf{z}$ (the orange and green dots in object space, Fig. 2.2). In the previous section we have already related $\Delta\mathbf{z}$ to $\mathbf{z}_{\text{defocus}}$. From the similar triangles formed by the relay lens, the diffuser and the sensor, we can calculate the lateral distance between the orange chief ray and the green chief ray on the sensor, $\Delta h = (f_{\text{ave}}/|\mathbf{z}_{\text{defocus}}|)h$. The minimum distance on the sensor for resolving the points is MR_{lateral}, which sets the minimum value for Δh, and the value of $\Delta\mathbf{z}$ we solve for is the axial resolution R_{axial}. Given the relation between the relayed pupil diameter and numerical aperture $N \cdot p = (f_{\text{RL}}/f_{\text{TL}})2\text{NA}_{\text{obj}}f_{\text{obj}}$, the axial resolution is:

$$R_{\text{axial}} = \frac{N}{N-1}\frac{1}{\text{NA}_{\text{obj}}}R_{\text{lateral}}. \tag{2.7}$$

When the pixel sampling is sufficient, such that the lateral resolution is diffraction-limited in Eq. 2.4, the diffraction-limited axial resolution becomes:

$$R_{\text{axial}} = \frac{N^2}{N-1}\frac{1.22\lambda}{2\text{NA}_{\text{obj}}^2}. \tag{2.8}$$

The axial resolution off-focus is determined in a similar way, except that the two point images must now be separated by a distance of at least the circle of confusion size. We derive this by replacing the R_{lateral} term in Eq. 2.7 with the radius of the circle of confusion Δc in Eq. 2.6. Note that the slope of the defocused axial resolution as a function of depth is proportional to that of defocused lateral resolution.

Field-of-View

We analyze the in-focus FOV for each of the three microlens designs, and assume that the FOV throughout the volume will be approximately the same as that at the native focal plane. The regular layout of the MLA results in a periodic PSF. When a point in the scene moves laterally by an amount that shifts the PSF by an integer number of pitches, the shifted PSF is nearly the same as the unshifted one; this creates ambiguities that cause the deconvolution to fail. To avoid this problem, a field stop is inserted to guarantee that the PSF shifts by less than one period over the FOV [67, 68]. The resulting FOV for the MLA-based FLFM is thus limited by the microlens pitch size:

$$\mathrm{FOV}_{\mathrm{MLA}} = \frac{p}{M}. \tag{2.9}$$

The randomly located lenses in the RUM and RMM create PSFs with randomly-located spots that do not suffer from the ambiguity caused by periodicity. So, both RUM and RMM are able to reconstruct images with the full objective FOV, giving

$$\mathrm{FOV}_{\mathrm{RUM}} = \mathrm{FOV}_{\mathrm{RMM}} = \mathrm{FOV}_{\mathrm{obj}}. \tag{2.10}$$

This is based on ideal optics; in reality, aberrations can break the shift-invariance of the PSF in the peripheral FOV so that the final reconstruction has a smaller FOV or reduced resolution near the edges. In practice, we determine the FOV for by calculating the similarity between on-axis and off-axis PSFs, described in more detail in Chap. 3.5.

Depth Range

The depth range describes the axial distance over which the object can be reconstructed with diffraction-limited resolution. For the uni-focal designs (MLA and RUM), the depth range is simply the depth-of-field (DOF) of a single microlens, since all microlenses have the same focal length. The microscope DOF expression is the sum of a wave optics term and a geometrical optics term [78], and we use the effective NA to account for the partitioning of the back pupil plane into multiple microlenses:

$$\mathrm{DOF}_{\mathrm{microlens}} = \frac{\lambda n_r}{\mathrm{NA}_{\mathrm{eff}}^2} + \frac{n_r \cdot s}{M \cdot \mathrm{NA}_{\mathrm{eff}}}. \tag{2.11}$$

The main advantage of using multi-focal microlenses in the RMM for Fourier DiffuserScope is that the depth range will be much larger, since the DOFs of different microlenses are offset. The RMM can be designed for the largest possible depth range by ensuring that the focus positions of different microlenses are separated axially by their DOF; thus, the upper bound of the depth range is the product of a single microlens' DOF and the number of microlenses.

$$\text{Depth Range}_{\mathrm{RMM}} \leq N^2 \times \mathrm{DOF}_{\mathrm{microlens}} \tag{2.12}$$

To design such a RMM to cover a depth range from $-z$ to $+z$, the maximum and minimum focal lengths are designed to focus on the farthest and nearest depth planes:

$$\begin{aligned}\frac{1}{f_{\max}} &= \frac{1}{f_{\text{ave}}} - \left(\frac{f_{\text{TL}}}{f_{\text{RL}} f_{\text{obj}}}\right)^2 \cdot (-z),\\ \frac{1}{f_{\min}} &= \frac{1}{f_{\text{ave}}} - \left(\frac{f_{\text{TL}}}{f_{\text{RL}} f_{\text{obj}}}\right)^2 \cdot (+z).\end{aligned} \tag{2.13}$$

The remaining focal lengths are dioptrically distributed between $f_{\min}$ and $f_{\max}$ to provide equally-spaced focus planes in the object space. In practice, because the microlenses have different sizes and shapes, there will be variation in the resolution of different microlenses. To ensure stability of performance, we design the DOFs to overlap by half, such that the depth range covers half of its upper bound.

2.4 Experimental Results

Experimental Setup

We build an experimental Fourier DiffuserScope system using the RMM design, with a 20×, 1.0 NA objective lens (Olympus XLUMPLFLN20XW water immersion). The system configuration and pictures are shown in Fig. 2.3. The fluorescent sample is excited with blue light from a Xenon lamp light filtered by a band-pass emission filter (Semrock FF01-474/27). The emitted green light is filtered using a dichroic mirror (Semrock FF495-Di03) and an emission filter (Semrock FF01-520/35). Since the back pupil diameter is larger than the sensor size (Andor Zyla 4.2, sensor size 13.3 × 13.3 mm, pixel size 6.5 µm), we need to demagnify the pupil using a relay lens paired with the tube lens. The size requirement is given by

$$\frac{f_{\text{RL}}}{f_{\text{TL}}} D + M \times \text{FOV}_{\text{obj}} \leq 13.3\,\text{mm} \tag{2.14}$$

hence the demagnification ratio $f_{\text{RL}}/f_{\text{TL}}$ should be at least $1:3$. We choose a demagnification rate of 3.75× so that the full FOV can be recorded and the sensor pixels are mostly utilized. Since the Olympus tube lens has a focal length of $f_{\text{TL}} = 180\,\text{mm}$, the relay lens focal length needs to be $f_{\text{RL}} = f_{\text{TL}}/3.75 = 48\,\text{mm}$. The relay lens design is optimized using Zemax OpticStudio to reduce aberration (see Chap. 3.2).

To fabricate our RMM diffuser, we make a negative mold by randomly indenting polished copper using ball bearings with diameters ranging from 10 mm to 16 mm. We then use polydimethylsiloxane (PDMS) to make a replica of the mold with convex-plano microlenses. This approximately achieves our diffuser design parameters of average focal length of 15.6 mm (after considering the relay system), with minimum and maximum focal lengths of 12.3 mm and 21.4 mm, respectively, giving a $\sim 200\,\mu\text{m}$ depth range using Eq. 2.13. The main fabrication errors come from deformation error during indentation and shrinkage of

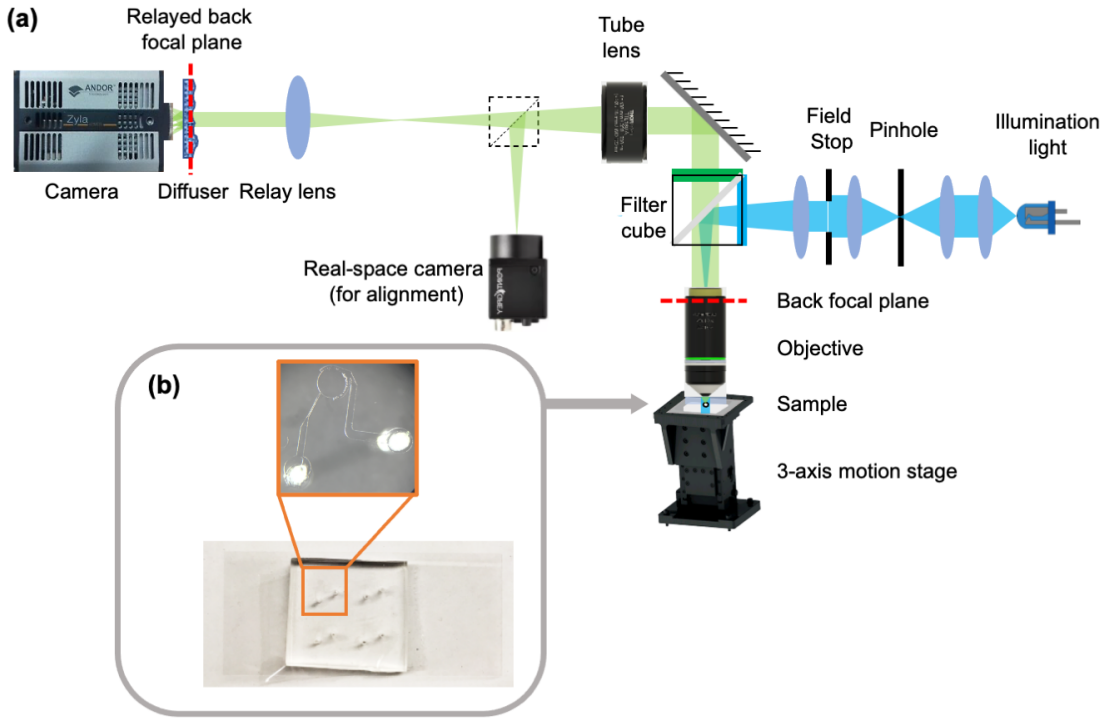


Figure 2.3: Experimental setup configuration. (a) Experimental setup configuration: a diffuser (Random multi-focal microlenses) in front of a sensor, placed in the Fourier plane, encodes depth information in a single exposure. A relay lens (custom lens group designed in Zemax) demagnifies the pupil while maintaining a shift-invariant PSF. Blue excitation light is aligned as epi-illumination for easier sample handling. There is an optional real-space camera port in the system for alignment purposes. (b) Sample holder: microfluidic chip for *in vivo C. elegans* imaging. The *C. elegans* will be injected into a $1000 \times 1000 \times 100\,\mu m^3$ arena which constrains the live worm to move within the FOV of the objective.

the PDMS material. These have opposite effects, since the indented deformation will have bigger diameter than the indenter while the material shrinkage gives smaller diameter, so they offset each other to some extent. More details on different fabrication methods will be discussed in Chap. 3.3.

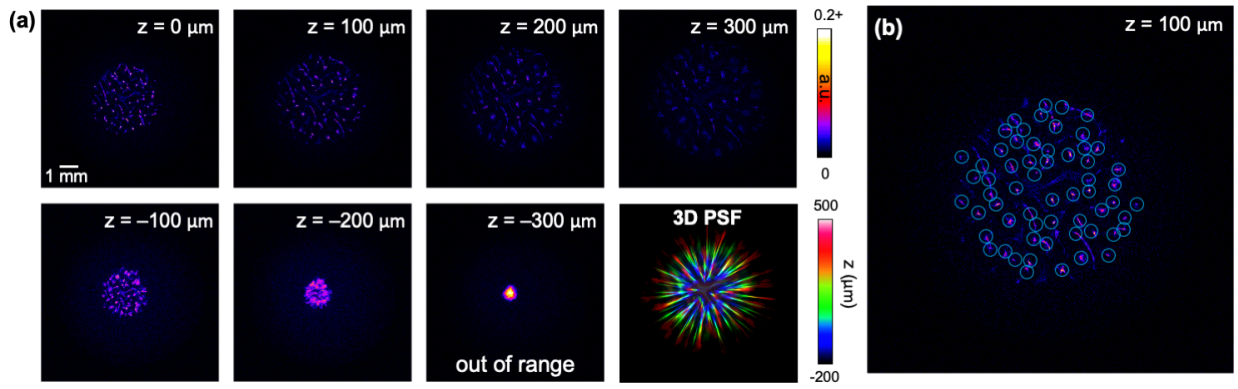


Figure 2.4: Experimental 3D PSF stack. (a) Sample PSFs measured with a point source from $z = -300\,\mu\text{m}$ to $z = 300\,\mu\text{m}$ depth planes, as well as the 3D PSF plotted with different depth planes color-coded according to the color bar. Positive z direction means moving towards the objective. The region more than 200 µm below the native focal plane is out of range since the overall PSF becomes so small that the neighboring microlenses will merge into each other. (b) In the measured PSF, every focal spot noted in a blue circle corresponds to a small microlens. There are ~ 60 microlenses in the illuminated region of the diffuser.

Calibration

Fabrication errors should be accounted for during calibration, such that they do not cause model mismatch during deconvolution. The calibration PSF stack is acquired by using a sub-resolution fluorescent bead as a point source. We use the 1 µm 468/508 nm green bead from Thermo Fisher Fluoro-Max product line [79] because it has a similar excitation and emission wavelength as GFP and its brightness is improved for sensitivity and detectability. We place a single bead at different depths, controlled by a motorized stage (Fig. 2.3), then record the intensity with a sensor located at the average back focal plane of the diffuser.

The 3D PSF stack is shown in Fig. 2.4. In total, 350 PSF images were recorded with a 2 µm axial increment from 200 µm below the native focal plane to 500 µm above. When the point source moves more than 200 µm below the native focal plane, the overall PSF becomes so small that the out-of-focus blur from neighboring microlenses will merge into each other, which causes very noisy reconstruction, so we avoid placing objects in this region. In the measured PSF, every focal spot corresponds to a small microlens on the diffuser. We manually find and circle the sharp spots in Fig. 2.4(b). There are 60 microlenses in the illuminated region of the diffuser. We apply the theory in Chap. 2.3, Eq. 2.4 and Eq. 2.7, to get the predicted lateral resolution of 2.4 µm and axial resolution of 2.8 µm.

Experimental Two-Point Resolution

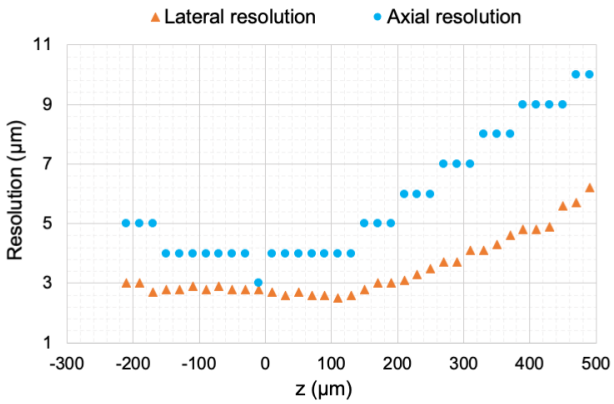


Figure 2.5: Experimentally measured resolution, defined as the minimum separation distance at which two sub-resolution fluorescent beads are resolvable. Across the 280 µm depth range from $z = -150$ µm to $z = 130$ µm, the lateral resolution is < 3 µm and the axial resolution is less than or equal to 4 µm.

We experimentally measured the two-point resolution (Fig. 2.5) in order to benchmark the resolution and depth range of our Fourier DiffuserScope prototype. For practicality, measurements were synthetically generated by summing two experimental PSFs at different locations. The image was recovered by solving Eq. 2.2, and we then calculated the smallest distance at which the two points were still resolved, both laterally and axially, at each depth z. The increment of separation distance is 0.1 µm laterally and 1 µm axially. Across a depth range of 280 µm (from $z = -150$ µm to $z = 130$ µm), the lateral resolution fluctuates between 2.5 µm and 2.9 µm and the axial resolution is mostly 4 µm, close to their theoretical predictions. The depth range is larger than the design, suggesting that the actual diameter range of the microlenses is wider than the ball bearings used.

Free-Moving *C. Elegans* Reconstruction

We next demonstrate our system on a live adult *C. elegans* organism that is pan-neuronally expressing a GCaMP fluorescent indicator. The *C. elegans* is anesthetized by levamisole in M9 buffer and then loaded into a $1000 \times 1000 \times 100\,\mu m^3$ arena on a microfluidic chip (Fig. 2.3(b)) which constrains the worm to move within the FOV of the objective. Since our method is able to reconstruct a 3D object from a single shot, the frame rate is only limited by the sensor. We recorded a raw video at 25 fps while the worm was freely moving (Fig. 2.6(c)). There is one *C. elegans* image behind every microlens and in total we see ~ 60

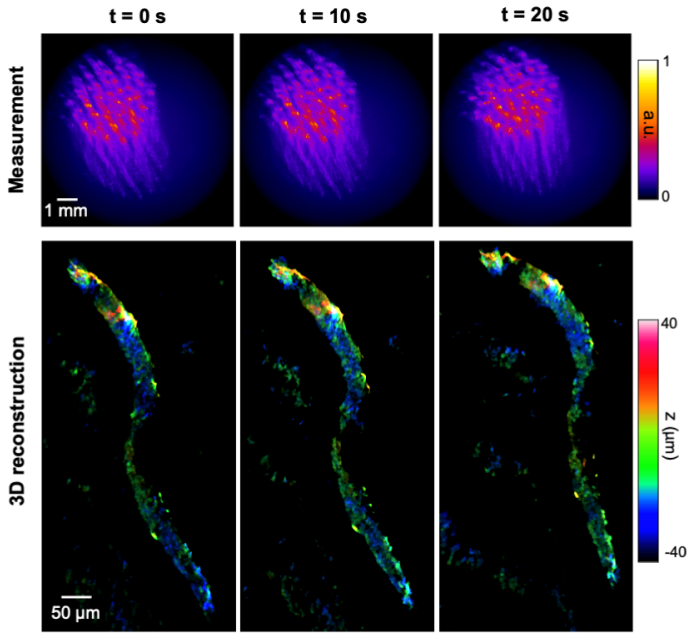


Figure 2.6: Raw measurements and 3D reconstructions of a GFP-tagged *C. elegans* recorded at 25 fps. The depth across a 80 µm range is color-coded according to the color bar.

overlapping *C. elegans* images, each from a different angle. Given that every location in the object space has a unique PSF on the sensor, we are able to deconvolve the overlapping images. Our deconvolution algorithm applies ADMM due to its fast convergence rate [80]. To save memory, we did not deconvolve the measurement with all the calibrated PSFs, instead we firstly use a coarse axial sampling of the PSFs to locate the object occupied depth range and then a small subset of fine sampling PSFs to reconstruct the object. The *C. elegans* in our raw video moves within a 80 µm depth range and we use 50 PSFs with 2 µm axial increment to cover the whole object. The reconstructed *C. elegans* from the corresponding frame is displayed in a color-coded depth image in Fig. 2.6, showing the potential of our method to locate the neurons of the whole animal simultaneously in 3D. The randomness of the diffuser also enables compressed sensing reconstructions with more voxels in the 3D result than pixels on the sensor. From a 4.2 mega pixel sensor, the reconstructed *C. elegans* volume contains 50× more voxels and the gain could increase to 140× if we deconvolve with all the available PSFs within the 280 µm depth range. With regards to the resolvable voxels, for our experimental system the lower bound equals the imaging volume divided by the worst lateral (2.9 µm) and axial (4 µm) resolution, which gives 10 mega resolvable voxels per frame.

Chapter 3

Optics Design of Fourier DiffuserScope

3.1 Introduction

In the previous chapter, I analyzed the performance of Fourier DiffuserScope, assuming that each lens is an ideal thin lens without aberration. However, in reality, each lens has a finite thickness, and we must carefully evaluate the aberration before building an experimental setup.

In this chapter, I present the optical design behind the scene. Fourier DiffuserScope is built on top of a traditional fluorescence microscope, by adding a relay lens to relay the pupil onto a diffuser. I choose a 20X, 1.0NA Olympus life-science objective because of its big FOV (1 mm) and high resolution, paired with a 180 mm focal length tube lens to match Olympus industry specs. The high-end industrial objective and tube lens are corrected for aberration, so we only need to design the optical elements behind the tube lens, namely the relay lens and the diffuser. In Chap. 3.2, I show that using an off-the-shelf relay lens will result in enormous aberration [81], and the performance can be optimized by designing a customized aberration-correction lens group using Zemax OpticStudio. Later in Chap. 3.3, I compare the diffuser locations (Fourier space vs. real space), different types of the diffuser (Gaussian diffuser, microlens array, and random multi-focal microlenses). The benefit of using random multi-focal microlenses in the Fourier space has been briefly discussed in Chap. 2. Here, I will provide a comprehensive analysis to rationalize the design and further investigate how to optimize the random multi-focal microlenses in Chap. 5. After designing the aberration-corrected Fourier DiffuserScope, in Chap. 3.4 I build a wave-optical simulation to model the light propagation from the object to the sensor, assuming no aberration. Using the optical simulator, we can predict the system performance for an arbitrary phase mask. In Chap. 3.5, I present the simulation results of system resolution, FOV and depth range. The numerical simulations match well with the previous theoretical derivations in Chap. 2.3. To end this chapter, I introduce several methods of fabricating customized phase masks in Chap. 3.6.

3.2 Optical System Design with Zemax OpticStudio

Let's examine the design of the relay lens in this section. The relayed pupil plane cannot have huge aberrations, otherwise the shift-invariant assumption does not hold. The back pupil of the objective has a diameter of $D = 2 * f_{\text{obj}} * \text{NA}_{\text{obj}} = 18\,\text{mm}$. To relay and demagnify a pupil, off-the-shelf achromatic doublets are commonly used due to their low price and convenience. However, at a pupil size of 18 mm, these achromatic doublets introduce optical aberrations at the peripheral pupil region, breaking the shift-invariance assumption. Although some companies provide carefully designed pupil-relay lenses, Thorlabs "Scan Lenses for Laser Scanning Microscopy" [82] for example, the choices of focal lengths and wavelengths are limited.

To investigate how aberration affects the system, I model the system in Zemax OpticStudio (Fig. 3.1). The objective is represented by an ideal paraxial lens since its lens data is not publicly available. The tube lens data is a black box model downloaded from Thorlabs (Thorlabs TTL180-A). The relay lens, together with the tube lens, relay a demagnified back pupil onto the diffuser plane. In a desired situation, the relayed pupil is flat and stationary for all the field points within the objective's FOV.

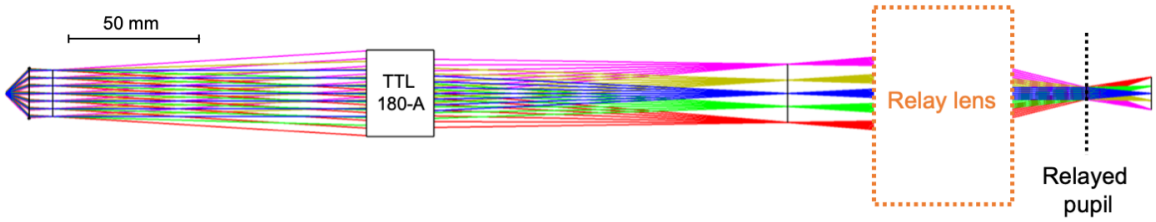


Figure 3.1: Fourier DiffuserScope model in Zemax OpticStudio. The objective is represented by an ideal paraxial lens. The tube lens data is a black box model from Thorlabs TTL180-A. The relay lens to be designed is inside the orange box.

Off-the-Shelf Achromats

There are two restrictions in choosing the relay lens: (1) the focal lengths should be < 60 mm to have the full FOV recorded based on Eq. 2.14, (2) the clear aperture is at least 30 mm to prevent vignetting. After searching the Lens Catalogs, I pick an off-the-shelf achromatic doublets from Edmund Optics (model number 89682, effective focal length 50 mm, clear aperture 39 mm). However, there is a huge amount of aberration in both the Zemax simulation and the experimental measurements.

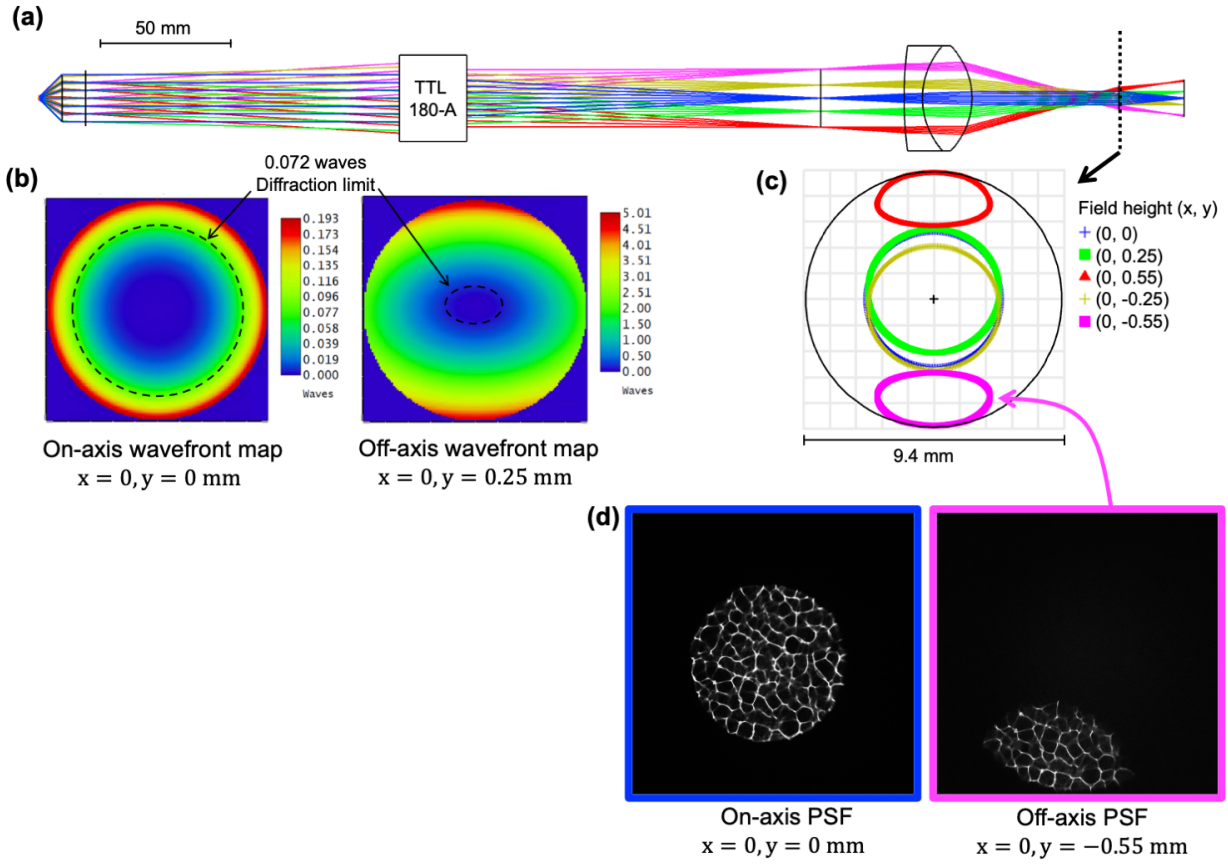


Figure 3.2: Fourier DiffuserScope model in Zemax OpticStudio using a single achromatic doublet as the relay lens. The wavefront map at the relayed pupil in (b) shows huge aberration for off-axis field point at $x = 0, y = 0.25\,\text{mm}$, with only $\sim 1/25$ of the pupil being diffraction-limited. The simulated footprints in (c) are drifting. The same drifting issue applies to the experimental measured PSF in (d), where the on-axis and off-axis PSFs contain different regions of the Gaussian diffuser. Both simulation and experimental results warn that the shift-invariance assumption does not hold.

To demonstrate, Fig. 3.2(a) shows the layout of an achromatic lens with 50 mm focal length. The relayed pupil position is located using the Zemax built-in "Thickness solve" function. Fig. 3.2(b) is the wavefront map at the relayed pupil for on-axis field point and off-axis field point at $x = 0, y = 0.25\,\text{mm}$. The ideal wavefront would be a plane wave since the object is placed at the native focal plane. However, the simulated wavefront map contains huge aberration, especially for off-axis points. I draw a dotted circle to represent a wavefront error of 0.072 waves, and only $\sim 1/25$ of the pupil is diffraction-limited for the

off-axis relayed pupil. Fig. 3.2(c) is the resulting footprint of the relayed pupil plane at different field heights. Since the phase mask is placed at the relayed pupil plane, the drifting footprints mean that different areas on the diffuser will be utilized for different field locations, which breaks the shift-invariant forward model. I build an experimental system with this off-the-shelf achromatic doublet and measured the PSFs with a 0.5 degree Gaussian diffuser in Fig. 3.2(d). The on-axis PSF contains the whole caustic pattern at the sensor center. The off-axis PSF for a point source on the edge of the FOV, on the other hand, crops the caustic pattern to half, agreeing with the simulated footprint from Zemax.

The cause of the drifting footprints is mostly due to spherical aberration. Spherical aberration is related to the nature that the lens has spherical surfaces. Most of the time, we see spherical aberration in the context of focusing (or imaging) instead of collimating. Noticing that focusing and collimating are two reciprocal light paths, the effects of spherical aberration are similar. The two situations are illustrated in Fig. 3.3.

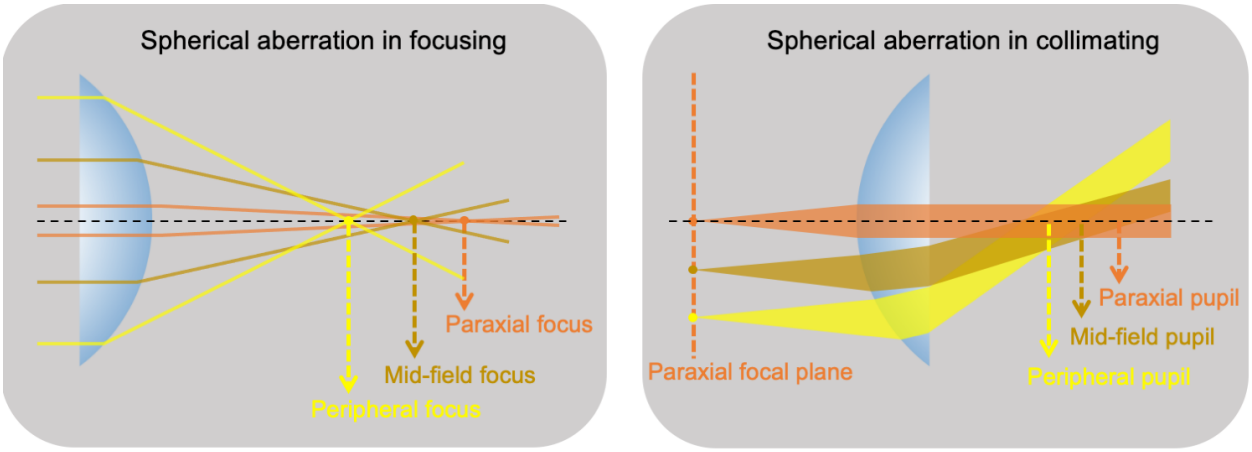


Figure 3.3: Illustration of spherical aberration in focusing and collimating. When the lens focuses a plane wave to a focal point, the marginal rays bend at a steeper angle and focus before the paraxial focus, causing a blurred point spread function. When the lens collimates point sources into plane waves, the collimated rays from the peripheral point source bend at a steeper angle, and the corresponding back pupil shift closer towards the lens, causing a drifting footprint from the paraxial back pupil plane.

In the focusing case, the lens aims to focus a plane wave, a bundle of parallel rays, to a focal spot. The paraxial rays are converged to the paraxial focus denoted by the focal length in the spec sheet. The marginal rays, however, bend at a steeper angle and focus before the desired focal length. The overall effects cause a blurred point spread function. If a camera sensor is placed at the paraxial focus, the edge of the field-of-view will appear to be out-of-focus.

In the collimating case, the lens aims to collimate point sources on the front focal plane into plane waves. The collimated beam from the paraxial point source is at a small angle to the optical axial, with the back pupil located at the denoted back focal plane on the spec sheet. The collimated beam from the peripheral point source bends at a steeper angle, and the corresponding back pupil shifts closer towards the lens. If a camera sensor is placed at the paraxial back pupil plane, the footprint of collimated beams will depend on the field location, as seen in the simulation result in 3.2(c). Moreover, the point sources at the outer zones might be vignetted or even not visible.

Customized Lens Group

Lens model	**Lens type**	**Focal length**
Edmund Optics 49-286	Achromatic Doublet	200mm
Air gap 0.5 mm		
Newport KBX163	Bi-Convex Lens	175 mm
Air gap 0.5 mm		
Thorlabs AC508-075-A	Achromatic Doublet	75 mm
Air gap 3 mm		
Newport KPC064	Plano-Concave Lens	−500 mm

Table 3.1: Relay lens prescription table.

In general, to correct a complex aberration, the lens group will need more elements to counterbalance different types of aberrations, at the price of increased cost, size, scattering, and absorption of light. Beyond achromatic doublets, we can consider scan lenses [82] and eyepieces [83] that consist of multiple lenses. In fact, the function of the relay lens in Fourier DiffuserScope is the same as the scan lens in laser scanning microscopy and the eyepiece in traditional microscopy: all these three lenses' function is to relay the back pupil of the objective. Since the choice of off-the-shelf scan lens and eyepiece are limited, I decide to design my own one. Additionally, since fabricating customized lens elements is expensive and time-consuming, to implement an aberration-free, affordable and readily available relay lens, the customized lens group will only consist of off-the-shelf elements.

As a starting point, I pick the Erfle eyepiece design due to its wide FOV, relatively long eye-relief and small distortion [84]. A typical Erfle eyepiece (Fig. 3.4(d)) contains 5 elements, with a biconvex singlet sandwiched by two doublets. Since the Erfle contains one piece of uncommon convex-concave doublet, I replace it with a convex-convex achromatic doublet

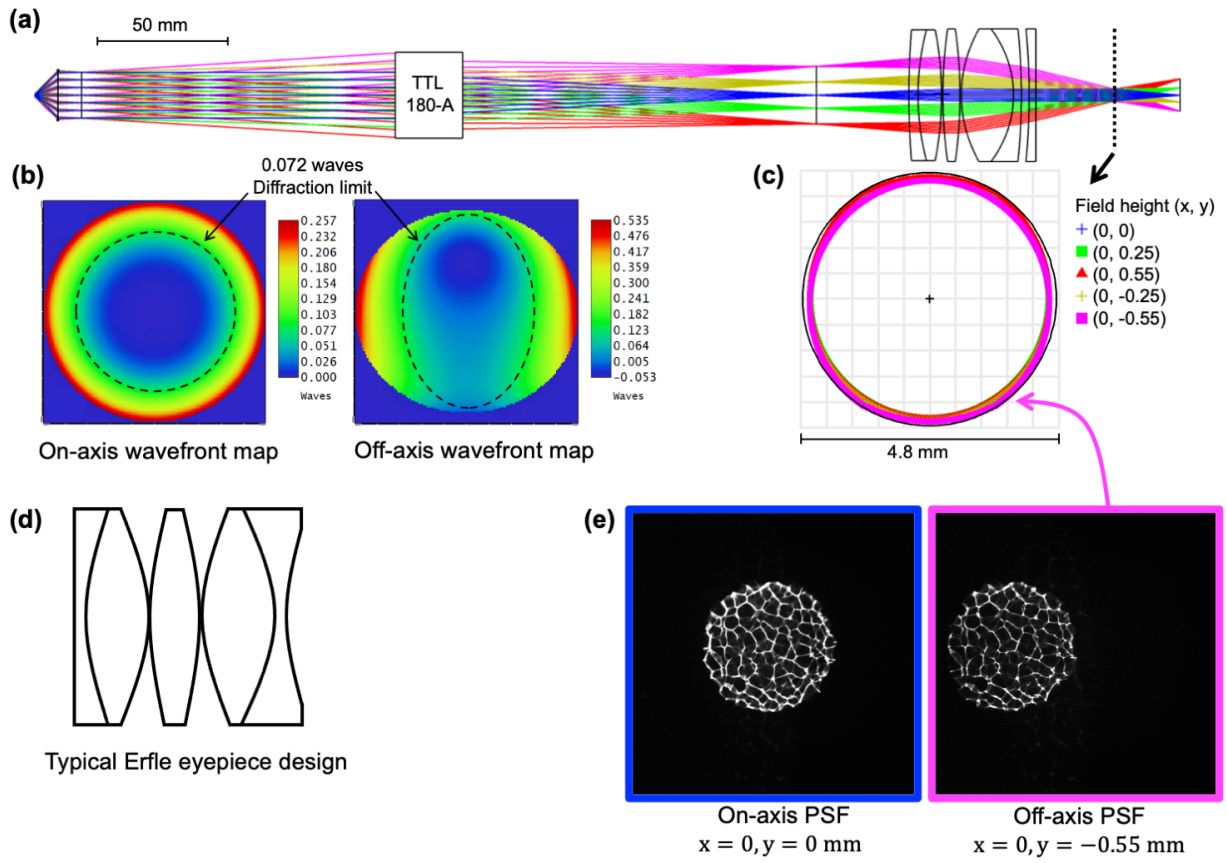


Figure 3.4: Fourier DiffuserScope model in Zemax OpticStudio using a customized lens group with off-the-shelf elements. The designed lens group is based on the Erfle eyepiece design in (d), and all the elements are from the off-the-shelf lens catalog. The wavefront map at the relayed pupil in (b) shows a much smaller aberration compared to Fig. 3.2. For off-axis field point at $x = 0, y = 0.25\,\text{mm}$, more than 1/2 of the pupil is diffraction-limited. In (c), the footprints from different field points mostly overlap. The same improvement shows in the experimental measured PSFs in (e), where the on-axis and off-axis PSF contain mostly the same region of the Gaussian diffuser. Both the simulation and the experimental results meet the shift-invariance assumption.

plus a concave-plano lens. I alternately optimize the radii of all the surfaces and replace every lens with the most similar counterpart in the catalog. I also optimize the air gap between every two components which can be controlled by using a spacers inside the lens tube. After many iterations, I arrive at the final design described in the Table 3.1. The wavefront map at the relayed pupil in Fig. 3.4(b) shows a much smaller aberration compared

to Fig. 3.2(b) . For off-axis field point at $x = 0, y = 0.25\,\text{mm}$, more than 1/2 of the pupil is diffraction-limited. In Fig. 3.4(c), the footprint of the peripheral field from the new design mostly overlaps with the center footprint and the aberration is greatly reduced. The relayed pupil size is 4.8 mm which means that the effective de-magnification rate of the back pupil plane is 3.75× and the effective focal length of the lens group is 48 mm.

3.3 Phase Mask Design and Modeling

To encode 3D information with a phase mask, the only requirement is that each voxel in the 3D object space generates a unique PSF, and thus the voxel location can be unambiguously identified during reconstruction. Under this minimum requirement, a wide variety of system designs would work, and each has its pros and cons in resolution, volumetric FOV, noise sensitivity, computation cost, and manufacture feasibility.

In this section, I analyze different systems designs under the topic of single-shot 3D imaging with a phase mask. I first compare placing the phase mask in the real space (light field microscopy [5]) versus in the Fourier space (Fourier light field microscopy [68], Fourier DiffuserScope [44]). I then compare different types of phase mask: Gaussian diffuser in DiffuserCam [28], microlens array in the light field microscopy [5] and the Fourier light field microscopy [68]), random microlenses in the flat diffuser microscope [29], and random multifocal microlenses in Fourier DiffuserScope [44] and Miniscope3D [32]. Following the analysis, I will also provide PSF comparisons between these phase masks.

Real Space vs. Fourier Space

To compare the pros and cons of placing a phase mask in the real space versus the Fourier space, I use the microlens array in the light field microscope as an example. Light field camera [85] and light field microscope [5] use a microlens array in the native image plane, or real space, to simultaneously capture 2D spatial and 2D angular information. The resulting 4D light field can be used for digital refocusing, perspective synthesis, or 3D reconstruction. The resolution of LFM can be improved by the wave-optics model [61], allowing for reconstruction of high-spatial frequencies through simulated PSF deconvolution. However, LFM with the real-space configuration has several major disadvantages:

① The PSFs are laterally spatially-varying, resulting in a computationally-intensive deconvolution procedure.

② LFM suffers from reconstruction artifacts at the native focal plane.

③ The resolution degrades rapidly with defocus.

④ The numerical aperture (NA) of the objective and the microlens array must be matched to avoid aliasing, which hinders design flexibility.

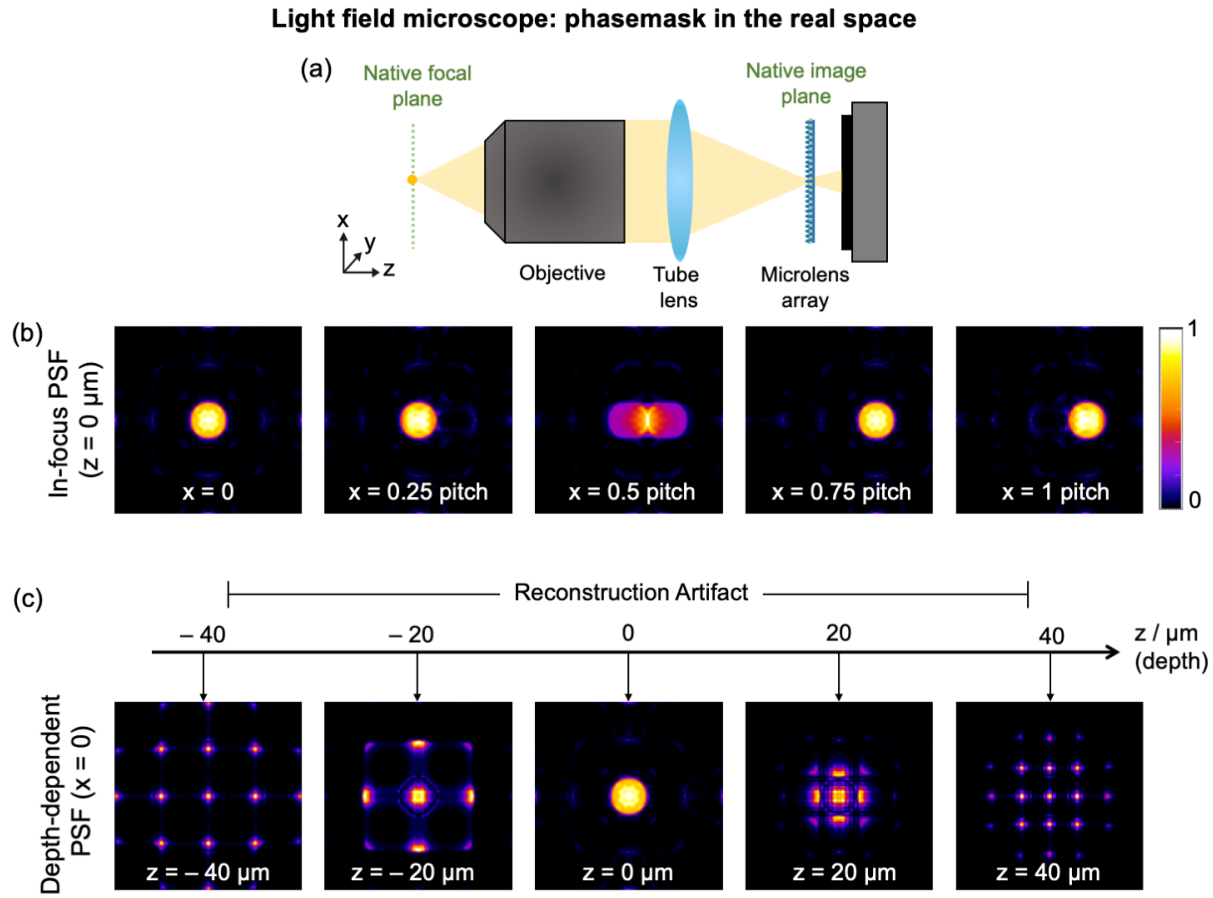


Figure 3.5: Light field microscope configuration and simulated PSFs. (a) Traditional light field microscope contains a microlens array in the native image plane and a sensor at the back focal plane of the microlenses. The system is simulated with a 20X 0.4NA objective and a microlens array with 125 µm pitch and 3125 µm focal length. All PSFs are plotted with a gamma correction of 0.6 to enhance visibility. (b) In-focus PSFs (z = 0 µm) with various lateral shifts. The PSFs are spatially-varying in the lateral dimension with a period of the microlens pitch, and the forward model cannot be modeled as a convolution with a single PSF, resulting in a computationally-intensive deconvolution procedure. (c) Depth-dependent PSFs of on-axis point source (x,y = 0 µm). The size of the PSF around z = 0 µm equals the microlens pitch, and this degradation of resolution causes reconstruction artifacts around the native focal plane.

I present simulated PSFs of an LFM system in Fig. 3.5 to elaborate on the above-mentioned disadvantages. The configuration of a traditional light field microscope is shown

in Fig. 3.5(a), containing a microlens array in the native image plane and a sensor at the back focal plane of the microlenses. I simulate an LFM system with a 20X 0.4NA objective, a microlens array with a 125 µm pitch and a focal length of 3125 µm, and the wavelength is 510 nm. The microlens NA ($\frac{125}{2\times3125} = 0.02$) is set to match the objective NA, and the NA-match relationship is given by the following Eq. 3.1. The NA-match requirement limits the choice of microlens arrays as described in disadvantage ④.

$$\frac{\mathrm{NA}_{\mathrm{obj}}}{M_{\mathrm{obj}}} = \mathrm{NA}_{\mu\mathrm{lens}} \tag{3.1}$$

I follow the wave-optics model in [61] and plot the simulated PSFs in Fig. 3.5(b)(c) with a gamma correction of 0.6 to enhance visibility. The disadvantage ① is demonstrated in Fig. 3.5(b) where the PSFs are spatially-varying in the lateral dimension. Although the translational variance has a period of the microlens pitch, the forward model cannot be modeled as a convolution with a single PSF, ending up with increasing both memory consumption and the computation speed during deconvolution. Moreover, in disadvantage ②, when the point source locates at the native focal plane, the in-focus PSFs are of the size of the microlens pitch, so that the in-focus lateral resolution equals the microlens pitch. This degradation of resolution causes artifacts around the native focal plane during deconvolution and the axial artifact region is given by $|z| \leq \frac{\mathrm{pitch}^2}{2\lambda M_{\mathrm{obj}}^2}$ in [61]. In the case of my simulation, the diffraction-limited resolution of the objective is $\frac{1.22\times\lambda}{2\mathrm{NA}} = 0.78\,\mu\mathrm{m}$, while the in-focus lateral resolution is 125 µm, 160× worse than the objective's diffraction-limited resolution. The reconstruction artifact will appear in the axial region between $z = \pm\frac{125^2}{2\times0.51\times20^2}\mu\mathrm{m} = \pm38.30\,\mu\mathrm{m}$. Existing LFM systems circumvent this issue by placing the sample on one side of the native focal plane and wasting half of the volumetric FOV [6, 61, 86], placing the MLA in an off-focus plane [63–65], or applying wavefront coding [62]. However, even with these methods applied, the performance degrades quickly when the point source moves out of a single microlens' depth of field because all microlenses have the same focal length (disadvantage ③).

To solve some of these problems, an alternative configuration, termed Fourier light field microscopy (FLFM), places the MLA at the relayed Fourier (pupil) plane of the objective, with the sensor one microlens focal length away [66–69]. The configuration of a Fourier light field microscope is presented in Fig. 3.6(a). Placing the phase mask in the Fourier space can extend the range of depths with high resolution. Since a point source in the objective's focal plane will become a plane wave and shine on the whole diffuser, an in-focus point source attains a sharp PSF, getting rid of the artifacts around the native focal plane. In contrast to LFM, where often only one side of the focal plane is employed, we can now use diverging and converging waves coming from both sides of the focal plane. The Fourier-space configuration effectively splits the 2D sensor into a grid of sub-images, with each microlens imaging the sample from a different perspective view. I simulate an FLFM system with a 20X 0.4NA objective, the same one used in the previous LFM simulation. The back pupil has a diameter of 8 mm and is 1:1 relayed onto the microlens array with 1.6 mm pitch and

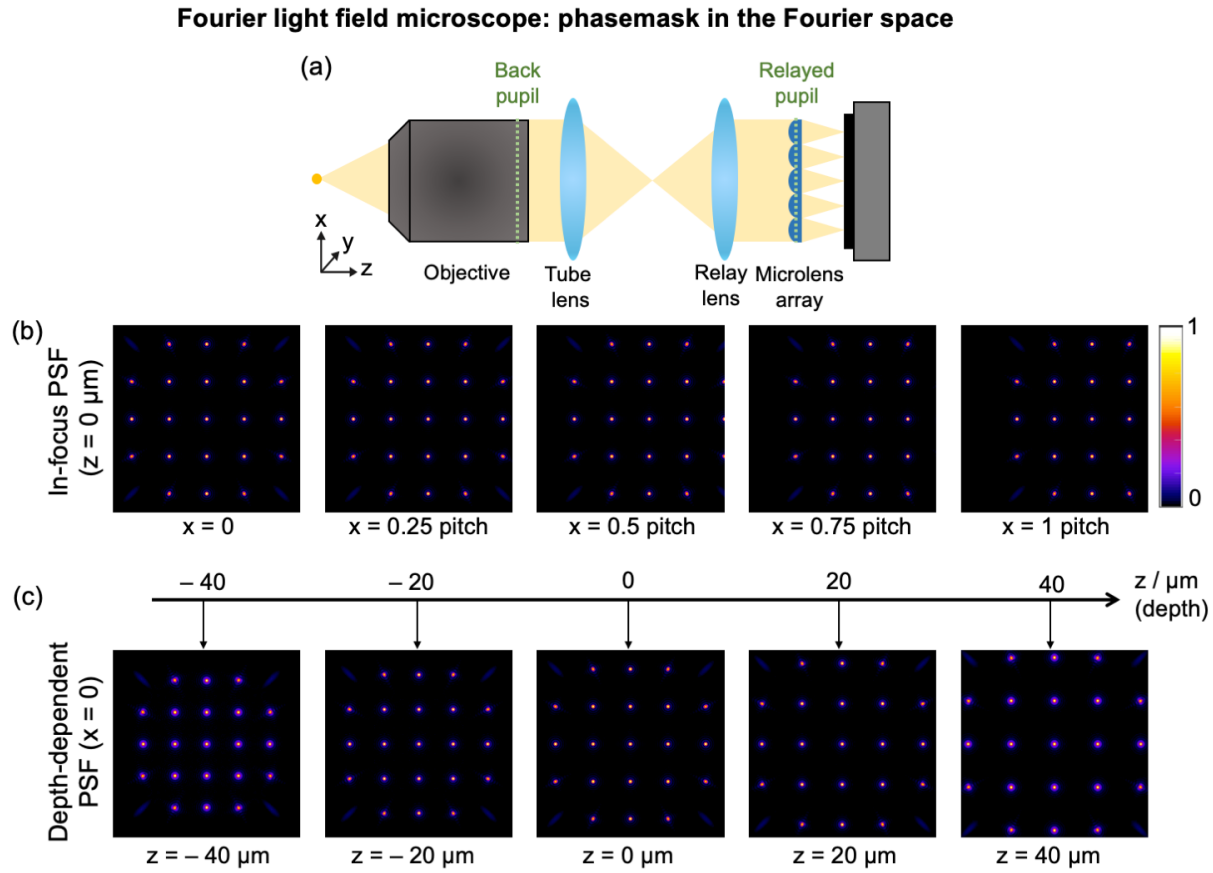


Figure 3.6: Fourier light field microscope configuration and simulated PSFs. (a) Fourier light field microscope contains a microlens array in the relayed Fourier space and a sensor at the back focal plane of the microlenses. The system is simulated with a 20X 0.4NA objective, a 1:1 relay system and a microlens array with 1.6 mm pitch and 65 mm focal length. All PSFs are plotted with a gamma correction of 0.6 to enhance visibility. (b) In-focus PSFs (z = 0 μm) with various lateral shifts. The PSFs are spatial-invariant in the lateral dimension, and the forward model can be modeled as a convolution with a single PSF for each depth, resulting in a computationally-efficient deconvolution procedure. (c) Depth-dependent PSFs of on-axis point sources (x,y = 0 μm). All the focal spots beneath each microlens become out-of-focus simultaneously, resulting in a shallow axial imaging range.

65 mm focal length. I build a wave-optics simulation which will be described in Chap. 3.4 and plot the lateral and axial PSFs in Fig. 3.6(b) and (c). The sharp in-focus PSFs are diffraction-limited by the microlens' NA, containing high-frequency information to recover

high-resolution during reconstruction. The in-focus resolution is $1.22 \times \frac{\lambda}{2\times \mathrm{NA}_{\mathrm{obj}}/5} = 3.89\,\mu\mathrm{m}$, only $5\times$ worse compared to the objective's diffraction-limited resolution because there are 5 microlenses dividing the pupil. Unlike the traditional LFM, FLFM fundamentally resolves the artifacts near the native focal plane. Furthermore, the PSFs are shift-invariant in the lateral dimension for improved computational efficiency.

FLFM solves the disadvantages ①②④ of LFM, but the disadvantage ③ of LFM remains because the axial PSFs become blurry due to defocus and therefore result in a shallow axial imaging range. Besides, placing the periodic MLA in the Fourier plane causes a new disadvantage. When a point in the scene moves laterally by an amount that shifts the PSF by an integer number of pitches, the shifted PSF is nearly the same as the unshifted one; this creates ambiguities that cause the deconvolution to fail. To avoid this problem, usually a field stop is inserted to guarantee that the PSF shifts by less than one period over the FOV [67, 68]. In summary, LFM with the Fourier-space configuration, or FLFM, has two major disadvantages:

① The resolution degrades rapidly with defocus, resulting in a shallow axial imaging range.

② The lateral FOV is limited in order to avoid overlap of the sub-images at the sensor

Different Types of Phase Mask

If we use a phase mask in the Fourier space or far-field, the computational-efficient convolutional forward model applies to various phase mask designs. Here, I compare three popular types of phase mask (smooth Gaussian diffuser, microlens array, random multifocal microlenses) from resolution, noise sensitivity, and cost. I also provide numerical models to evaluate each type of phase mask in simulation. To model a pure phase mask with a given surface height map $T(x, y)$ and refractive index n, I assume there is no absorption (fill factor $= 100\%$, transmittance $= 1$) and the complex transmission function $t(x, y)$ only depends on the phase delay $\phi_T(x, y)$:

$$\begin{aligned} t(x,y) &= 1\, e^{i[\phi_T(x,y)]}, \\ \text{and } \phi_T(x,y) &= \frac{2\pi(n-1)}{\lambda} T(x,y). \end{aligned} \tag{3.2}$$

Given an incident wave E_{in}, the output wave exiting the phase mask E_{out} will be the complex multiplication between the incident wave and the phase mask's transmission function:

$$E_{\mathrm{out}}(x, y, z = 0) = E_{\mathrm{in}}(x, y)t(x, y). \tag{3.3}$$

To simulate the PSF of a phase mask with plane-wave illumination, we have $E_{\mathrm{in}} = 1$, and propagate the output field from the back of the phase mask to its focal plane and acquire $E_{\mathrm{out}}(x, y, z = f)$. The intensity measurement at the sensor is the square modulus of the complex field:

$$I(x, y, z = f) = \left|E_{\mathrm{out}}(x, y, z = f)\right|^2. \tag{3.4}$$

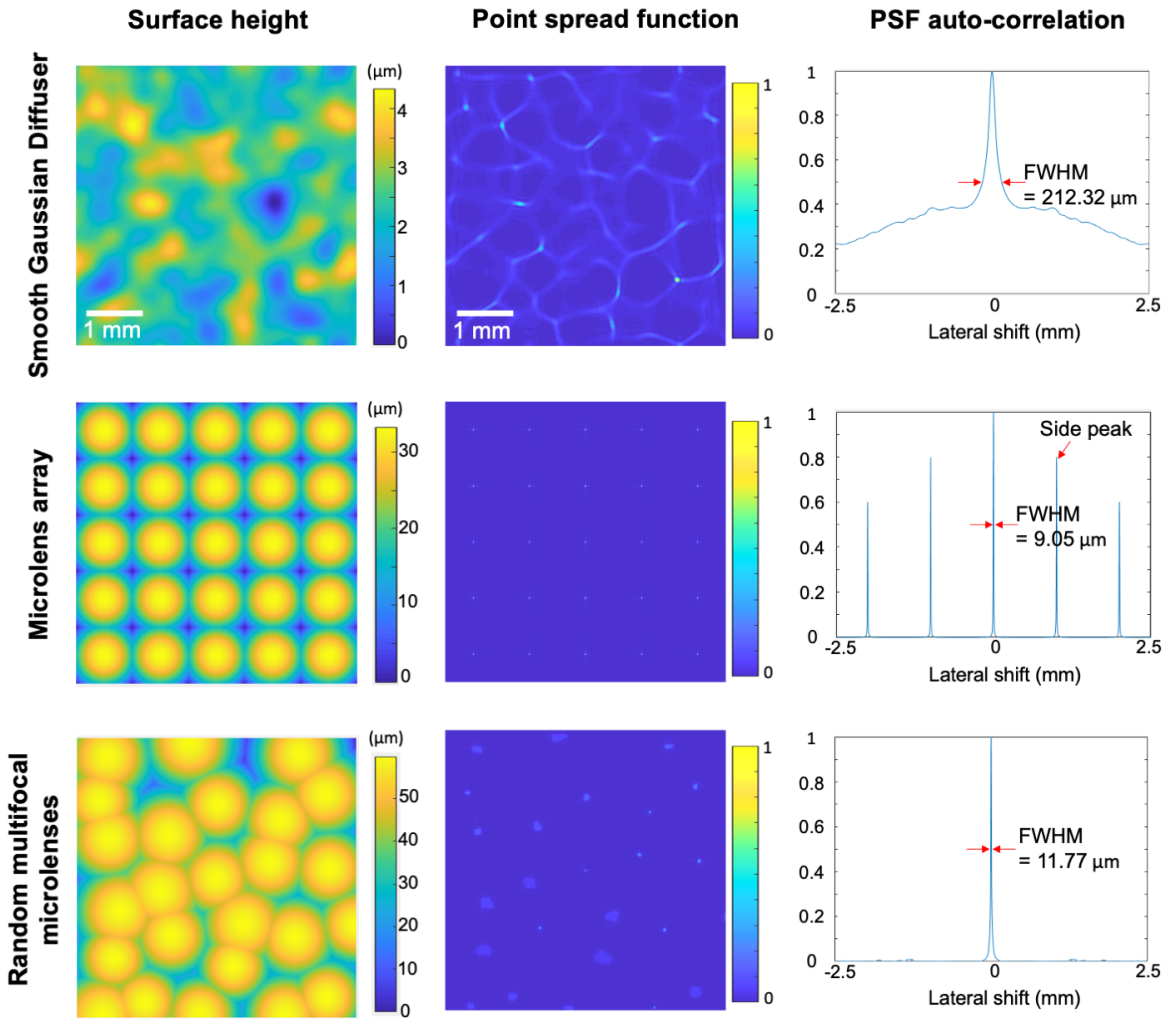


Figure 3.7: Compare the surface height, PSF, and the PSF auto-correlation of different phase masks. All three phase masks have 5 mm side length and 15 mm focal length. The simulated point spread function assumes plane-wave illumination and uses angular spectrum method to propagate the light field to the focal plane of the phase mask. The PSF auto-correlation versus lateral shift is used to quantify the reconstruction performance. The PSF of the smooth Gaussian diffuser has noticeable background light, resulting in noise amplification during reconstruction. The PSF of the microlens array has a periodic structure, resulting in ghost images in reconstruction. The PSF of the random multifocal lenslets has high-contrast focal points and pseudo-random structure, enabling a reconstruction with high SNR and large FOV.

Smooth Gaussian Diffuser

Off-the-shelf diffusers are being used for lensless single-shot 3D imaging in several systems [28, 87]. These pseudorandom planar phase masks are inexpensive, convenient and easily scalable. The surface height $T(x, y)$ can be modeled as a random matrix convolved with a Gaussian kernel:

$$T(x, y) = A\left[rand(x, y) * N(0, \sigma)\right] \tag{3.5}$$

where A is a scaling parameter to adjust the surface height. $rand$ is a uniform distributed random matrix with values between $[0, 1]$. $N(0, \sigma)$ is a 2D Gaussian Kernel with a standard deviation of σ. '$*$' is a 2D convolution operator. Since the smooth Gaussian surface does not contain spherical surfaces, we cannot directly apply the Lens-Maker's Formula to calculate its focal length. Alternatively, we can approximate the surface as a summation of many 2D Gaussian kernels, $g(x)$,and each 2D Gaussian surface is a plano-convex lens. The "radius" of a Gaussian is the radius of curvature of its peak, where the surface has the strongest focusing power. The radius of curvature can be calculated based on the first and second derivatives of the function.

$$\begin{aligned} g(x) &= A\ \exp[-\frac{x^2}{2\sigma^2}] \\ g'(x) &= A\frac{-x}{\sigma^2}\ \exp[-\frac{x^2}{2\sigma^2}] \\ g''(x) &= A\frac{-1}{\sigma^2}\ \exp[-\frac{x^2}{2\sigma^2}] + A\frac{x^2}{\sigma^4}\ \exp[-\frac{x^2}{2\sigma^2}] \\ R(x) &= \left|\frac{[1+g'(x)^2]^{\frac{3}{2}}}{g''(x)}\right| \\ \Longrightarrow R(x=0) &= \frac{\sigma^2}{A} \end{aligned} \tag{3.6}$$

Once we have the radius of curvature of the Gaussian surface ($R_1 = \sigma^2/A$), with the knowledge that the other side is flat ($R_2 = \infty$), we can use the Lens-Maker's Formula to get the approximated focal length:

$$f_{\text{Gaussian}} = \frac{\sigma^2}{A(n-1)} \tag{3.7}$$

where n is the refractive index of the material.

Using the above-mentioned equations, I simulate a smooth Gaussian diffuser surface with 5 mm side length, 15 mm focal length in Fig. 3.7. The surface contains random convex and concave bumps, directing incoming light into a caustic pattern, or PSF of the diffuser. The simulated PSF assumes plane-wave illumination and uses angular spectrum method (See details in Chap. 3.4) to propagate the light field to the focal plane of the phase mask. The caustic pattern contains random sharp lines spread in all directions, which enable good resolution at all depths and a highly structured PSF for deconvolution. To quantify the

sharpness and the matrix condition of the caustic pattern, I calculate the auto-correlation versus lateral shift and then examine the full-width at half-maximum (FWHM) of the peak and the sidelobes. A smaller FWHM value stands for a sharper PSF, which contains higher frequency information and will provide better spatial resolution for an imaging system. The high sidelobes represent an ill-conditioned matrix for sparsity-constrained reconstruction. Because when the shifted PSF is similar to the un-shifted version, it is challenging for an inverse algorithm to determine the true location of an object, causing higher noise amplification during deconvolution. From the PSF auto-correlation comparison in Fig. 3.7, the smooth Gaussian diffuser has an FWHM of 212.32 µm and a sidelobe of more than 20%. Compared to lens-based phase masks, the smooth Gaussian diffuser ends up with a worse spatial resolution and a fair matrix condition.

Microlens Array

Microlens Array (MLA) is a phase mask consisting of a regular array of microlenses, being widely used in light field imaging. Each microlens is a convex-plano spherical lens that concentrates all incoming light into a diffraction-limited spot beneath. The PSF of this type of phase mask contains an array of high-contrast and sharp focal spots.

To model the complex transmission function of a microlens array with focal length $f_{\mu lens}$ and period d, we start from a single microlens, and the rest of them are all identical. This microlens, $t_{\mu lens}(x, y)$, can be modeled as an amplitude mask $t_A(x, y)$ representing the aperture and a phase mask $\phi_T(x, y)$ representing the refraction of light.

$$t_{\mu lens}(x, y) = t_A(x, y)e^{i[\phi_T(x,y)]}. \tag{3.8}$$

Using parabolic approximation when the radius of a lens is much larger than its aperture size, namely $R \gg d$, the phase shift of the lens is

$$\phi_T(x, y) = -\frac{k}{2f_{\mu lens}}(x^2 + y^2). \tag{3.9}$$

The amplitude term $t_A(x, y)$ is related to the microlenses' shape and filling factor. If the filling factor = 100%, meaning each microlens' aperture is square with a side length equals the period d,

$$t_A(x, y) = \text{rect}(\frac{x}{d})\text{rect}(\frac{y}{d}). \tag{3.10}$$

If the microlenses' apertures are circular with a diameter equal to the period,

$$t_A(x, y) = \text{circ}(\frac{\sqrt{x^2 + y^2}}{d/2}). \tag{3.11}$$

Once we have the profile of a single microlens, the whole microlens array is a summation of shifted microlenses and can be described as a convolution between $t_{\mu lens}(x, y)$ and a 2D comb function with period d.

$$t(x, y) = t_{\mu lens}(x, y) * [\text{comb}(\frac{x}{d})\text{comb}(\frac{y}{d})] \tag{3.12}$$

Plotted in Fig. 3.7, I simulate a microlens array with 25 microlenses, 5 mm side length and 15 mm focal length. The PSF has very low background light and very sharp focal points. The PSF auto-correlation diagram has a narrow central peak with a FWHM of 9.05 µm, and the sidelobes close to the central peak are nearly 0. However, when the regularly-spaced microlens array is translated by exactly one period, the shifted PSF largely overlaps with the on-axis PSF in FLFM, resulting in an increased auto-correlation and higher noise amplification. In my simulated case with 25 focal spots, laterally shifting one period will end up with 20 overlapping focal spots, and thus 2 side peaks with an auto-correlation value of 0.8. In order to avoid the side peaks, the field-of-view of FLFM is limited in order to avoid overlap of the sub-images at the sensor; otherwise, the reconstruction shows periodic ghosting effect.

Despite the downsides mentioned above, having a regular structured phase mask makes it easier for fabrication and calibration. There are plenty of off-the-shelf microlens arrays with varying parameters from Thorlabs, RPC photonics, Holographix, etc, and it is a standard procedure to customize one with specific parameters. Knowing the focal length and period of a high-quality microlens array, it is also possible to use simulated PSFs for 3D reconstruction without calibration, which eases out the experiment procedure [6, 61].

Random Multifocal Microlenses

Random multifocal microlenses (RMM) acquire the benefits of MLA and resolve several shortcomings. Like MLA, RMM encodes depth information with microlenses in front of the sensor, providing an in-focus PSF with a high SNR. However, unlike MLA, RMM is randomly-spaced with varying focal length, more similar to a diffuser. Randomizing the location of the microlenses breaks the ambiguity caused by periodicity, enabling full-FOV imaging with a sparsity-constrained inverse solver. Assigning different focal lengths to the microlenses, as in the RMM, extends the imaging depth range, within which there is always a subset of microlenses in focus. This trades SNR near the native focal plane for an increased depth range due to spreading high-frequency information across the whole volume. The RMM PSFs form nearly orthogonal columns of the design matrix, enabling a compressed sensing 3D reconstruction with more voxels than the pixels in the 2D measurement.

To simulate an RMM surface with side length L containing N microlenses, I begin with finding out the location and focal length of each microlens. The centers of the randomly-spaced microlenses, $(x_i, y_i)_{i=1,\dots,N}$, are independent, identically distributed and is generated one-by-one from a random distribution (I use continuous uniform distribution here), under the constraint that the distance between adjacent centers is at least 70% of the averaged

microlens pitch $d = M/\sqrt{N}$. The value 70% is a tuning parameter to prevent randomly generated microlenses from overcrowding.

$$\begin{aligned} x_i, y_i &\sim \mathcal{U}_{[0,L]} \text{ for } i \text{ in } 1, ..., N \\ \text{s.t.} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} &\geq 0.7 * d \text{ for any } i, j \text{ in } 1, ..., N. \end{aligned} \tag{3.13}$$

The minimum and maximum focal lengths affect the axial range of the imaging volume, and the rest focal lengths are evenly spaced within that range. Based on the lens maker's formula of thin lenses, and knowing that one surface is flat (radius = ∞), I acquire the following relationship between the ith microlens' surface radius R_i and focal length f_i:

$$R_i = (n-1)\frac{1}{f_i} \text{ for } i \text{ in } 1, ..., N \tag{3.14}$$

I draw the spherical sag of the microlenses in Eq. 3.15 and Fig. 3.8. The final diffuser surface is a point-wise maximum over the individual spherical lens profiles with a maximum thickness of t. The aperture shape of the microlenses is irregular, determined by the centers and radii of the adjacent microlenses, similar to a weighted Voronoi diagram.

$$\mathbf{s}(x_i, y_i, R_i) = t + R_i \left(\sqrt{\max\left(1 - \left(\frac{x - x_i}{R_i}\right)^2 - \left(\frac{y - y_i}{R_i}\right)^2, 0\right)} - 1 \right) \tag{3.15}$$

$$\mathbf{T}(x, y) = \max_i \mathbf{s}(x_i, y_i, R_i) \tag{3.16}$$

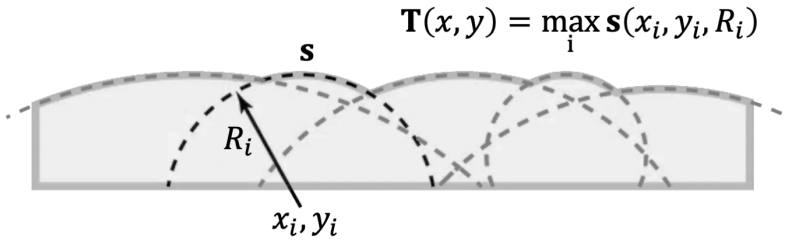


Figure 3.8: Random multifocal microlenses parameterization. I draw the spherical sag $\mathbf{s}$ of the microlenses with varying locations (x_i, y_i) and focal lengths R_i. The final diffuser surface $\mathbf{T}$ is a point-wise maximum over the individual spherical lens profiles.

Similar to the smooth Gaussian diffuser and microlens array, I simulate a random multifocal microlenses phase mask containing 25 microlenses, with a side length of 5 mm and an average focal length of 15 mm (ranging from 12 mm to 18 mm). Like the MLA, the

PSF of RMM has high contrast and a central peak with a FWHM of 11.77 µm. Due to some microlenses being defocused, the FWHM of RMM is slightly wider than the MLA, but still much narrower than the smooth Gaussian diffuser. Unlike the periodic MLA, the auto-correlation no longer has high side peaks, with all the side lobe areas being close to zero. This effectively avoids ambiguity during reconstruction, and the FOV of RMM is only limited by shift-invariance.

3.4 Wave-optical Forward Model

I have mentioned the forward model in Chap. 2 Eq. 2.1, in which the image formation of Fourier DiffuserScope is compressed into a linear measurement matrix **H**. To determine **H** that contains the PSFs of the optical system, I introduce the wave-optical forward model based on Fourier optics [88] in this section. This simulation framework, not limited to Fourier DiffuserScope, works for a general configuration of phase-mask-based single-shot 3D microscope including Fourier LFM.

In summary, the simulation is ordered as follows. From a point source location in the object space, I calculate the spherical wavefront at the the back focal plane, then multiply by the apodization function of the objective to get the wavefront distribution at the pupil plane. The wavefront at the pupil is then scaled by the relay system, multiplied by the transmission function of the phase mask, and propagated to the sensor using the angular spectrum method. I first define the simulation parameters in Table 3.2, then dive into the implementation of each function.

Sampling Requirements

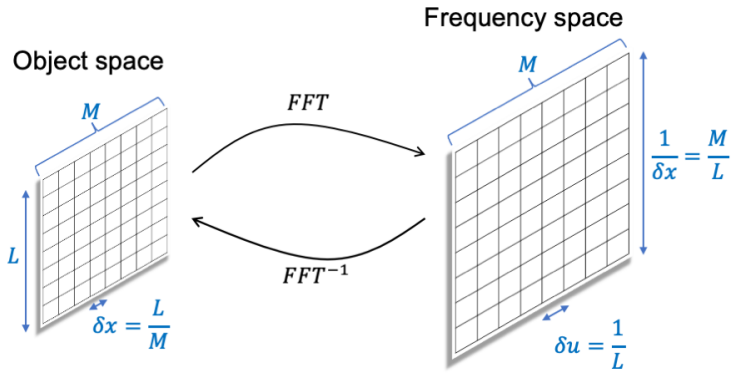


Figure 3.9: Sampling rate in the object space and frequency space.

M	size of the square simulation grid in one axis
L	width of the simulation grid (mm)
(x, y)	lateral coordinate in the object space
(u, v)	lateral coordinate in the frequency space
(u', v')	lateral coordinate in the relayed frequency space
(x_0, y_0, z_0)	point source location in the object space
λ	wavelength
$k = \frac{2\pi}{\lambda}$	wavenumber
f_{obj}	focal length of the objective
n	refractive index of the immersive media
NA_{obj}	Numerical aperture of the objective
f_{TL}	focal length of the tube lens
f_{RL}	focal length of the relay lens
$\text{NA}_{\mu\text{lens}}$	Average numerical aperture of the microlenses
$f_{\mu\text{lens}}$	Average focal length of the microlenses
s	sensor pixel size

Table 3.2: Parameter definitions for numerical simulation.

The object space and the frequency space are connected by the following Fourier transform,

$$u_{\text{Fourier}}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_{\text{Real}}(x, y) \exp[-j \frac{k}{f_{\text{obj}}}(ux + vy)] dx dy \tag{3.17}$$

where (x, y) are the Cartesian coordinate in the object space, or real space, and (u, v) are the Cartesian coordinate in the pupil space, also called Fourier space, or frequency space. In Fourier transforms, the resolution at one space is inversely proportional to the field width at another space. In other words, the field width L in the real space determines the pixel size δu in the frequency space, and vice versa. Since the simulation grid size M stays constant during Fourier transform, a larger field width at the real space provides a higher sampling resolution at the frequency space.

To Nyquist sample the minimum resolvable spot in the real space, simulation pixel size must satisfy

$$\delta x = \frac{L}{M} \leq \frac{1}{2}\frac{\lambda}{2\mathrm{NA}_{\mathrm{obj}}}. \tag{3.18}$$

From the inverse relationship, the width of the frequency space is $\frac{1}{\delta x} = \frac{M}{L} \geq 2 \times \frac{2\mathrm{NA}_{\mathrm{obj}}}{\lambda}$, within which the back pupil occupies a center circle with a diameter of $\frac{2\mathrm{NA}_{\mathrm{obj}}}{\lambda}$. Another sampling requirement that resides in the Fourier space is to Nyquist sample the minimum resolvable spot on the sensor, because the sensor is behind a Fourier-space phase mask. The width of the simulated field in the sensor plane is demagnified from the pupil plane to be $\frac{M}{L}\frac{f_{\mathrm{RL}}}{f_{\mathrm{TL}}}$. So we have

$$\delta x' = \frac{M}{L}\frac{f_{\mathrm{RL}}}{f_{\mathrm{TL}}} \leq \frac{1}{2}\frac{\lambda}{2\mathrm{NA}_{\mu\mathrm{lens}}}. \tag{3.19}$$

Combining Eq. 3.18 and Eq. 3.19 yields the final constraints on the sampling grid,

$$2\frac{f_{\mathrm{RL}}}{f_{\mathrm{TL}}}\frac{2\mathrm{NA}_{\mu\mathrm{lens}}}{\lambda} \leq \frac{L}{M} \leq \frac{1}{2}\frac{\lambda}{2\mathrm{NA}_{\mathrm{obj}}}. \tag{3.20}$$

It is worth noting that Eq. 3.20 gives a good starting point to determine the sampling rate. However, it's always recommended to check the beam profile at every surface in the optical path after running the whole simulation. Due to some irregular surface shapes (e.g., the RMM) or fast diverging/converging beams (e.g., defocused point sources), the phase changes quickly in certain area and a higher sampling rate might be necessary to avoid aliasing. Users need to plot the phase profile of each surface to make sure the optical field is adequately sampled.

Point Sources

The point source to an optical system is an impulse function, namely Dirac delta function, to a linear system. An ideal point source, in theory, should be infinitely small. But in numerical simulation, the point source must have a finite spatial extent and the smallest size available is the grid size in the object space. If a point source must be acquired, previous work used a narrow Gaussian function with a quadratic phase, sinc-Gaussian function, and so on [89].

Fortunately, in this case, we can utilize the Fourier transform feature of a lens to avoid the sampling issue. Since a point source has infinite spatial bandwidth, its Fourier spectrum is constant across all spatial frequencies. Knowing the point source location, we can derive the pupil function using Fourier optics. In the following paragraphs, we decouple the pupil function for an arbitrary point source location (x_0, y_0, z_0) into three parts: defocus corresponds to the axial location z_0, lateral shift corresponds to the lateral displacement (x_0, y_0), and apodization corresponds to the nature of the objective lens.

Pupil Functions

① Defocus and Spherical Wavefront

In this discussion, I only consider the effect of defocus z_0, and assume all point sources are on-axis, namely $x_0 = y_0 = 0$. A delta function at origin is Fourier transformed into a constant. Similarly in optics, a point source at the front focal point ($z_0 = 0$) is transformed into a plane wave in the back pupil, $u_{\text{defocus}} = c$.

When $z_0 \neq 0$, the light emitted from the point source forms a spherical wave at the front focal plane, which is then Fourier transformed. One way to simulate the defocused pupil is to apply a Fourier transform to the object-space spherical wave numerically, but computing an FFT is slow. Here, I use thin lens equation to derive an analytical defocused pupil function. From the Newtonian form of the thin lens equation, a point source at z_0 away from the front focal plane is imaged to a point at $z' = -f_{\text{obj}}^2/z_0$ away from the back focal plane. With the knowledge of the image location, we can infer the spherical wavefront at the back pupil to be

$$u_{\text{defocus}} = \frac{E_0}{r} \exp[-\text{sign}(z')jkr], \tag{3.21}$$

where

$$\begin{aligned} r &= \sqrt{u^2 + v^2 + z'^2}, \\ z' &= -\frac{f_{\text{obj}}^2}{z_0}. \end{aligned}$$

u, v are the frequency coordinates in the Fourier space.

② Lateral Displacement and Wavefront Tilt

In Fourier transform, a lateral displacement in the real space converts to a phase shift in the Fourier space; and correspondingly, an off-axis point source in the object space results in a tilted wavefront in the pupil space. I build a coordinate system shown in Fig. 3.10 with the origin being the front focal point of the objective lens and the positive z pointing towards the light propagation direction. In the spherical coordinate systems, a tilt angle is decomposed into a polar angle θ and an azimuthal angle ϕ. The relationship between the Cartesian coordinate (x, y, z) and the tilt angle (θ, ϕ) is expressed as

$$\begin{aligned} \theta &= \arctan\left(\frac{\sqrt{x^2 + y^2}}{f_{\text{obj}} - z}\right), \\ \phi &= \text{arctan2}(y, x). \end{aligned} \tag{3.22}$$

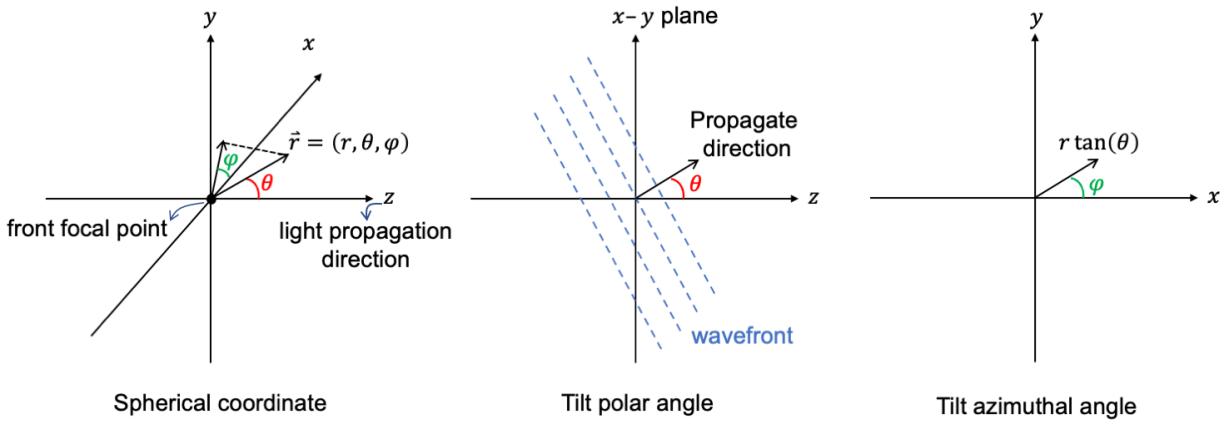


Figure 3.10: Wavefront tilt angle diagram.

For an off-axis point source at (x_0, y_0, z_0), I apply a phase ramp $\Phi(u, v)$ to account for the tilt due to $(x_0, y_0) \neq (0, 0)$ [90], such that the output field

$$u_{\text{tilt}} = u_{\text{in}} \exp\left[j\Phi(u, v)\right], \tag{3.23}$$

where

$$\Phi(u, v) = k(x\cos\phi + y\sin\phi)\tan\theta. \tag{3.24}$$

③ Apodization Function

The previous discussions on defocus and tilt ignore the physical size of the lens and assume the back pupil extends infinitely. However, the lens back pupil has a finite size in the frequency space limited by its NA and wavelength. The diameter of the back pupil in the frequency space is $\frac{2\text{NA}_{\text{obj}}}{\lambda}$. The area outside of the back pupil is physically cropped and the pixel value should be set to zero.

Within the back pupil, we consider 'apodization', a tapering function that gradually brings a sampled signal down to zero at the edges of the pupil. In low-NA objectives, paraxial approximation holds well and apodization can be ignored. In high-NA objective with NA$>$ 0.7, the apodization effect becomes noticeable and must be included in the simulation. For a lens designed to meet the Abbe sine condition, the apodization function is $P(\theta) = P(r)\sqrt{\cos(\theta)}$, where $r = \sqrt{u^2 + v^2}$ and $\theta = \arcsin(r\frac{\lambda}{n})$ [91]. Namely, the field after

apodization becomes

$$u_{\text{apod}}(r) = \begin{cases} u_{\text{in}}(r)\sqrt{\cos[\arcsin(r\frac{\lambda}{n})]}, & \text{if } r \leq \frac{2\text{NA}_{\text{obj}}}{\lambda} \\ 0, & \text{otherwise} \end{cases} \tag{3.25}$$

where $r = \sqrt{u^2 + v^2}$.

$\Rightarrow$ In summary, with the consideration of defocus, lateral shift, and apodization, the final format of the optical field at pupil is

$$u_{\text{pupil}} = u_{\text{apod}} \cdot u_{\text{tilt}} \cdot u_{\text{defocus}}. \tag{3.26}$$

Magnification

The relay lens and the tube lens form a $4f$ system that relays the back pupil onto the phase mask with a demagnification rate of $\frac{R_{\text{RL}}}{R_{\text{TL}}}$. Since in the experimental system the demagnification rate is less than 5, we assume the $4f$ system is ideal and ignore any distortions. The relayed optical field u_{relay} has the same structure as u_{pupil}, but at a small scale.

$$u_{\text{relay}}(u', v'; L') = u_{\text{pupil}}(u, v; L), \tag{3.27}$$
$$\text{where } u' = u \times \frac{f_{\text{RL}}}{f_{\text{TL}}}, v' = v \times \frac{f_{\text{RL}}}{f_{\text{TL}}}, L' = L \times \frac{f_{\text{RL}}}{f_{\text{TL}}}.$$

Spherical Microlenses

The basic building block of an MLA and an RMM is convex-plano microlenses with spherical surfaces. Each spherical lens contains optically dense material with varying thickness, adding phase delays to the incident wavefront. In Eq. 3.9, I briefly mentioned the phase transformation of a lens. Here, I elaborate on the derivation behind the spherical lens model. The lens geometry is plotted in Fig. 3.11.

Let the thickness at (x, y) be $\Delta(x, y)$, the phase delay becomes

$$\Phi(x, y) = kn\Delta(x, y) + kn_0[\Delta_0 - \Delta(x, y)], \tag{3.28}$$

where n is the refractive index of the lens material, n_0 is the refractive index of the surrounding environment which is usually air, and Δ_0 denotes the lens on-axis thickness $\Delta(0, 0)$. So the transmission function of a lens is a phase transformation

$$t(x, y) = \exp[j\Phi(x, y)] = \exp[jk\Delta_0 + jk(n-1)\Delta(x, y)]. \tag{3.29}$$

Since the lens shape is known to be convex-plano, assuming the radius of the spherical surface is R, the thickness function is

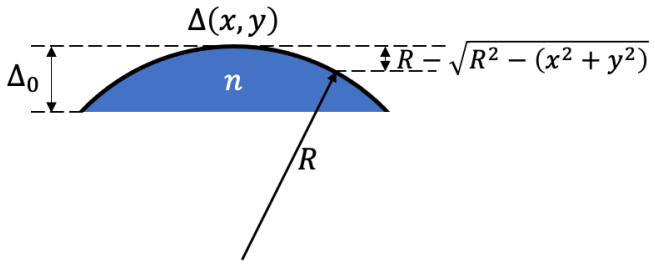


Figure 3.11: Geometry for a spherical lens to calculate the phase mask thickness function.

$$\Delta(x,y) = \Delta_0 - \left(R - \sqrt{R^2 - (x^2+y^2)}\right) = \Delta_0 - \left(R - R\sqrt{1 - \frac{x^2+y^2}{R^2}}\right) \tag{3.30}$$
$$\approx \Delta_0 - \left(R - R(1 - \frac{x^2+y^2}{2R^2})\right) = \Delta_0 - \frac{x^2+y^2}{2R}.$$

The approximation is based on paraxial rays only interacting with the very center area of the lens so that $\frac{x^2+y^2}{R^2} \ll 1$.

Substitute of Eq. 3.30 into Eq. 3.29, we have the expression of the lens phase transformation using the surface radius,

$$t_{\text{phase}}(x,y) = \exp\left[jk\Delta_0 + jk(n-1)(\Delta_0 - \frac{x^2+y^2}{2R})\right] \tag{3.31}$$
$$= \exp[jkn\Delta_0] \exp\left[-jk(n-1)\frac{x^2+y^2}{2R}\right].$$

The constant phase term $\exp[jkn\Delta_0]$ is often dropped since only the phase delay term matters. And the final phase transformation function becomes

$$t_{\text{phase}}(x,y) = \exp\left[-jk(n-1)\frac{x^2+y^2}{2R}\right]. \tag{3.32}$$

For thin convex-plano lenses, the radius and the refractive index can be converted into the focal length using $\frac{1}{f} = \frac{n-1}{R}$, yielding an equivalent expression of the phase transformation

function,

$$t_{\text{phase}}(x,y) = \exp\left[-jk\frac{x^2+y^2}{2f}\right]. \tag{3.33}$$

The last step is to account for the physical size of the lens and bring in the amplitude term. With a lens aperture of D, the final lens transmittance function is

$$t_{\mu\text{lens}}(x,y) = \text{circ}\left(\frac{\sqrt{x^2+y^2}}{D/2}\right)\exp\left[-jk\frac{x^2+y^2}{2f}\right]. \tag{3.34}$$

Until now, this derivation converges to Eq. 3.9. The RMM and MLA phase masks are simulated by composing multiple microlens phase transformation functions. I use $t_{\text{phase mask}}$ to represent a general transfer function of any kind of phase mask. The optical field right after the phase mask is

$$u_{\text{phasemask}}(u',v') = u_{\text{relay}}(u',v')\cdot t_{\text{phasemask}} \tag{3.35}$$

Physical Optics Propagation

The next step of the simulation pipeline is to propagate the optical field from the back of the phase mask to the sensor plane. When light propagate in free space, the wavefront coherently interferes with itself. To model the coherent propagation, two widely accepted algorithms are 'Fresnel diffraction propagation' and 'angular spectrum propagation', both with advantages and disadvantages. The rule of thumb to choose a propagation algorithm is to calculate the Fresnel number. The Fresnel number F_n depends on the diameter of the beam A and the distance Z between the source and the observation plane, $F_n = \frac{A^2}{\lambda Z}$.

When the propagation distance is large, that is $F_n < 1$, the observation plane resides in the far field and the Fresnel propagator should be chosen. In this case, the phase of the optical field is measured relative to a reference sphere with a radius equal to Z. Although the total number of simulation size M remains constant, the pixel size δu will change as the beam propagates, adjusted to the field size change during propagation. On the other hand, when the observation plane resides in the near field, namely $F_n > 1$, the angular spectrum propagator works well. Different from the Fresnel propagator, the phase of the optical field is measured relative to a plane. Both the simulation size M and the pixel size δu stay the same between the source and the observation planes [92].

In the experimental system, I use an Olympus 20×, 1.0NA objective, a $f_{TL} = 180\,\text{mm}$ tube lens, a $f_{RL} = 48\,\text{mm}$ relay lens, and a RMM with an average focal length of $f_{\mu\text{lens}} = 15.6\,\text{mm}$. The physical pupil size $D = 2\times f_{\text{obj}}\times \text{NA}_{\text{obj}} = 2\times 9\times 1.0 = 18\,\text{mm}$. After the $4f$ system, the beam size becomes $A = D\times\frac{f_{\text{RL}}}{f_{\text{TL}}} = 18\times\frac{48}{180} = 4.8\,\text{mm}$. The propagation distance equals the average focal length of the microlenses, $Z = f_{\mu\text{lens}} = 15.6\,\text{mm}$. So, the Fresnel number $F_n = \frac{4.8\,\text{mm}^2}{520\,\text{nm}\times 15.6\,\text{mm}} = 2.84 > 1$. Therefore, the angular spectrum propagator

is a better choice in this simulation. I apply the angular spectrum algorithm [88] to acquire the optical field on the sensor plane,

$$
\begin{aligned}
u_{\text{sensor}} &= u_{\text{phasemask}}(u', v') * \left[\frac{\exp(jkz)}{j\lambda Z} \exp\left(\frac{jk}{2Z}(u'^2 + v'^2) \right) \right], r = \sqrt{Z^2 + u'^2 + v'^2} \qquad (3.36) \\
&= \mathcal{F}^{-1}\left[\mathcal{F}\Big[u_{\text{phasemask}}(u', v') \Big] \mathcal{F}\Big[\frac{\exp(jkz)}{j\lambda Z} \exp\left(\frac{jk}{2Z}(u'^2 + v'^2) \right) \Big] \right] \\
&= \mathcal{F}^{-1}\left[\mathcal{F}\Big[u_{\text{phasemask}}(u', v') \Big] \exp(jkz) \exp\big(-j\pi\lambda Z(f_{u'}^2 + f_{v'}^2) \big) \right]
\end{aligned}
$$

where $*$ denotes 2D convolution, $\mathcal{F}$ denotes Fourier transform, $\mathcal{F}^{-1}$ denotes inverse Fourier transform, $f_{u'}$ and $f_{v'}$ are the frequency coordinates corresponds to u' and v'.

Sensor Measurement

Since camera sensors only capture intensity information and ignore phase, the final step is to convert the optical field at the sensor plane into a realistic sensor measurement. The intensity of the simulated field is proportional to the absolute square of the electrical field,

$$
I_{\text{sensor}} = u_{\text{sensor}} \cdot \overline{u_{\text{sensor}}}. \qquad (3.37)
$$

If the sensor pixel size s is larger than the simulation grid size at the sensor plane $\delta u'$, a downsampling procedure with a rate of $\frac{s}{\delta u'}$ is necessary. To better represent the physical process, the downsampling algorithm should bin all the values within a sensor pixel block instead of skipping. To this end, the final output of the simulation pipeline is

$$
I_{\text{output}} = \text{downsample}\big(I_{\text{sensor}}, \frac{s}{\delta u'}\big) \qquad (3.38)
$$

3.5 Simulation Results

I use simulations to numerically validate the design theory derived in the previous section and to demonstrate the advantage of using RMM over MLA and RUM. I set the target performance to be $\sim 2\,\mu\text{m}$ resolution across a $\sim 200\,\mu\text{m}$ depth range using a $20\times$, 1.0NA objective lens ($f_{\text{obj}} = 9\,\text{mm}$, $\text{NA}_{\text{obj}} = 1.0$, $\text{FOV}_{\text{obj}} = 1.1\,\text{mm}$, $D = 18\,\text{mm}$). The design wavelength is $\lambda = 510$ nm for common green fluorescent calcium indicators. The tube lens and the relay lens form a $1:1$ relay system to conjugate the back pupil plane onto the phase mask, so the diffuser side length equals the pupil diameter ($N \cdot p = 18\,\text{mm}$). Calculated from Eq. 2.4 and Eq. 2.7, the diffuser has at most $N = 5$ microlenses in one transverse direction with an average pitch size $p = 3.6\,\text{mm}$, resulting in an effective NA

of 0.2 and predicted resolution $R_{\text{lateral}} = 1.56\,\mu\text{m}$ and $R_{\text{axial}} = 1.94\,\mu\text{m}$. The average focal length of the RMM ($f_{\text{ave}} = 58.5$ mm), matched to the focal lengths of the MLA and RUM, is chosen to achieve a total magnification of $M = 6.5\times$. For the RMM, with the goal of $\pm z = \pm 100\,\mu\text{m}$, the microlens focusing at the nearest and farthest depth planes have $f_{\text{min}} = 54.6\,\text{mm}$ and $f_{\text{max}} = 63.1\,\text{mm}$, respectively. The focal lengths of the remaining 23 microlenses are dioptrically distributed between f_{min} and f_{max}.

The surface height of the three phase mask designs are shown in Fig. 2.1. The centers of the randomly-spaced microlenses are generated from a uniform distribution, under the constraint that the distance between adjacent centers is at least 70% of the microlens pitch. A spherical surface is placed at the center of each microlens location, taking into account the focal lengths and refractive index ($n_r = 1.56$ for photopolymer). Then, I take the point-wise maximum surface height to form the final diffuser with 100% fill factor. The sensor is located at the distance of the average focal length behind the diffuser, with the 2 µm pixel size ensuring Nyquist sampling of the diffraction-limited pattern.

The resulting PSFs are shown in Fig. 2.1, with the in-focus PSF being the intensity pattern at the sensor for a point source at the native focal plane of the objective, and the defocus PSFs being the intensity for a point source off-focus by 100 µm towards the objective. For both uni-focal designs (MLA, RUM), all the lenslets are in-focus or out-of-focus simultaneously, while for RMM each microlens comes into focus at a different plane.

Resolution

To characterize the lateral and axial resolution, which vary with depth, I reconstruct volumes from acquisitions with two point sources at varying separation distances. For this very sparse scene, the reconstruction converges in only 5 iterations of Richardson-Lucy without regularization [93, 94]. I stop after 8 iterations and consider the two points resolved when there is at least a 20% intensity drop between them, as in the Rayleigh criterion. For lateral resolution, the two points are placed on the same depth plane with separation only in the x-y direction; for axial resolution, the two points are both on the optical axis and symmetrically set apart from the designated depth plane. The results in Fig. 3.12 compare the reconstruction resolution for the three diffuser/MLA designs, with comparison to the theory presented in Chap. 2.3.

At the native focal plane ($z = 0\,\mu\text{m}$) the MLA has a lateral resolution of 0.6 µm and the RUM has a lateral resolution of 1.1 µm (Fig. 3.12(a)), somewhat better than the predicted $R_{\text{lateral}} = 1.56\,\mu\text{m}$ owing to deconvolution. However, the resolution of both uni-focal designs (MLA, RUM) degrades rapidly with depth; based on Eq. 2.6, the slope of the resolution with depth is 0.13 laterally and 0.1625 axially. The lateral resolution of Fourier DiffuserScope (RMM) remains relatively steady over a large depth range ($z = -80\,\mu\text{m}$ to $z = 90\,\mu\text{m}$), varying between $1.4 \sim 2.6\,\mu\text{m}$.

The axial resolution (Fig. 3.12(b)) follows similar trends. The highest axial resolution for both MLA and RUM is 1.75 µm at the native focal plane, which is somewhat better than the theoretical prediction of $R_{\text{axial}} = 1.94\,\mu\text{m}$ (Chap. 2.3). The axial resolution of RMM

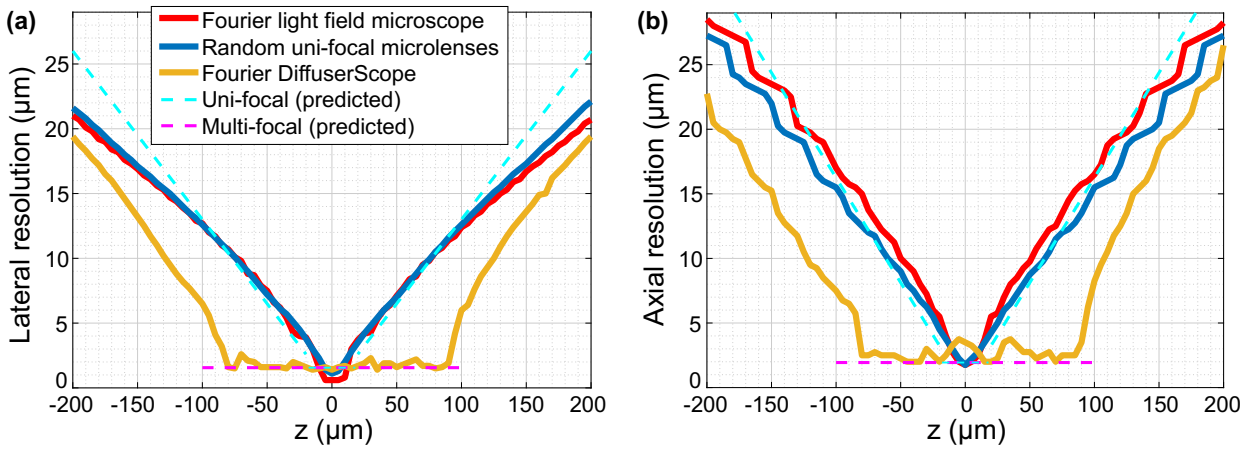


Figure 3.12: Comparison of simulated and theoretical two-point resolution at different depth planes for the three cases of Fourier microlens designs: Fourier light field microscope's MLA, RUM and Fourier DiffuserScope's RMM. (a) Lateral resolution and (b) axial resolution at different depth planes. The MLA used in Fourier light field microscope (red solid line) and the random uni-focal microlenses (blue solid line) have the best resolution at the native focal plane ($z = 0$) but the performance degrades rapidly outside a small range of depth planes ($z = -10\,\mu\text{m}$ to $z = 10\,\mu\text{m}$), as predicted by theory (cyan dashed line). The RMM used in Fourier DiffuserScope (orange solid line) has slightly worse resolution at $z = 0$, but achieves good resolution across a much larger depth range ($z = -80\,\mu\text{m}$ to $z = 90\,\mu\text{m}$). Within this range, the resolution stays fairly close to the predicted multi-focal resolution (magenta dashed line).

oscillates between $2.0 \sim 3.8\,\mu\text{m}$ within a $170\,\mu\text{m}$ depth range. Thus, I conclude that the RMM design, relative to the MLA and RUM, slightly sacrifices lateral and axial resolving power at the native focal plane, but gains uniformly high performance across a large imaging volume.

Field-of-View

To compare the FOV of the three different designs, I simulate and reconstruct a 2D phantom that fills the objective FOV ($1.1 \times 1.1\,\text{mm}^2$), placed at the native focal plane of the objective (where the uni-focal microlenses have the best performance). The theory in Chap. 2.3 predicts that the random diffusers (RUM, RMM) should be able to reconstruct the whole object, while the MLA will only reconstruct $\text{FOV}_{\text{MLA}} = 554\,\mu\text{m}$.

To simulate the imaging pipeline accurately, I take into account the aberration from plano-convex microlenses, which means that the PSF at the edges of the FOV will have subtle differences from the center PSF. I divide the object into $10 \times 10\,\mu\text{m}^2$ blocks, convolve each

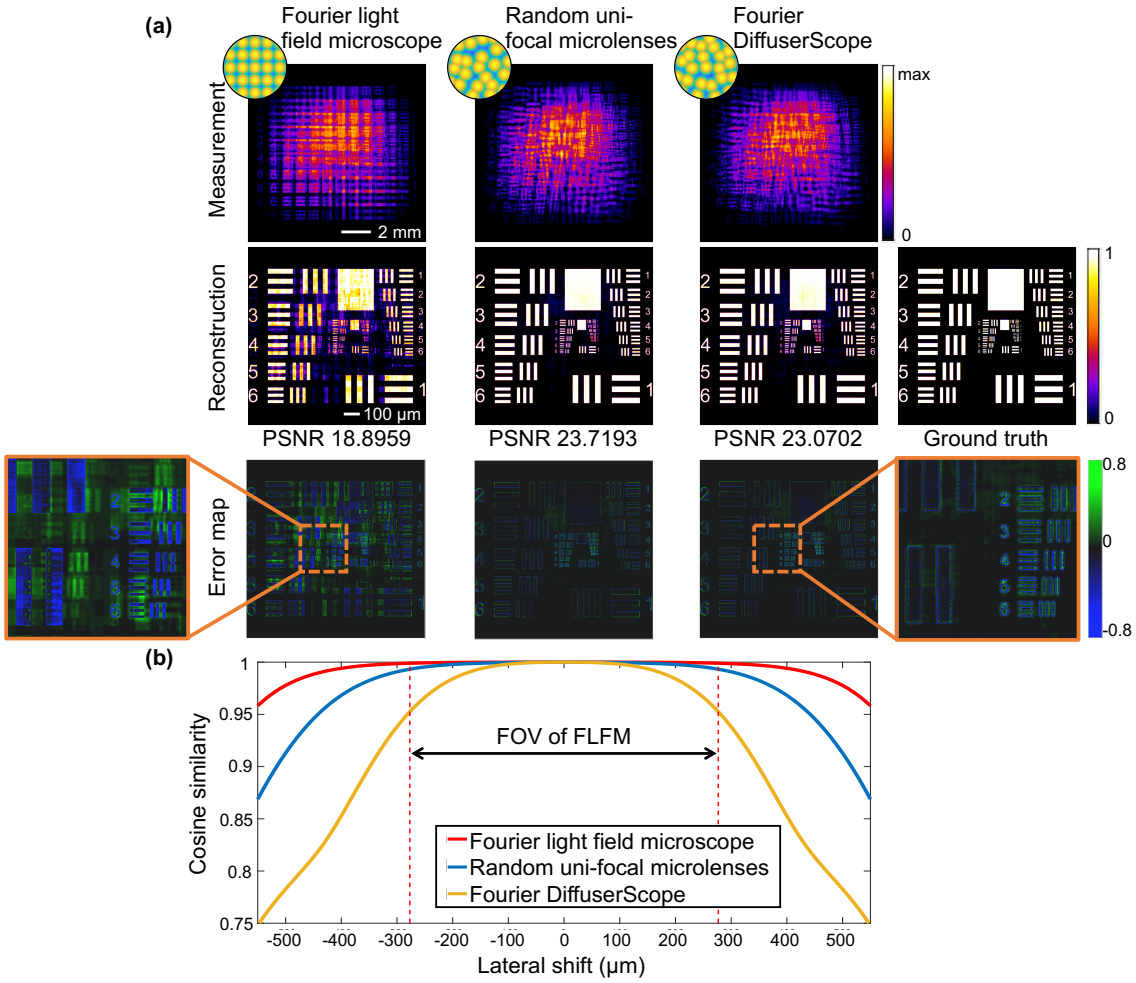


Figure 3.13: Simulations comparing field-of-view (FOV) for different microlens designs. (a) The FLFM (with MLA) reconstruction suffers from ghosting replicas (green regions in the error map) due to its periodic structure. Both the RUM and the RMM reconstruct the phantom successfully. The error of the random diffusers mainly occurs at sharp edges, which can be fixed by adding total variation regularization. Error = reconstruction − ground truth. (b) Cosine similarity between the on-axis PSF and off-axis PSFs is used to quantify the shift-invariance assumption. The MLA has the highest similarity value (red), but its FOV is limited by the microlens pitch. The similarity of RUM (blue) and RMM (orange) are all above 75% across the full objective FOV.

block's content with its corresponding PSF (calculated at the center of the block) and then sum up the convolution result from all the blocks to get the simulated measured image. After the spatially-variant block-wise convolution is done, I add 5% Gaussian noise to generate the

final measurement shown in Fig. 3.13(a), first row. The simulated MLA measurement has a periodic pattern because of its periodic PSF, while the diffuser measurements are more random.

To reconstruct the image, I deconvolve the simulated acquisition with a single on-axis PSF (assuming shift invariance) using Richardson-Lucy deconvolution [93, 94]. The result is shown in Fig. 3.13(a), second row. No regularization is added ($\tau = 0$ in Eq. 2.2) in order to compare the worst-case performance. The reconstruction using the MLA shows periodic replicas and large errors, due to the ambiguity of its PSFs. Restricting the FOV with a field stop eliminates this ambiguity at the cost of a reduced FOV. Both random diffusers, which do not have ambiguities in their PSFs, are able to reconstruct the whole object faithfully. The RUM has a slightly better peak signal-to-noise ratio (PSNR), since at the native focal plane all its microlenses are in focus, while only some of the RMM microlenses are. The error maps (error = reconstruction − ground truth) in Fig. 3.13(a) show significantly less error for the random microlenses designs than for the MLA, and errors for the random designs are mainly at edges of objects, which can be reduced by adding TV regularization to the reconstruction.

The shift-variance introduced by the aberrations in the simulation will cause model-mismatch that reduces the performance of the system when using a single-PSF reconstruction. To quantify the shift-variance, I examine the cosine similarity (normalized cross-correlation) between the on-axis and off-axis PSFs (Fig. 3.13(b)). At each lateral shift location, I register the off-axis PSF to the on-axis PSF [95] and calculate the inner product between them. The similarity value for randomly-located microlenses is at least 75% across the FOV, which is sufficient for single-PSF deconvolution [28]. At the edges of the FOV, the similarity goes down because the aberration and distortion are most severe at the periphery. The MLA provides the highest values because all microlenses have a regular shape and are of the same size, but the benefits are not useful because the FOV is actually limited by periodicity, as described in Chap. 2.3. The randomly distributed microlenses have irregular borders where the surfaces of neighboring microlenses are merged, which increases the aberration, and the multi-focal diffuser adds additional defocus aberration as compared to the uni-focal diffuser.

If high accuracy near the periphery is important, we can correct model mismatch with a spatially-varying deconvolution algorithm [29]. This algorithm calibrates the PSFs at multiple points across the FOV and interpolates them to find the PSFs at each position. It should give better reconstructions, but at a cost of significantly longer computation times and larger memory requirements. In my experimental system, the highest angle incident onto the diffuser (13 degrees) is much smaller than the highest angle (50 degrees) in [29], and the shift-invariant assumption holds well. Thus, I choose to use only a single PSF for each depth for computational efficiency.

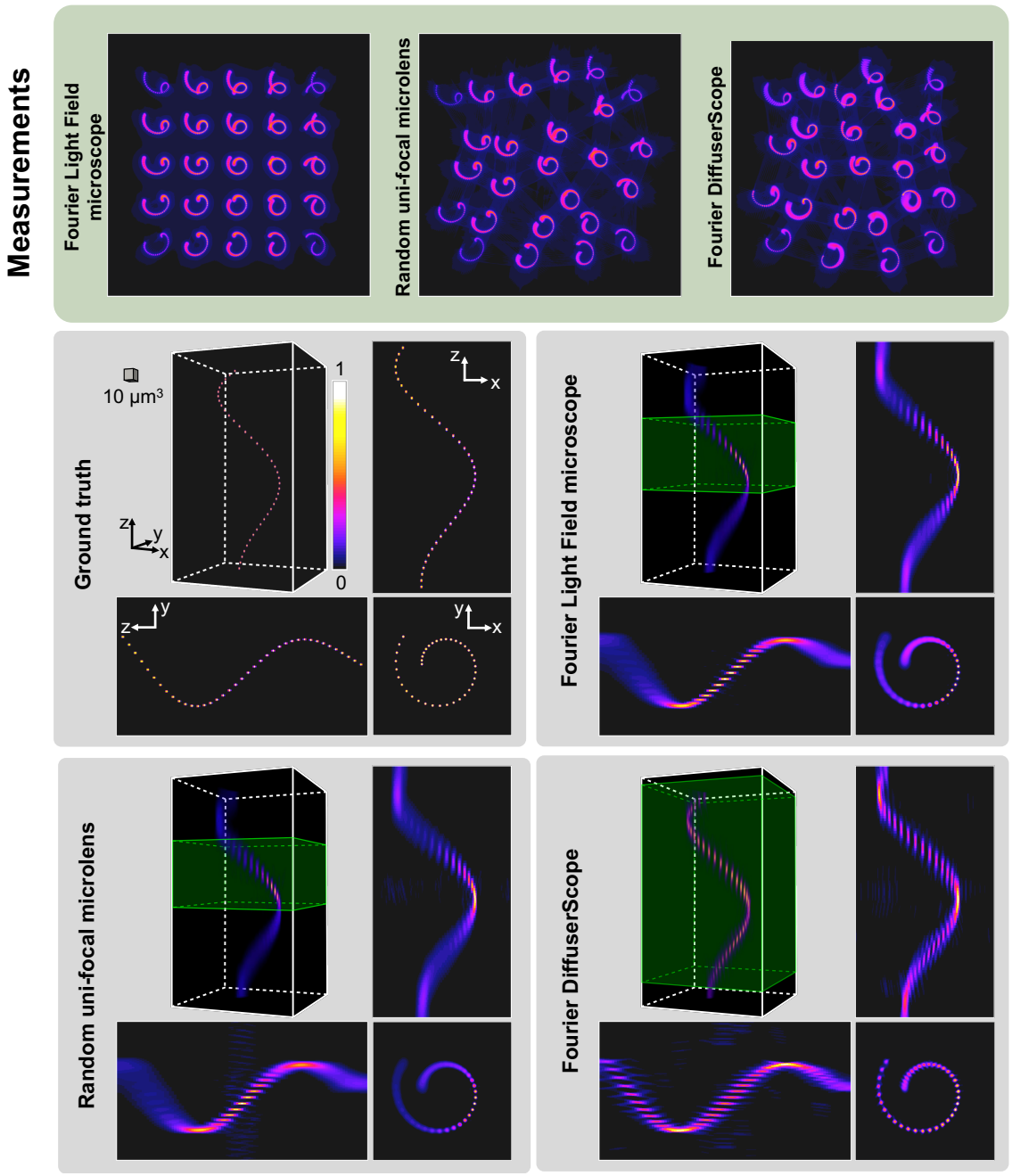


Figure 3.14: Simulated 2D measurements and 3D reconstructions of a sparse spiral object with different microlens designs. The ground truth object is a 200 µm-long spiral made of spheres. The Fourier light field microscope (MLA) and the RUM only resolve the spheres in the area around the native focal plane (green shaded area), whereas Fourier DiffuserScope (RMM) extends the depth range to cover almost the entire volume.

Depth Range

The two-point resolution in Fig. 3.12 can be used to estimate the depth range. For the uni-focal designs (MLA, RUM), the lateral resolution remains below its predicted in-focus value over a range of $\sim 20\,\mu m$, which is in agreement with the depth range predicted by Eq. 2.11 using the system parameters: $\mathrm{DOF}_{\mathrm{microlens}} = \frac{0.51\times1.33}{0.2^2} + \frac{1.33\times2}{6.5\times0.2} = 19\,\mu m$. The multi-focal design has stable performance from $z = -80\,\mu m$ to $z = 90\,\mu m$, demonstrating the improvement of depth range over uni-focal designs.

To demonstrate the depth range differences, I reconstruct a long 3D spiral of point sources covering a $200\,\mu m$ depth range (Fig. 3.14). This phantom contains 39 spheres of $2\,\mu m$ diameter, with the first one at $z = 95\,\mu m$ and the last one at $z = -95\,\mu m$, spaced axially by $5\,\mu m$ (resulting in a $3\,\mu m$ gap between spheres axially). The lateral distance between the spheres starts from $3\,\mu m$ (gap is $1\,\mu m$) at the center of the spiral and increases up to $7\,\mu m$ (gap is $5\,\mu m$) at the outer circle of the spiral. The lateral extent of the spiral ($66\,\mu m$) stays within the restricted FOV of the MLA to avoid ghosting artifacts. I divide the $200\,\mu m$-long object into 200 layers of 2D slices, implement the forward model in Eq. 2.1 and add 5% Gaussian noise to the simulated measurement (Fig. 3.14). The measurement contains 25 sub-images of the spiral object, one for each microlens which observes the spiral from a specific angle; in this way the 3D information is encoded into a single 2D acquisition. The simulated measurements highlight why the depth range of the multi-focal RMM (Fourier DiffuserScope) is much larger than the uni-focal design cases. For the uni-focal (MLA, RUM) cases, only the waist area of the spiral is sharp in all the sub-images. For the multi-focal RMM, different spiral sub-images contain different sharp areas; hence, more in-focus information about the entire depth range passes into the measurement.

The 3D reconstructions for each of the three cases are shown in Fig. 3.14. I use a PSF calibration stack with fewer (100) PSFs than were used in the forward simulation, to mimic practical axial sampling rates of continuous objects. The sparsity parameter ($\tau = 10^{-5}$) is hand-tuned and remains the same for all cases. From the reconstructions, the benefit of using multi-focal microlenses is obvious. 36 spheres are clearly resolved (from $z = -80\,\mu m$ to $z = 95\,\mu m$) with the RMM design, while only up to 13 spheres are resolved with the uni-focal designs (green shaded regions). The depth range of the three cases matches the depth range where the axial two-point resolution is under $5\,\mu m$. However, from both the two-point resolution result and the 3D object reconstruction result, the depth range of the RMM is slightly worse than predicted. This is likely due to most microlenses being out-of-focus at both ends of the targeted depth range, causing a lack of high-frequency information that is difficult to deconvolve.

3.6 Diffuser Fabrication Methods

When we consider purchasing a phase mask, off-the-shelf products are always the first choice whenever possible. However, to the best of my knowledge, there is no off-the-shelf

random multifocal lenslets available so far. The closest off-the-shelf product I found is the random pitch micro lens array from Nalux CO., LTD[96], but unfortunately these microlenses all have the same focal length.

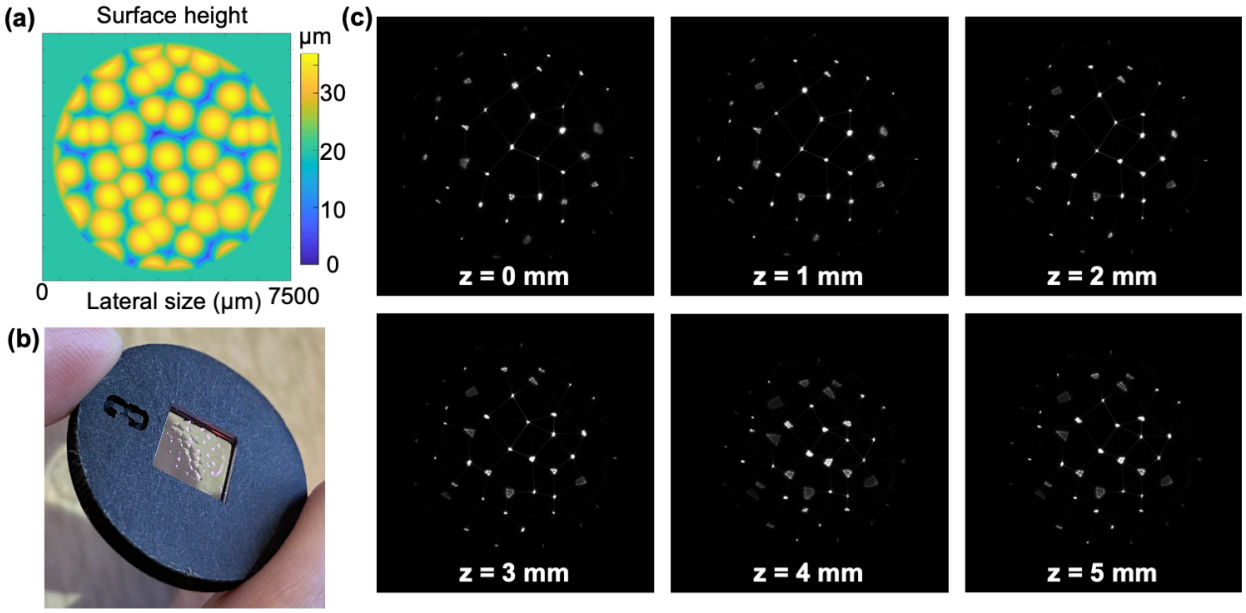


Figure 3.15: Random multifocal microlenses manufactured by PowerPhotonic Ltd. using the LightForge^{TM} fabrication service. (a) Surface height map sent for fabrication. The maximum surface sag is 38 µm, within the 50um range from the LightForge^{TM} spec sheet. The side length of the phase mask is 7500 µm and is plotted onto a 10 µm x 10 µm sampling grid as a 750×750 matrix. The edge of the circular outline is tapered to avoid sharp edges. (b) Manufactured RMM in a 1-inch sample holder. The surface is highly smooth with extremely low roughness. (c) Experimentally measured PSF stack with a 1 mm axial interval. Each microlens goes into and out of focus within the 5 mm range.

The industrial manufacturers use numerous ways to fabricate micro free-form optics with high quality at high costs. For example, diamond turning, laser writing and lithography are suitable for fast prototyping with few pieces, and injection molding with machined mold is suitable for mass production. If the free-form optics is composed of microlenses with a filling factor $< 100\%$, meaning that neighboring microlenses do not attach, technologies based on silicon processing including wet etching, melt and re-flow, are also popular choices. Here, in Fig. 3.15 I show an RMM sample manufactured by PowerPhotonic Ltd. using the LightForge^{TM} fabrication service. The maximum surface sag is 50 µm, and the surface map is plotted onto a 10 µm × 10 µm sampling grid. The surface shape matches perfectly with my design, and is highly smooth with extremely low roughness. There are sharp edges on

the original design where two microlenses merge. Due to physical limitations, the fabricated surfaces are rounded off into a concave or convex feature, which avoids light diffraction on the edge and doesn't affect the resulting PSFs. I then measure the PSF stack of the RMM sample in the experiment, by placing the RMM in front of a camera sensor and gradually moving a point source in the axial dimension. The point source is a fiber head with 0.2 NA. The laser light outcoming from the fiber head directly incidents on the RMM, which is then formed into PSFs on the sensor. In the PSF stack shown in Fig. 3.15 (c), the fiber head is moving away from the RMM at a 1 mm increment. Each microlens goes into and out of focus within the 5 mm range.

Since those commercial methods are either high-cost or require special skills, they are often not accessible to individuals. I also explore how to fabricate a random multifocal microlenses phase mask at low cost and medium quality. In the following section, I present four methods: indentation, droplet, two-photon polymerization, and inkjet printing. For each method, I explain the working principle, pros and cons, and provide initial results of a sample RMM phase mask.

Indentation and Replication

Indentation stands for deforming the surface of a material by indenting with a probe that has a know geometry and load. This method is widely used to measure the mechanical properties, such as hardness, of a material. Since the deformed surface remains the shape of the indenter, with a ball-shape indenter, indented material can serve as a negative mold for microlenses. Cirino et al [97] employed a CNC machine with a ball-nose indenter to precisely control the location and load of indenting and subsequently replicated the surface shape by injection molding on PMMA. Without modifying a CNC machine or acquiring special tools, we instead used hand indentation to fabricate an in-house RMM in [44].

We make a negative mold by randomly indenting polished copper using ball bearings with varying radii (Fig. 3.16 : Indentation). The choice of radii comes from the designed focal lengths and the refractive index of RMM material. The number of microlenses on the RMM is controlled by the number of hand indenting, but the location of each indenting is random. We then pour a thin layer of polydimethylsiloxane (PDMS) or UV-curing epoxy onto the copper mode, which is then placed inside a vacuum pump to remove air bubbles (Fig. 3.16 : Replication). For solidification, the PDMS material is baked in the oven, and epoxy is cured by UV light using a nail lamp. Finally, we carefully peel off the RMM containing convex-plano microlenses (Fig. 3.16 : Random multifocal microlenses). The main fabrication errors come from deformation error during indentation and shrinkage of the material during solidification. These have opposite effects, since the indented deformation will have bigger diameter than the indenter while the material shrinkage gives smaller diameter, so they offset each other to some extent. With some trial and error, we use ball bearings with diameters from 10 mm to 16 mm and PDMS material to approximate RMM with focal lengths from 12 mm to 21 mm.

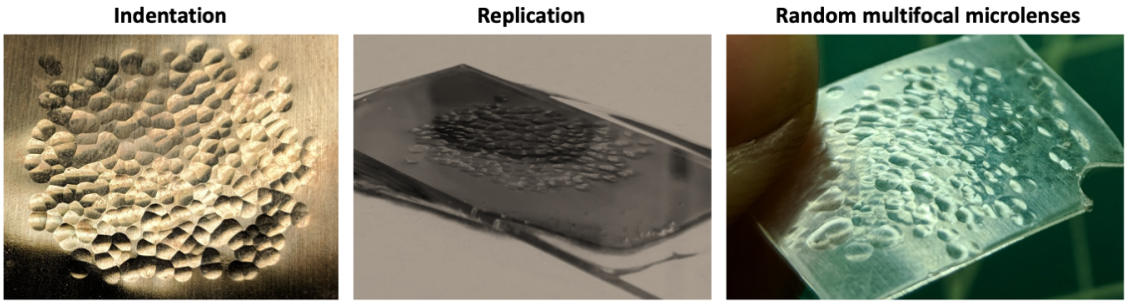


Figure 3.16: Hand indentation and replication of random multifocal microlenses. We make a negative mold by randomly indenting polished copper using ball bearings. The surface shape is then replicated into convex-plano microlenses using PDMS or UV-curing epoxy.

Droplet

Droplets make use of the surface tension of the material to form microlenses[98], and they have better surface quality compared to the replicated one. We fabricated a phase mask with a droplet-based technique for the flat diffuser microscope in [29]. Drops of optical epoxy (Norland 63 [99]) were formed and cured on a hydrophobic surface. We then handpick the droplets with suitable sizes under a microscope, and transfer those onto a glass coverslip to form the RMM, generating lenslets with approximately 250 µm diameter. Due to the strong surface tension of the epoxy, the droplets have smaller radii and thus smaller focal lengths than desired. The final RMM is index-matched with PDMS to increase the microlens focal length to about 1.5 mm.

Two-Photon Polymerization

3D printing, also known as additive manufacturing, is a fabrication process by adding material layer by layer, which is in contrast to traditional methods involving subtracting material. For typical 3D printers based on stereolithography (SLA), material jetting, or selective Laser sintering (SLS) technologies, currently the minimum layer thickness is bigger than 15 µm. However, the RMM requires a surface height resolution of 1 µm over a FOV of at least 5 mm.

Two-photon polymerization is a newly developed type of 3D printing to fabricate micro-optics on the sub-micron resolution scale [100]. Although photoresist is only sensitive to UV light (350–400 nm) and the two-photon polymerization uses a near-infrared femtosecond laser (780–800 nm Ti:sapphire), the material absorbs two photons simultaneously in a small focus area that leads to a polymerization process. In contrary to the SLA technique which directly uses a UV laser, the two-photon technique can achieve a resolution beyond the

diffraction limit of the objective due to the nonlinear nature of the two-photon polymerization process [101]. In a subsequent step, similar to other 3D printing technologies, the redundant photoresist is washed out to reveal the 3D printed structure.

We fabricate RMM samples with the Nanoscribe two-photon polymerization machine, Photonic Professional GT2, located in the Stanford Nanofabrication Facility. We upload a 3D STL file of the RMM design to DeScribe, a print job developing software for Nanoscribe, then a workflow is generated with user-chosen printing parameters including scan speed, laser power, stitch overlapping, and so on. Nanoscribe uses planar galvanometric scanning to precisely control the polymerization location within the FOV of the objective and the biggest FOV has a diameter of ∼400 µm with the 25× Nanoscribe objective. If larger objects need to be printed, several blocks need to be stitched together by moving the substrate with a mechanical stage.

As shown in Fig. 3.17, we apply Nanoscribe to two print jobs: a large RMM design with a 4.2 mm side length and an average microlens diameter of ∼900 µm, and a small RMM design with a 1.8 mm side length and an average microlens diameter of ∼300 µm used in [32]. Both RMMs are printed with the 25×, 0.8 NA objective and 1 µm layer thickness, paired with the Nanoscribe IP-S photoresist since it provides smooth surfaces and low shrinkage. For the large RMM design (Fig. 3.17 (a)), several blocks are needed to piece together a single microlens. The offset and tilt error from the mechanical stage cause uneven seams cutting through the center of the microlenses. The stitching artifacts can be very detrimental to the RMM's focusing capability. As seen in the experimental PSF, the focal points are blurry or even split into 2-3 sub-focal points. Fortunately, for the small RMM design where each microlens stays within the objective's FOV, we can employ the adaptive stitching algorithm [102] to arrange the stitching seam to be at the microlenses' edges. The method to find out the edges is similar to the Voronoi segmentation. Given a set of centers and radii of circles in a plane, the task is to assign each coordinate in the plane to one of the given circles. For each coordinate in the plane, the distance to all centers is calculated, and the closest center with the minimum distance is the assignment. As a result, a dividing line is defined by the fact that the distance to two or more centers is equal. Once the adaptive stitching mask is obtained, the writing instructions per block can be generated using TipSlicer [103]. Fig. 3.17 (b) compares the designed and experimental PSFs at three depth planes, showing a good match with some degradation at the end of the volume.

Conclusion

The indentation manufacturer method and droplet method are imprecise and thus lacks reproducibility. While as-built surface shape is calibrated and computationally accounted for, using more precise manufacturing methods could improve the surface quality of the diffuser and guarantee the system performance, especially for microscopic applications. Two-photon polymerization are suitable when the size of the microlenses is smaller than the objective's FOV. Otherwise, professional industrial manufacturers should be the top choice if budget allows.

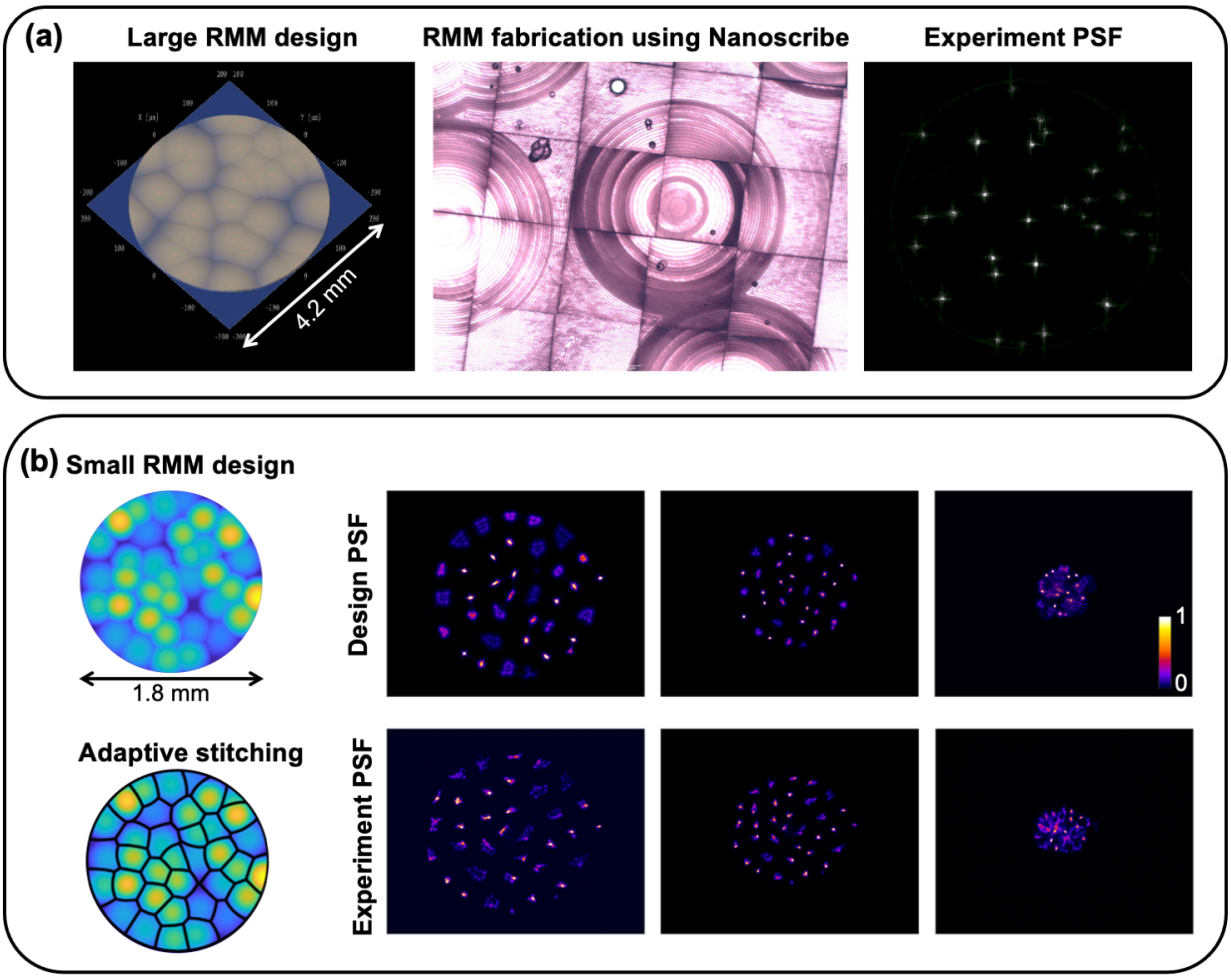


Figure 3.17: Random multifocal microlenses fabrication with Nanoscribe. (a) For the large RMM design with a 4.2 mm side length and an average microlens diameter of ∼900 µm, several blocks are needed to piece together a single microlens, and the stitching artifacts detriments the RMM's focusing capability. (b) For the small RMM design with a 1.8 mm side length and an average microlens diameter of ∼300 µm, adaptive stitching algorithm is employed to put the seams at the boundaries of the microlenses to mitigate artefacts. The designed and experimental PSFs show good agreement, with slight degradation at the edge of the volume.

Chapter 4

Algorithm Design of Fourier DiffuserScope

The raw acquisition on the camera encoded by the phase mask is a multiplexed measurement that cannot be directly interpreted for 3D imaging. So a reconstruction procedure is necessary to recover the depth information. If we know a physics-based forward model of the optical system, the reconstruction is an inverse of the forward when well-conditioned, which can be solved using one-step inverse algorithms, iterative convex optimization, and neural networks. In this chapter, I will discuss different reconstruction algorithms used for 2D, 3D, and 4D imaging (3D space plus 1D time).

For 2D imaging, since the forward model is a 2D convolution between the object and the PSF, the most straightforward method is to do one-step deconvolution algorithms, for example, inverse filtering, regularized inverse filtering, and Wiener filtering. However, when the camera is not big enough to capture the whole convolved image, the forward model becomes a 2D convolution followed by a crop function. In this case, partial information is lost during cropping so that the measurement can no longer be directly inverted. Iterative algorithms with regularizers (e.g., Richardson-Lucy, gradient descent, FISTA, ADMM, etc.) can successfully constrain the reconstruction to an optimal solution using compressed sensing theory. Pre-trained neural networks also contain prior knowledge of the system and is able to predict a reconstruction at a faster speed, without iterations. For 3D imaging, the measurement is a summation of multiple convolutions. Each pixel on the camera is traced back to more than one voxels in the volume. For example, in the dataset in Fig. 2.6, there are 50 $\times$ more voxels than the captured pixels. The information is hugely compressed such that the inverse procedure is ill-posed unless adding priors. For 4D imaging, the space-time model becomes critical to decouple spatially overlapping signals in the temporal domain, and vice versa. Making use of the temporal information between frames can help with the reconstruction by separating the background from the dynamic signals.

4.1 2D Plane Reconstruction Algorithms

The 2D image acquisition model is $\mathbf{y} = \mathbf{h} * \mathbf{x} + \mathbf{n} = \mathcal{F}^{-1}[\mathcal{F}(\mathbf{h})\mathcal{F}(\mathbf{x})] + \mathbf{n}$, where $\mathbf{y}$ is the sensor acquisition, $\mathbf{x}$ is the object of interest, $\mathbf{h}$ is the PSF, and $\mathbf{n}$ is the noise. The sources of noise are often divided into two general components: signal-dependent noise, namely shot-noise following a Poisson distribution; signal-independent noise, such as read-out noise, dark current, quantization noise, which are modeled as a Gaussian distribution. The aim of image reconstruction is to produce an estimate of the object $\hat{\mathbf{x}}$ from a noisy acquisition $\mathbf{y}$.

2D deconvolution methods fall into three categories: one-step filtering, iterative reconstructions, and neural networks.

- **One-Step Filtering**
  One-step filtering is fast, but tends to amplify noise. It is difficult to incorporate various types of priors, and only works for well-conditioned problems [104, 105]. The most direct approach of deconvolution is *inverse filtering* corresponding to the least-squares solution, $\hat{\mathbf{x}} = \mathcal{F}^{-1}[\mathcal{F}(\mathbf{y})/\mathcal{F}(\mathbf{h})]$. Inverse filtering is parameter-free, but the denominator term $1/\mathcal{F}(\mathbf{h})$ might contain close-to-zero values and drastically amplifies noise in the frequency domain. A simple fix is to add a constant to the denominator to avoid dividing by zero issues, termed *regularized inverse filtering*, where $\hat{\mathbf{x}} = \mathcal{F}^{-1}[\frac{\mathcal{F}(\mathbf{y})}{\mathcal{F}(\mathbf{h})+\lambda\mathbf{I}}]$. Another method to improve inverse filtering is to add a damping factor related to the SNR, yielding the well-known Wiener filtering [106],

$$\hat{\mathbf{x}} = \mathcal{F}^{-1}\left[\frac{|\mathcal{F}(\mathbf{h})|^2}{|\mathcal{F}(\mathbf{h})|^2 + 1/\text{SNR}} \cdot \frac{\mathcal{F}(\mathbf{y})}{\mathcal{F}(\mathbf{h})}\right], \text{ where SNR} = \frac{\text{signal mean}}{\text{noise std}}. \tag{4.1}$$

  The SNR term requires prior knowledge of the signal and noise, which is often impractical. A common approximation is to use the mean value of $\mathbf{y}$ as the signal term and use the standard deviation of $\mathbf{y}$ as the noise term. Wiener deconvolution with hand-tuned parameters generally achieves acceptable results when the inverse problem is well-conditioned, but fails in compressed sensing when the problem is ill-conditioned.

- **Iterative Reconstructions**
  Iterative reconstructions can solve ill-conditioned problems by combining prior information. The main drawback of iterative approaches is that the convergence speed is often too slow for real-time imaging.

  With an extended FOV, the convolved image, $\mathbf{h} * \mathbf{x}$, becomes wider than the camera size. Only the center part will be captured and the forward model becomes a 2D convolution followed by a crop function, $\mathbf{y} = \text{crop}(\mathbf{h} * \mathbf{x}) + \mathbf{n}$. In this case, the inverse problem becomes ill-posed, meaning that there exist an infinite number of solutions that would result in the same measurement. The general formulation is to use convex

optimization to iteratively minimize a loss function, which consists a data-fidelity term and a regularization term.

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\| \mathbf{y} - \text{crop}(\mathbf{h} * \mathbf{x}) \right\|_2^2 + \Gamma(x). \tag{4.2}$$

The data-fidelity term enforces that the reconstructed object, after a known imaging model, matches the measurement. The regularizer term, $\Gamma(x)$, enforces prior knowledge of the latent image. Some common priors on the object include non-negativity, native sparsity $||\cdot||_1$, gradient sparsity using total variation $||D(\cdot)||_1$, smoothness $||\cdot||_2$, wavelet sparsity $||W(\cdot)||_1$, etc. Once a physical problem is formatted into an optimization problem with constraints, various solvers can be applied to find the optimal solution. The detailed implementation of FISTA and ADMM of 2D diffuser imaging can be found in this tutorial [107], to name a few.

- **Neural Networks**
  Neural networks-based methods have been widely used in solving inverse problems, such as denoising, deconvolution, and super-resolution. In convolutional neural networks (CNN), we apply a sequence of linear filtering operations alternating with simple point-wise nonlinear activation functions. These filter weights are optimized during the training process with a minimization of the cost function. After training, the forward operation of a CNN is very fast with the help of hardware (GPU in our case) acceleration, and the gradients w.r.t. these weight parameters are computable via the chain rule and the backpropagation algorithm. Efficient implementation of CNN models has been realized in standard libraries and frameworks, such as TensorFlow [108] and PyTorch [109]. Once training is finished and the parameters are pinned down, the inference step is lightning fast. However, a CNN model is not a silver bullet without a large amount of high-quality data (i.e., garbage in, garbage out). Fortunately, in the case of 2D plane reconstruction, this is not a problem. Generating simulated training data is fairly straightforward because the forward model (convolution + cropping) is readily available and easy to compute. As for the network architecture, there are a great variety of choices [110–112], such as the number of convolutional layers, filter sizes and types of nonlinear activation functions, etc. The number of learnable parameters of a CNN model is a rough indicator of its complexity as well as the computational cost. In fact, neural networks can also be built up with unrolling iterative optimization algorithms (e.g., ADMM, FISTA, etc.), turning each iteration into a layer of neural network [30, 113, 114].

In this section, I propose a novel neural network architecture, termed *FourierNet*, to tackle 2D diffuser imaging reconstruction. The architecture of FourierNet is based on a physical understanding of the optical forward model and is tailored for image deconvolution with a large PSF.

FourierNet Overview

Because the size of the diffuser PSF can be as big as the camera sensor, the measured image could potentially encode signals from any pixel of the original image. Therefore, a reconstruction neural network must take into consideration all the pixels of the measured image, and the conventional small kernels (3×3 or 5×5) for CNN models are not enough. Here, I instead apply a kernel with the same size of the input image to achieve such a global context.

Note that such a choice of image-size kernel is computationally heavy in conventional real-space convolution, but can be efficiently implemented in the Fourier space, where the convolution is converted to a element-wise multiplication of the Fourier-transformed image and the weight matrix. Compared with a 2D real-space convolution with a computational complexity of $\mathcal{O}(N^4)$, the Fourier-space multiplication cuts the computational complexity down to $O(N^2 \log^2 N)$. For input images with 256×256 pixels, we can achieve a speedup of two orders of magnitudes. We name this architect *FourierNet*.

Since the neural network is a universal function approximation, adding more weights to the neural network leads to better approximation of high-dimensional functions, namely more complicated objects in our case. The extra weights can be added in series (e.g., deep ResNet [115]) or in parallel (e.g., UNet [110]). As shown in Fig. 4.3, UNet, through cascading downsampling and upsampling convolutional layers to capture features at different scales, is especially suitable for imaging processing tasks. Similar approaches can be applied in the Fourier space [46, 116, 117], where downsampling in real-space corresponds to a cropping in the Fourier space, and vice versa. Such a multi-scale Fourier convolution network is called *Fourier-UNet*, as shown in Fig. 4.3.

FFT-based Conv2d

Conv2d is a 2D convolutional layer that outputs the 2D cross-correlation between the input signal and the layer's weights. In TensorFlow and PyTorch libraries, the built-in Conv2d layers are implemented using conventional sliding inner-product methods, suited for small kernel size. If the input image has $N \times N$ pixels and the kernel size is $M \times M$, the time complexity of conventional Conv2d is $\mathcal{O}(N^2M^2)$, considering only 1 kernel. Alternatively, if the cross-correlation is calculated using Fourier transform, the time complexity of Fourier transforming the input image is $\mathcal{O}(N^2 \log^2 N)$, the time complexity of element-wise multiplication in the Fourier space is $\mathcal{O}(N^2)$. To make our life easier, I directly use the kernel matrix in the Fourier space. In total, the time complexity is $\mathcal{O}(N^2 \log^2 N + N^2) = \mathcal{O}(N^2 \log^2 N)$. The time complexity of conventional Conv2d increases at a rate of M^2, while the time complexity of FFT-based Conv2d is independent of the kernel size. For $N = 256$ and $M = 256$, $\frac{N^2M^2}{N^2 \log^2 N} = 11300$, so it is essential to use FFT-based Conv2d for big kernel sizes.

Our FFT-based Conv2d layer implementation is shown in Fig. 4.1. The input image is zero-padded to double its size in the real space such that the frequency resolution is increased. Because the images are real signals, I choose to do Fourier transform with the *rfft*

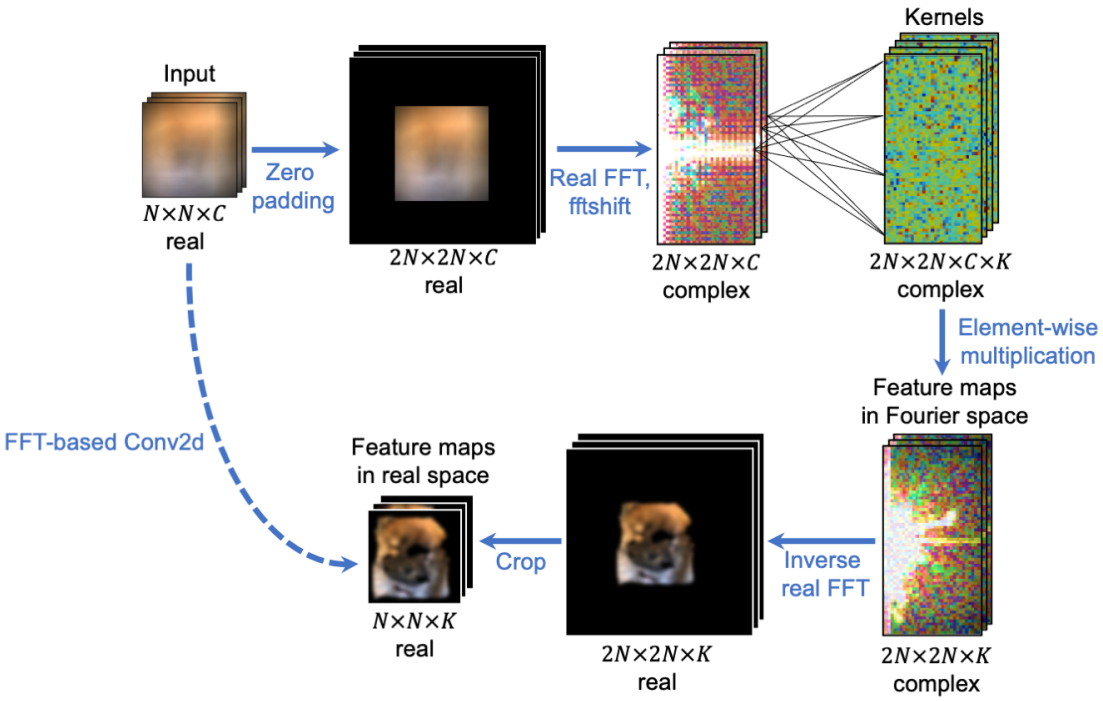


Figure 4.1: FFT-based Conv2d layer implementation

(discrete Fourier Transform for real input) algorithm implemented in PyTorch. A follow-up element-wise multiplication is done in the Fourier space. Note that the usage of *rfft* not only keeps the symmetry of the Fourier coefficients ($C_\omega = C^*_{-\omega}$), but also reduces the number of complex parameters in the Fourier-space weight matrix. The products after the multiplication is converted back to the real space using *irfft* (inverse discrete Fourier Transform for real input), and cropped to match the size of the input images. I implement the FFT-based Conv2d as a neural network module with the same input and output structure as conventional Conv2d.

Dataset

In this work, I use a modified Stanford Dogs dataset [118] to train and test our neural network models. The original dataset contains 20,580 images with varied sizes (different resolution and aspects) covering 120 breeds of dogs. I crop them to square images and resize to 256×256 pixels, as shown in Fig. 4.2(a). In addition to the ground-truth images, I use the forward model to generate simulated measurements by performing a convolution with respect to the PSF and then cropping the result, $\mathbf{y} = \text{crop}(\mathbf{h} * \mathbf{x})$. This procedure is illustrated in Fig. 4.2(b). Note that the PSF in the forward model is acquired from an actual

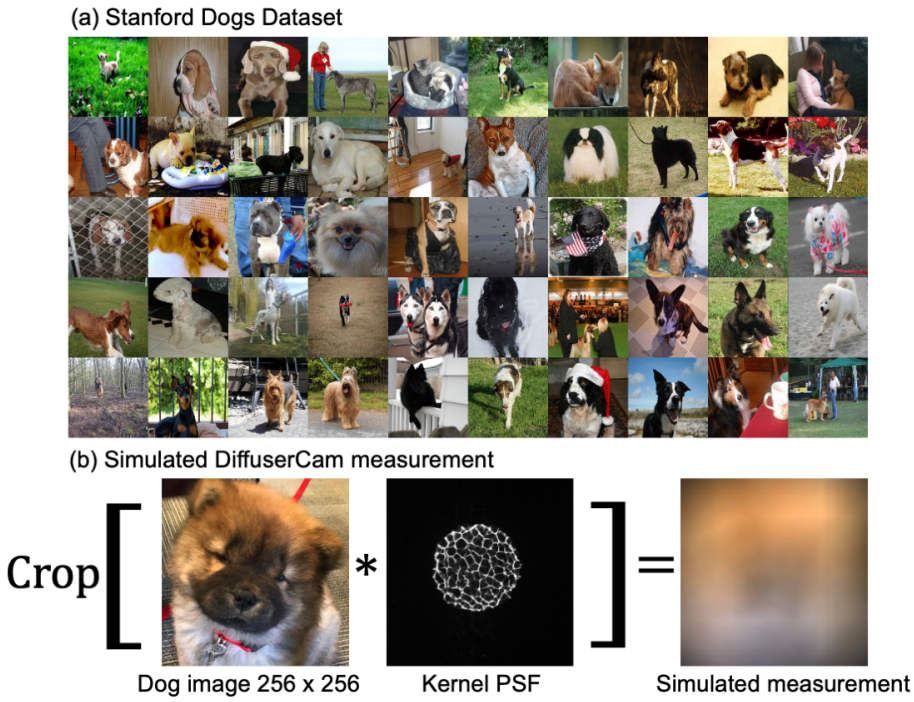


Figure 4.2: The dataset contains 20,580 image pairs, of which 80% (16464 images) are training samples, 19% (3910 images) are validation samples, and 1% (206 images) are testing samples. Each pair of data contains a ground-truth image and a simulated measurement image. (a) Ground-truth data. I modified the Stanford Dogs dataset [118] to square images with 256×256 pixels. (b) I generate simulated measurements by convolving the ground-truth image with the PSF and then cropping the result, assuming no noise.

Fourier DiffuserScope using a random Gaussian diffuser, and no noise is considered in the forward model. I use 80% of data (16464 images) as training samples, 19% (3910 images) as validation samples, and 1% (206 images) as testing samples. A batch size of 20 is used during training.

Neural Network Architecture

I built 4 different neural network models to investigate the effect of kernel sizes and the total number of parameters. First, I define the *Vanilla CNN* to be two identical blocks of 2D real-space convolution with 32 different 3×3 kernels followed by a ReLU, as shown in Fig. 4.3. To capture features at different scales, I beef up the Vanilla CNN into a 3-layer UNet, where the input images pass through a downsampling and then upsampling path with skip connections. Note that the 3-layer UNet still adopts a 3×3 kernel in the Conv2D layer.

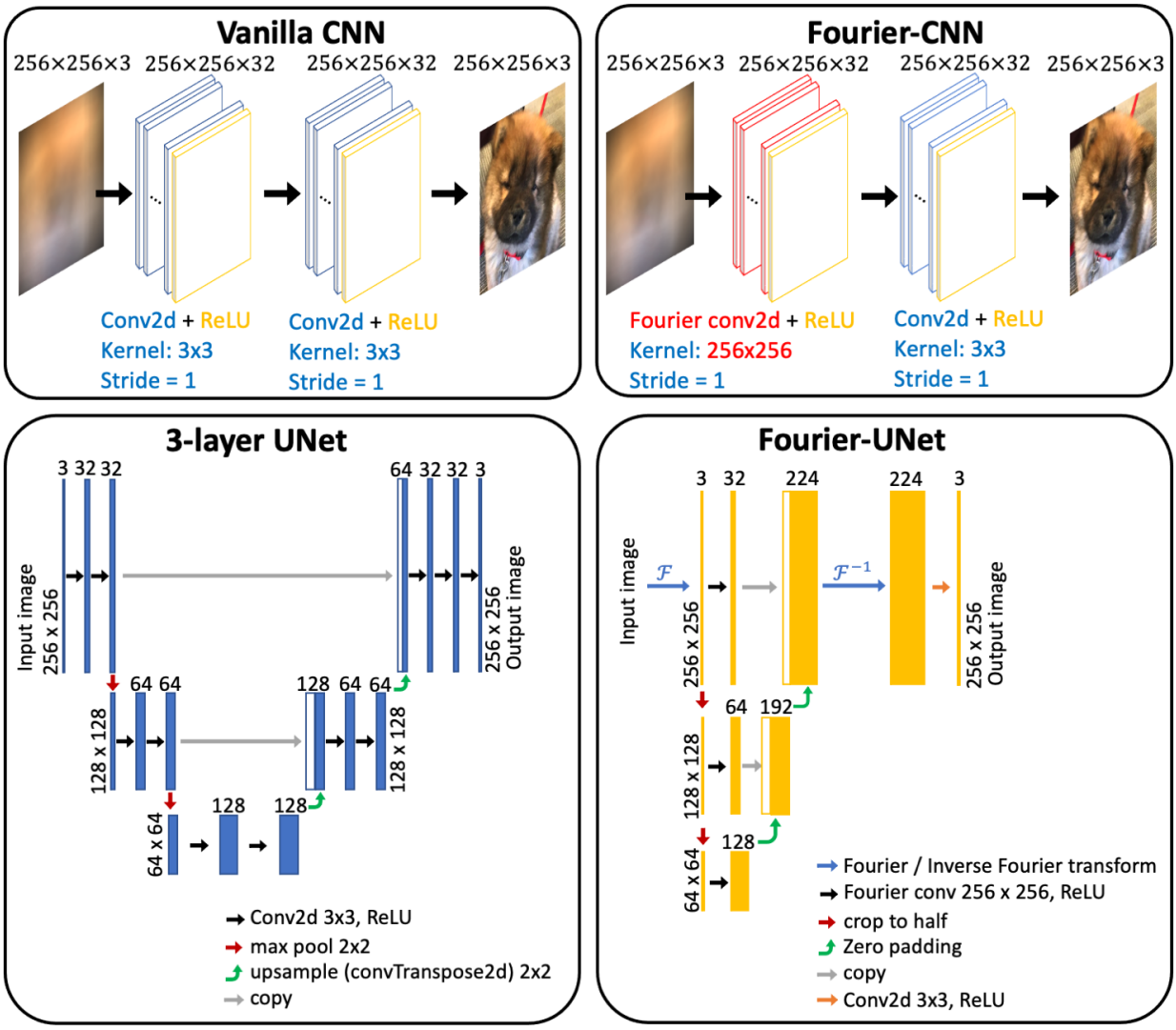


Figure 4.3: Neural network architecture of (a) Vanilla CNN, (b) 3-layer UNet, (c) Fourier-CNN, and (d) Fourier-UNet. Both input and output images have a size of 256×256 pixels with RGB channels.

Fourier-CNN, unlike Vanilla CNN, first performs the convolution in the Fourier space to learn the features with a global context, then uses a conventional Conv2D layer to manage fine tuning. An UNet-style generalization of Fourier-CNN is the so-called Fourier-UNet in Fig. 4.3, where Fourier-transformed input images are distilled further with cropping and multiplication with a correspondingly smaller weight matrix. The results are packed into a large

matrix in the Fourier space before an inverse Fourier transform. A follow-up conventional Conv2D layer is again used to learn the features. Note that our architecture is simpler and faster than the one in [46] where an inverse Fourier transform is performed at each level.

Results

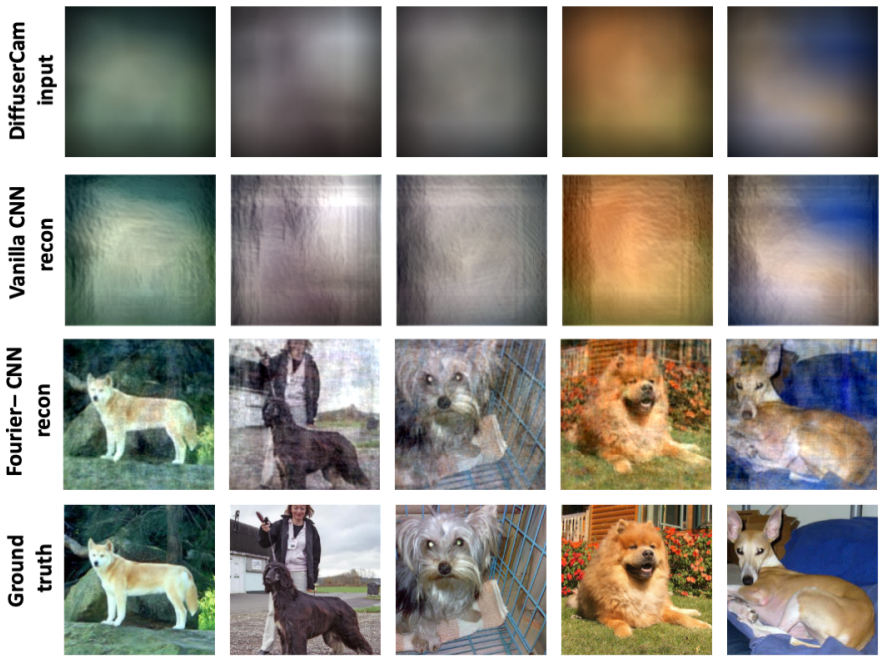


Figure 4.4: Sample reconstructed images with neural network models comparing Vanilla CNN and Fourier-CNN. (a) Input images simulated with the DiffuserCam forward model. Output images with (b) Vanilla CNN, (c) Fourier-CNN. (d) Ground-truth of the input images.

As expected, the small kernel size of Vanilla Net leads to poor performance of reconstruction, as illustrated in Fig. 4.4, while Fourier-CNN achieves much better results with the global context. Applying downsampling and upsampling techniques in UNet enables multi-scale feature recognition, leading to an effectively larger kernel size. The details of images can thus be reproduced in Fig. 4.5. It is beneficial to have both a large kernel in the first place and multi-scale feature recognition, for which the Fourier-UNet is designed.

To better assess the neural network models, I plot the validation loss as a function of epoches, as shown in Fig. 4.6. Fourier CNN and Fourier-UNet are much more stable and converge faster than Vanilla CNN and 3-layer UNet. Fourier-UNet achieves the best

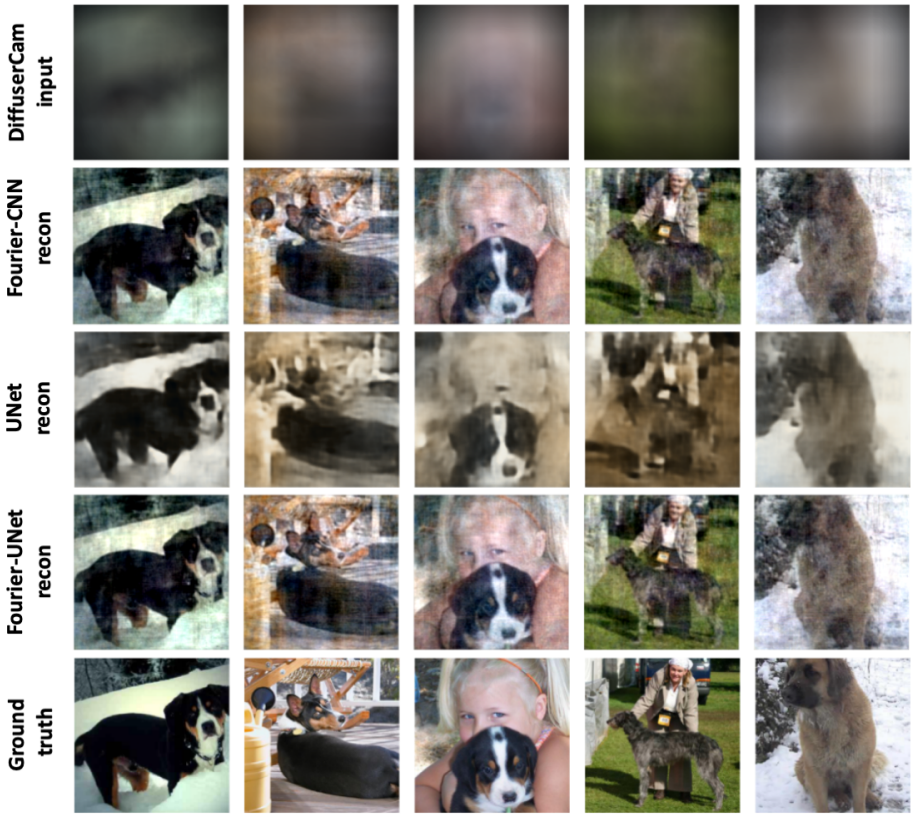


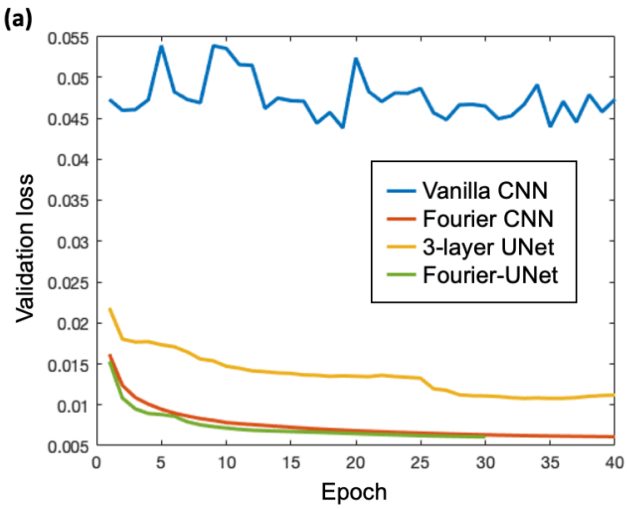
Figure 4.5: Sample reconstructed images with neural network models comparing Fourier-CNN, UNet and Fourier-UNet. (a) Input images simulated with the DiffuserCam forward model. Output images with (b) Fourier-CNN, (c) UNet, and (d) Fourier-UNet. (e) Ground-truth of the input images.

reconstruction performance, however, the usage of multi-scale feature recognition in Fourier-UNet only have marginal gain compared with flat Fourier CNN. Fourier CNN is therefore the most efficient neural network model in this study.

I demonstrate the power of Fourier CNN using natural images and smooth Gaussian diffusers. This is a general workflow for deconvolution with big kernels, and can be applied to microscopic applications as well.

4.2 3D Volume Reconstruction Algorithms

To reconstruct a 3D volume from a single 2D measurement, the inverse problem is highly ill-posed, and we must add priors to recover a reasonable solution. Numerous methods



(b)

	Learnable parameters	Training Time per epoch
Vanilla CNN	1798	35 sec
Fourier-CNN	1.26×10^7	1 min 55 sec
3-layer UNet	1.22×10^5	3 min 26 sec
Fourie- UNet	1.47×10^7	3 min 18 sec

Figure 4.6: Validation loss comparison

have been applied to 3D diffuser imaging and light-field imaging reconstruction, including Richardson-Lucy [61], FISTA [29], ADMM [28], CNN [119], etc. In this section, I introduce how to use the Richardson-Lucy (RL) optimization solver as an example for 3D volume reconstruction. More details of the FISTA and ADMM implementation of diffuser image reconstruction can be found in [120].

Richardson-Lucy as an Inverse Solver

RL [93, 94] is a maximum-likelihood estimation assuming that the noise follows a Poisson distribution, $\mathbf{y} \sim \mathcal{P}(\mathbf{Ax})$, where $\mathcal{P}$ models a Poisson process and $\mathbf{A}$ summarizes the image formation model. The log-likelihood of having a measurement $\mathbf{y}$ given an object $\mathbf{x}$ is

$$\log(L(x)) = \log(p(\mathbf{y}|\mathbf{x})) = \mathbf{y}^T \log(\mathbf{Ax}) - \mathbf{1}^T(\mathbf{Ax}) - \sum_{i=1}^{M} \log(\mathbf{y}_i!), \tag{4.3}$$

where the log operation is applied element-wise and M represents the total number of pixels on sensor. Since for a given measurement, the last term $\sum_{i=1}^{M} \log(\mathbf{y}_i!)$ is a constant, the gradient of the log-likelihood has the following expression,

$$\nabla \log(L(x)) = \mathbf{A}^T\left(\frac{\mathbf{y}}{\mathbf{Ax}}\right) - \mathbf{A}^T\mathbf{1}, \tag{4.4}$$

where the division $\frac{\mathbf{y}}{\mathbf{Hx}}$ is element-wise. Once the iterative algorithm converges to the optimal solution, the log-likelihood won't change, and its gradient equals zero: $\nabla \log(L(x)) = 0$.

From this conclusion, the multiplicative update of RL writes

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} \times \frac{\mathbf{A}^T\left(\frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(k)}}\right)}{\mathbf{A}^T\mathbf{1}}. \tag{4.5}$$

Here, both the multiplication $\times$ and the division $\frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(k)}}$, $\frac{\cdot}{\mathbf{A}^T\mathbf{1}}$ are implemented element-wise. If the initialization $\mathbf{x}_0 \geq \mathbf{0}$, during the multiplicative update, the non-negative nature is preserved in the following iterations.

Forward and Backward Function

Although the derivation of RL update in Eq. 4.5 is carried out in matrix form, this algorithm is found useful in a wide range of matrix-free problems, as long as we replace $\mathbf{A}$ with the forward function and replace $\mathbf{A}^T$ with the backward function.

The 3D forward model has been introduced previously in Eq. 2.1. With FFT acceleration, the forward function implementation is as follows.

$$\text{Forward function: } \mathbf{A}\mathbf{x} = \sum_z \mathcal{F}^{-1}\big(\mathcal{F}(\mathbf{h}_z)\mathcal{F}(\mathbf{x}_z)\big) = \sum_z \mathcal{A}_z\mathbf{x}_z. \tag{4.6}$$

For a given depth z, the convolution operator can be represented by a Toeplitz matrix, $\mathcal{A}_z = \mathcal{F}^{-1}diag(\widetilde{\mathbf{h}_z})\mathcal{F}$. $\widetilde{\mathbf{h}_z}$ is the discrete Fourier transform of $\mathbf{h}_z$, and is also the eigenvalues of the Toeplitz matrix $\mathcal{A}_z$. $\mathcal{F}$ is the discrete FFT matrix.

The backward function, on the other hand, is $\mathbf{A}^T\mathbf{x}$. For each z layer, the operator turns into $\mathcal{A}_z^T = \mathcal{F}^\dagger diag(\widetilde{\mathbf{h}_z})^\dagger(\mathcal{F}^{-1})^\dagger$, where † stands for conjugate transpose of complex matrices. Since the discrete FFT matrix is unitary with a scaling factor, $\mathcal{F}^{-1} = \frac{1}{n}\mathcal{F}^\dagger$. Therefore, $\mathcal{A}_z^T = \mathcal{F}^{-1}diag(\widetilde{\mathbf{h}_z})^\dagger\mathcal{F}$. This means that the backward function in the Fourier space has the following implementation.

$$\text{Backward function: } \mathbf{A}\mathbf{x} = \sum_z \mathcal{A}_z^T\mathbf{x}_z = \sum_z \mathcal{F}^{-1}\Big(\big(\mathcal{F}(\mathbf{h}_z)\big)^\dagger\mathcal{F}(\mathbf{x}_z)\Big). \tag{4.7}$$

In summary, Algorithm 1 outlines the pseudo-code for the RL deconvolution for 3D diffuser image reconstruction. It is worth noting that, similar to FISTA and ADMM, regularizers can be added to RL to help with convergence speed and suppress noise. Interested readers can refer to [121] for details.

4.3 4D Space-Time Reconstruction Algorithms

The most straightforward approach to reconstructing a video is to apply the reconstruction algorithm frame by frame, as it does in Fig. 2.6. However, the background fluorescence signal originating from static components breaks down the sparsity assumption and decreases the signal-to-noise ratio of the dynamic fluorescence signals. Moreover, the temporal behavior of calcium indicators can be employed as a constraint in the space-time domain [86,

Algorithm 1 Richardson-Lucy 3D reconstruction

$\mathbf{x} \leftarrow ones(W, H)$; ▷ Initialization
while $k \leq Iter_{max}$ **do**
  $\hat{\mathbf{y}} \leftarrow zeros(W, H)$
  for $z = 1$ to z_{max} **do** ▷ Forward function
    $\hat{\mathbf{y}} \leftarrow \hat{\mathbf{y}} + ifft\ [\ fft(\mathbf{h}_z)\ fft(\mathbf{x}_z)\ ]$
  end for
  $\mathbf{r} \leftarrow \mathbf{y}/\hat{\mathbf{y}}$
  $\hat{\mathbf{r}} \leftarrow ones(W, H)$
  $\boldsymbol{\mu} \leftarrow zeros(W, H)$
  $\hat{\boldsymbol{\mu}} \leftarrow zeros(W, H)$
  for $z = 1$ to z_{max} **do** ▷ Backward function
    $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + ifft\ [\ conj(\ fft(\mathbf{h}_z))\ fft(\mathbf{r})\ ]$
    $\hat{\boldsymbol{\mu}} \leftarrow \hat{\boldsymbol{\mu}} + ifft\ [\ conj(\ fft(\mathbf{h}_z))\ fft(\hat{\mathbf{r}})\ ]$
  end for
  $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}/\hat{\boldsymbol{\mu}}$ ▷ Normalization
  $\mathbf{x} \leftarrow \mathbf{x} \times \boldsymbol{\mu}$ ▷ Update
end while

122], which can demix the spatial and temporal footprints of each neuron. In this section, I make use of the temporal information between frames to help with the reconstruction by separating the background from the dynamic signals.

Inspired by background removal algorithms based on Robust Principle Component Analysis (RPCA) [123, 124], I perform sparse and low-rank decomposition on our raw video, which is a spatial-temporal matrix. The temporal sparse component contains the dynamic moving objects, and the temporal low-rank component contains the static object and noise which serve as background. This sparse and low-rank decomposition separates the raw video into a temporal sparse video and a temporal low-rank video. I then use the inverse solver to reconstruct the two components separately with different sparsity priors: native sparsity for the sparse component and total variation sparsity for the low-rank component. In the experiments with a hydra sample tagged with GCaMP, the tentacle movements are exaggerated and the mostly stationary body is separated. The resulting reconstruction highlights the hydra movements more than the frame-to-frame method.

Methods

Fourier DiffuserScope is a fluorescence microscope with a diffuser (random multifocal lenslets) in the Fourier space and the sensor placed immediately after (Fig. 4.7a). The raw camera measurement contains overlapping sub-images; each sub-image comes from a lenslet imaging the 3D object from a different perspective and with a different focal plane. From the captured raw video of a dynamic sample, I use non-negative RPCA to perform a sparse

and low-rank decomposition:

$$\min_{\mathbf{L}\geq\mathbf{0},\ \mathbf{S}\geq\mathbf{0}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \quad \text{subject to} \quad \mathbf{A} = \mathbf{L} + \mathbf{S}, \tag{4.8}$$

where matrix A is a spatial-temporal matrix representing the raw video. Each column of A represents a 2D frame in the video. I aim to decompose A into a low-rank matrix L and a sparse matrix S, by minimizing the nuclear norm of L and the $l1$ norm of S with tuning parameter λ [123].

Next, I feed the sparse video and the low-rank video into the sparsity-constrained reconstruction algorithm in Eq. 4.9. Each frame, a column in L or S, is reconstructed independently. Here, y is a 2D frame, v represents the reconstructed 3D volume, H is a matrix containing the calibrated 3D point spread function (PSF) stack, and Φ transforms v into a domain where the 3D object is sparse. The temporal sparse component is sparse in the native 3D space and Φ is an identity matrix; the gradient of the temporal low-rank component is sparse and Φ is a finite-difference operator. After reconstruction, I use open-source software, Icy [125], to render and visualize the final results.

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}\geq\mathbf{0}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{v}\|_2^2 + \beta\|\Phi\mathbf{v}\|_1. \tag{4.9}$$

Results

I recorded a freely-moving hydra, tagged with GCaMP, at 25 fps using our Fourier DiffuserScope [44]. The volumetric field-of-view is $1\times1\,\text{mm}^2$ laterally and 400 µm in depth. The voxel size in the reconstructed 3D volume is 2 µm laterally and 3 µm axially. A total of 500 frames (20 seconds) were captured. The non-negative RPCA decomposition in Eq. 4.8 is solved using the alternating direction method of multipliers (ADMM) and the 3D reconstruction in Eq. 4.9 is optimized using the fast iterative shrinkage-thresholding algorithm (FISTA [126]) on an NVIDIA TITAN X (Pascal) GPU. We view the final 3D reconstruction by stacking layers of each cross-section of the object, then repeating for each time frame to generate a video. Results show improved visibility for the active and moving portions of the hydra, specifically the tentacles (near the top). In contrast, the pedal disc (near the bottom) is less actively moving and more prominent in the low-rank component.

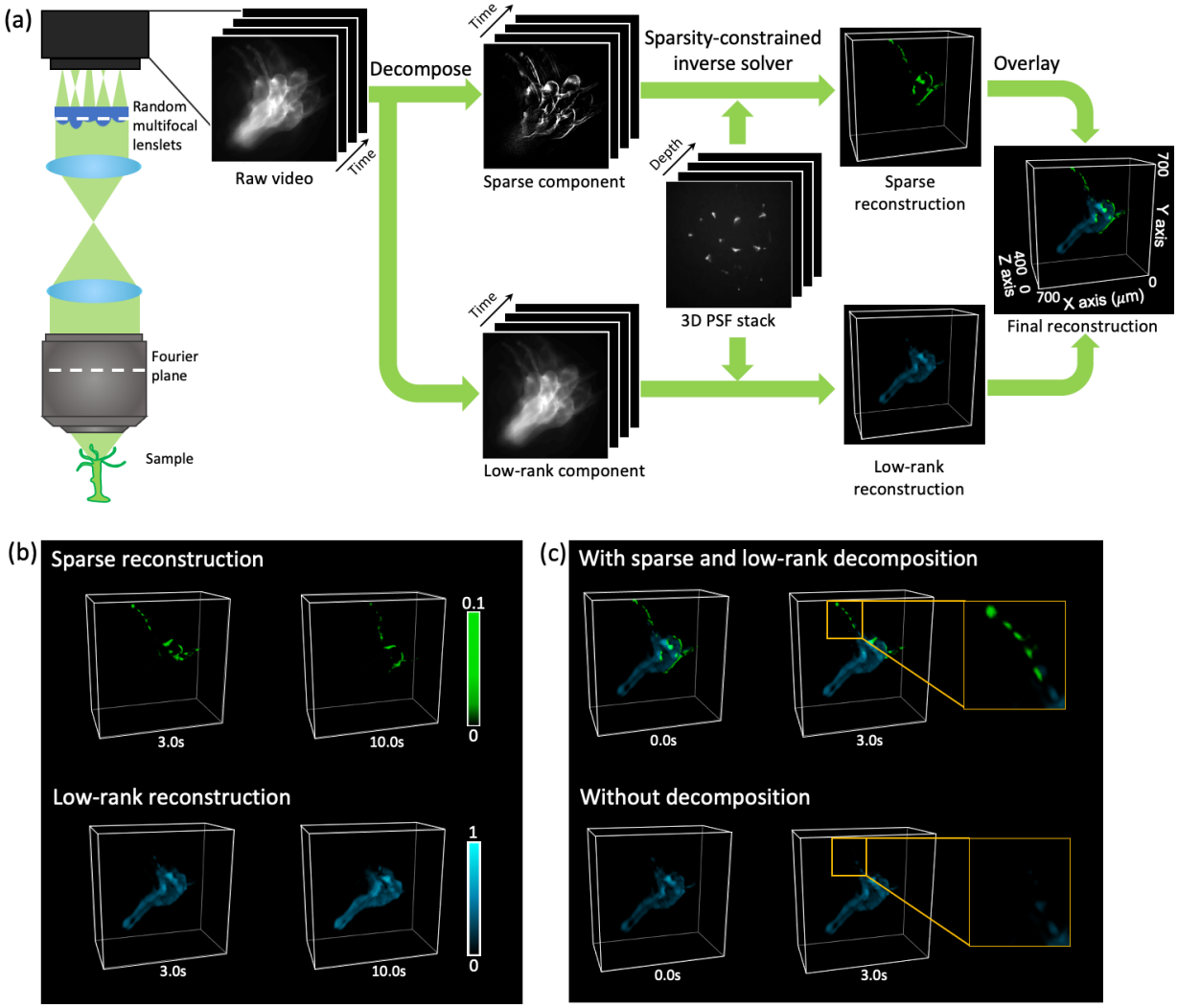


Figure 4.7: Sparse and low-rank video processing pipeline for Fourier DiffuserScope. (a) The optical system is a fluorescence microscope with a diffuser (random multifocal lenslets) in the relayed pupil plane. The video processing pipeline starts by decomposing the raw video into sparse and low-rank components, then reconstructs the corresponding 3D volumes separately by solving a sparsity-constrained inverse problem with knowledge of the system 3D PSF stack, and lastly overlays the two components to get the final reconstruction. (b) Selected frames from the reconstructed video with the sparse component (top) and the low-rank component (bottom) (c) Comparison between our new decomposition method (top) and the previous frame-by-frame method (bottom). The sparse reconstruction exhibits finer features including the tentacles.

Chapter 5

Optics and Algorithms Co-Design of Fourier DiffuserScope

5.1 Introduction

Many design methodologies have been used on single-shot 3D imaging systems [5, 19, 29, 32, 41, 43–57]. We categorize these methods into three group: heuristic designs based on first principles [5, 19, 29, 44, 45], merit-function-driven designs that optimize for a specific imaging task based on a system's point spread function's PSF properties [32, 40, 41], and end-to-end design methods that optimize directly for a given task through the use of a differentiable optical simulator in conjunction with a differentiable reconstruction algorithm [43, 46–59]. Physics-based end-to-end learning methods incorporate physical understanding about the system and therefore require fewer learnable parameters than deep neural network methods, enabling efficient training on smaller dataset as well as better generalizability. Our work in this chapter takes the last approach.

In this chapter, I (in collaboration with co-authors) demonstrate improved performance with an end-to-end optimized RML and reconstruction algorithm, using physics-based learning [57]. Our key contributions are as follows:

- Presented a physics-based learning framework to jointly optimize both the optical hardware (i.e., diffuser design) and the software (i.e., reconstruction algorithm).
- Demonstrated higher lateral and axial resolution over the imaging volume compared to heuristic designs.
- Implemented a memory-efficient backpropagation algorithm to fit a 3D volume of 6-million voxels and 10 reconstruction unrolls into a 24 gigabyte GPU.

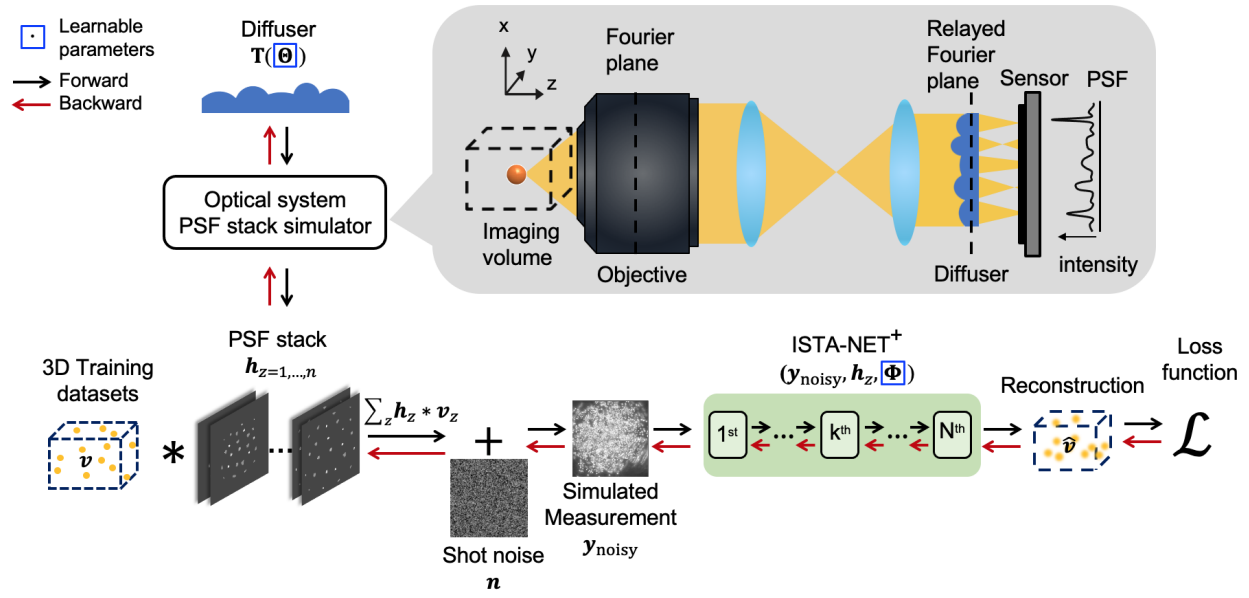


Figure 5.1: Overview of our physics-based learning pipeline. First, we simulate the point spread function (PSF) stack of the optical system with learnable diffuser parameters, $\boldsymbol{\Theta}$. The PSF stack is convolved with a 3D datasets, then a shot noise is added to produce a noisy, simulated measurement. The ISTA-Net$^+$ reconstruction network with learnable parameters, $\boldsymbol{\Phi}$, takes in the PSF stack and the noisy simulated measurement, and outputs a reconstructed volume which is fed into a loss function. The loss is backpropagated through the pipeline to update both the diffuser's and the reconstruction network's learnable parameters, $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$. The inset depicts our diffuser-based single-shot 3D microscopy setup. The optical system consists of a traditional fluorescence microscope with the addition of a relay lens to image the Fourier plane onto the diffuser, which is placed just before the sensor.

5.2 Physics-based Learning Pipeline

Our physics-based learning pipeline (Fig. 5.1) aims to jointly optimize the diffuser design (i.e., surface height of a refractive phase plate) and the reconstruction parameters. The optical simulator takes as input the current diffuser surface parameters, $\boldsymbol{\Theta}$, and outputs the microscope PSF corresponding to each depth plane. A noise-free 2D measurement is generated by laterally convolving each plane's PSF with each plane of the 3D training volume and then summing over depth. Noise is then added using the noise model in 5.2 and the noisy, simulated measurement is fed into an ISTA-Net$^+$ [114, 126] reconstruction network with learnable parameters, $\boldsymbol{\Phi}$. A loss function penalizes both ℓ_2 reconstruction error between the ground truth and reconstruction, and error in the learned inverse of the sparsifying transform. The optical system and reconstruction parameters ($\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$, respectively) are

jointly updated using gradient-based updates in a memory-efficient manner.

Diffuser Parameterization

Our RML diffuser (Fig. 5.1) consists of a number of plano-convex lenslets of varying focal lengths. The learnable parameters are each lenslet's lateral position coordinates and radius, $\mathbf{\Theta}_i = \{x_i, y_i, r_i\}$. Compared to a parameterization using a pixel-wise surface height [43, 51], our diffuser model dramatically reduces the number of learnable parameters, thus preventing overfitting and decreasing the amount of data required for training. Compared to a Gaussian diffuser [28], contour-shape diffuser [27] or a Zernike polynomial based phase mask [59], our lenslets-based surface focuses light to sharp points, providing higher signal-to-noise ratio (SNR) and frequency coverage over a wide depth range [29, 32, 44]. The final diffuser surface is a point-wise maximum over the individual spherical lens profiles with a maximum thickness of t (refer to Eq. 3.15 and Eq. 3.16):

Since we cannot take derivatives of the loss function with respect to the number of lenslets, we instead optimize the number of lenslets by starting with extra lenslets and allowing some to merge and exit the pupil during the learning process. In practice, the number of learned lenslets is consistent for each imaging scenario regardless of the initialized number, demonstrating robustness of the pipeline.

Imaging Model

In this section, we describe our imaging model to encode 3D information into a single 2D frame. Based on the Fourier DiffuserScope setup [44], we use a differentiable forward model followed by a differentiable noise model to generate simulated measurements, which are fed into Chap. 5.2 to reconstruct a 3D volume.

Optical System

The configuration of optical system is shown in the inset of Fig. 5.1. The system begins with a 20× 0.8 NA objective lens and a tube lens (with focal length $f_{TL} = 180$ mm), as in a traditional fluorescence microscope. Because the Fourier plane of the objective resides inside the objective tube and is inaccessible, we design a relay lens with $f_{RL} = 48$ mm to form a $4f$ system together with the tube lens and place the diffuser at the relayed Fourier plane. The sensor is 15.6 mm behind the diffuser. This setup results in an overall system magnification of 6.5×. The size of the imaging volume is $500 \times 500 \times 50\,\mu\text{m}^3$ consisting of 11 depth layers with an increment of 5 µm. We model our optics using wave-optics propagation in the following section.

Differentiable Forward Model

The wave-optical forward model is the same as the one described in Chap. 3.4. The 3D volume is divided into a stack of 2D slices of finite thickness in depth. Assuming our system

is laterally shift-invariant, we only need one PSF from each depth layer to fully characterize the system response. From an on-axis point source at each depth layer, we calculate the spherical wavefront at the objective back focal plane, and then multiply it by the apodization pupil function. The wavefront at the pupil is demagnified by the relay system and passes through the diffuser. The diffuser is modeled as a pure phase mask with phase delay $\phi = \exp\left[i\frac{2\pi}{\lambda}(n_r - 1)\mathbf{T}\right]$, with refractive index, n_r, and surface thickness, $\mathbf{T}(\mathbf{\Theta})$, embedding $\mathbf{\Theta}$ in the forward model. The electric field behind the diffuser is then digitally propagated to the sensor via angular spectrum method based on the Fresnel approximation [88]. The intensity images at the sensor from all the depth layers form a PSF stack, $\mathbf{h}_{z=1\ldots d}$, where $\mathbf{h}_z$ represents the simulated PSF at depth z and there are in total $d = 11$ depth layers in our training. The simulated measurement is modeled as the sum of all the lateral 2D convolutions of object slices and PSFs, one for each depth:

$$\mathbf{y} = \sum_z \mathbf{h}_z \overset{[x,y]}{*} \mathbf{v}_z = \mathbf{A}\mathbf{v}. \tag{5.1}$$

where, $\mathbf{y}$ is the noise-free intensity image, $\overset{[x,y]}{*}$ denotes the lateral 2D convolution, $\mathbf{v}_z$ is the object intensity at depth z. $\mathbf{v}$ represents the entire 3D volume and $\mathbf{A}$ is a matrix with columns containing the PSF stack, used to write our forward model in compact matrix form.

Noise Model

We expect to receive approximately 30k photons per fluorescent bead [43], allowing us to model the noise as a Poisson distribution and to neglect read noise. Unfortunately, the sampling of the Poisson distribution is not differentiable with respect to its input, complicating the computation of gradients with respect to the diffuser parameters. By the Central Limit Theorem, we can use the Gaussian approximation of shot noise (Eq. 5.2), giving a noise model that is differentiable with respect to the diffuser parameters. This noisy, simulated measurement $\mathbf{y}_{\text{noisy}}$ is then fed into our differentiable reconstruction algorithm:

$$\mathbf{y}_{\text{noisy}} = \mathbf{y} + \sqrt{\mathbf{y}} * \mathbf{n} \tag{5.2}$$

$$\mathbf{n} \sim \mathcal{N}(0, 1). \tag{5.3}$$

Physics-based Reconstruction

Our reconstruction algorithm aims to solve the following sparsity-constrained inverse problem:

$$\hat{\mathbf{v}} = \min_{\mathbf{v}} \|\mathbf{y} - \mathbf{A}\mathbf{v}\|_2^2 + \lambda \|G(\mathbf{v})\|_1, \tag{5.4}$$

where λ is a regularization parameter and $G(\cdot)$ is a transform that sparsifies the 3D volume. Traditional iterative optimization algorithms - for example, Fast Iterative Shrinkage-Threshold Algorithm (FISTA) [126] can be used to solve this problem, but suffer from slow

computation time (due to the large number of iterations required) and require hand-tuning of the hyper-parameters and proximal operators. Deep network based reconstruction methods are significantly faster than iterative optimization, but these black boxes lack an understanding of the physical system [30, 114]. Consequently, more training data is needed in order to achieve sufficient generalization at test time.

Here we apply a physics-based ISTA-Net$^+$ [114] of 10 unrolled ISTA iterations, each consisting of a gradient step followed by a proximal step, as shown in Fig. 5.2 (a). The learnable parameter set, $\mathbf{\Phi}$, includes the regularization parameter λ, the sparsifying transform $G(\cdot)$ and its left inverse $\tilde{G}(\cdot)$, as in $\mathbf{\Phi} = \{\lambda, G, \tilde{G}\}$. The loss function driving the training of the pipeline is:

$$\mathcal{L} = \frac{1}{M}\sum_{m=1}^{M} \|\hat{\mathbf{v}}_m^{(N)} - \mathbf{v}_m\|_2^2 + \gamma\|\tilde{G}(G(\mathbf{v}_m)) - \mathbf{v}_m\|_2^2, \tag{5.5}$$

where M is the number of training examples per batch, γ is a tuning parameter, subscript m denotes the m-th training example, superscript N denotes the total number of unrolls, and the second term encourages $\tilde{G}(\cdot)$ to be the left inverse of $G(\cdot)$.

A physics-based reconstruction approach is taken by including the forward model in the reconstruction algorithm to help minimize training data requirements and prevent overfitting. Additionally, by including the forward model in the reconstruction algorithm, the reconstruction takes in the PSF stack, which is a function of the diffuser's parameters $\mathbf{\Theta}$, allowing for the derivative of the loss to be taken with respect to them. We treat both update steps of this algorithm as forward Euler steps, allowing for inversion through backward Euler steps with a fixed point method. This enables the use of memory-efficient learning techniques (see Chap. 5.2) [56, 127, 128].

Memory-Efficient Backpropagation

Backpropagation is used to calculate the gradient of a single output with respect to multiple learnable variables of a system, with relatively low time complexity, by using a large amount of memory to store the full computational graph during the forward pass. Due to the 3D nature of our problem, GPU memory is at a premium. Therefore, we use memory-efficient backpropagation techniques to fit our problem in memory with a reasonable increase of compute time, allowing for calculating the necessary gradients from a series of unrolls while only having to fit a single unroll in memory. The primary memory-efficient backpropagation technique used is forward checkpointing based [129] but the model is also compatible with reverse checkpointing [128].

Forward checkpointing [129] consists of saving the output of each unroll of our network during the forward pass while deleting the computational graph. On the backward pass, we calculate the adjoint state by passing each checkpoint through its proceeding layer while saving the computational graph. Then auto-differentiation is used to take derivatives with

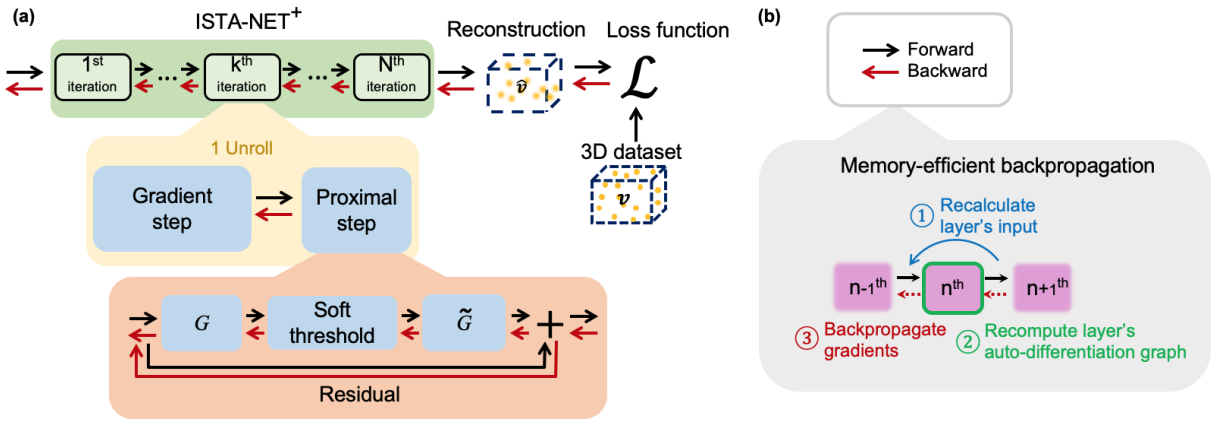


Figure 5.2: Physics-based reconstruction algorithm.**(a)** Our reconstruction network uses ISTA-Net$^+$ with N unrolled iterations of ISTA, each including a gradient step followed by a proximal step. The proximal step consists of a learned nonlinear sparsifying transform G, soft thresholding, and a learned left inverse of the sparsifying transform $\tilde{G}$. **(b)** Memory-efficient backpropagation calculates the n$^{\text{th}}$ layer's gradients in three steps: 1) recompute the layer's input from output. 2) recompute the layer's auto-differentiation graph. 3) recompute the gradients.

respect to the learnable parameters. For simple cases such as in Chap. 5.3, forward checkpointing works well when the checkpoint size and/or the number of checkpoints are small. When more unrolls are needed and checkpoints cannot be stored for each unroll, we implement a layer inversion framework [128] as shown in Fig. 5.2 (b). This framework requires each unroll to be invertible, then each layer's input can be recalculated by inverting its output, followed by recomputing the gradient of the loss function with auto-differentiation. By trading the time to calculate the inversion with the memory to store checkpoints, we can further reduce the memory requirements when necessary.

5.3 Learned Results

The size of the imaging volume is $500 \times 500 \times 50\,\mu\text{m}^3$ consisting of 11 depth layers in $5\,\mu\text{m}$ steps. The model was trained on a simulated dataset consisting of 200 volumes of $750 \times 750 \times 11$ voxels containing 1 and 2 micron beads. The beads were simulated using Gaussians with FWHM equal to the desired bead size. The distribution of the peak intensity of the beads is uniform from .8 to 1.2. The volumes are then multiplied by the desired photon emission level at training time to achieve the desired SNR. We also allow the density of the fluorescent beads to vary across training volumes. Testing was performed on 40 volumes generated in the same fashion as the training set. The model and training pipeline are written primarly in Pytorch and training was performed on a NVIDIA RTX 3090 GPU.

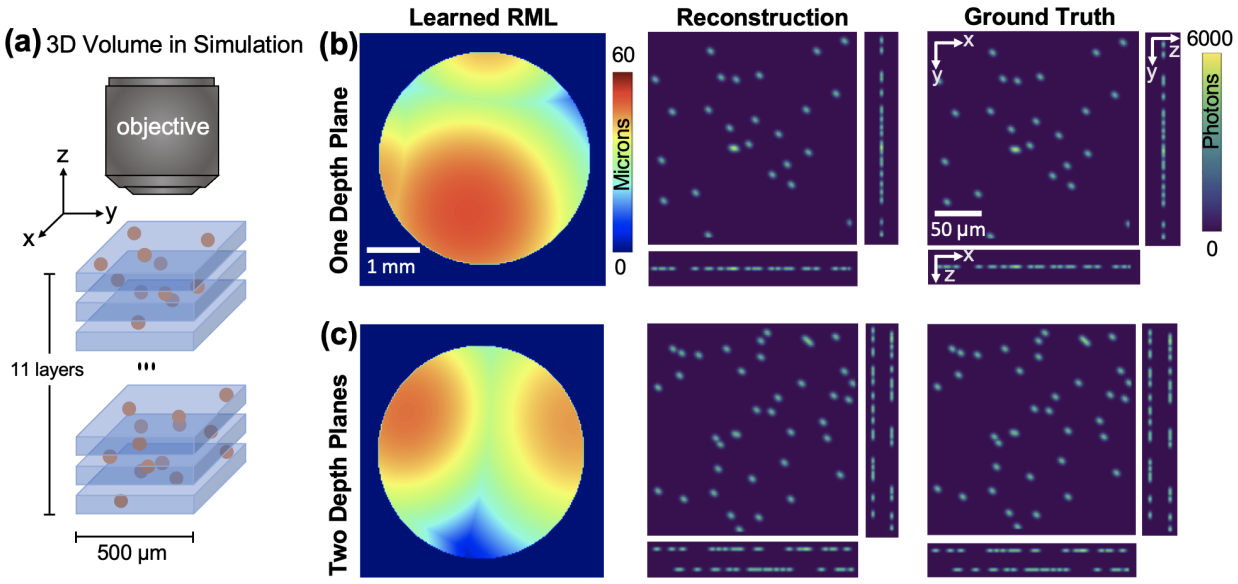


Figure 5.3: **(a)** Simulated 3D training volume with randomly placed fluorescent beads. **(b)** When the beads are constrained to a single depth plane, the learned random multi-focal lenslets (RML) consists of one dominating lens, as expected. **(c)** When the beads are constrained to two depth planes, the learned RML contains two lenses, each focusing at one depth plane. Maximum intensity projections show the reconstructed test volumes match well with ground truth.

Learned Diffuser for 2D and 2.5D Imaging

First, we first perform two 'sanity checks' to show that the pipeline functions as expected, by optimizing the RML for two imaging scenarios where a reasonable guess of the ideal RML is known. We use a small volume at 1/15 the size of our experimental system to speed up computation time and reduce our carbon footprint.

In the first scenario, all beads are constrained to the center depth plane of the volume. This simplifies the problem to 2D imaging and a single lenslet focusing at that depth should be ideal. As seen in Fig. 5.3 (b), a single lenslet dominates the RML but more than one lenslet is present due to the non-convexity of the problem. In the second scenario, the beads are constrained to 2 depth planes at opposite ends of the volume. In this scenario, a RML with two lenslets, each focusing at one of the depth planes, is expected. As seen in Fig. 5.3 (c), two lenslets dominate the RML. In both scenarios, the learned RML matches our intuition, and the reconstruction of the testing volume matches well with the ground truth. We conclude the pipeline is functioning sufficiently and move to more complex imaging scenarios.

Learned Diffuser for 3D Imaging

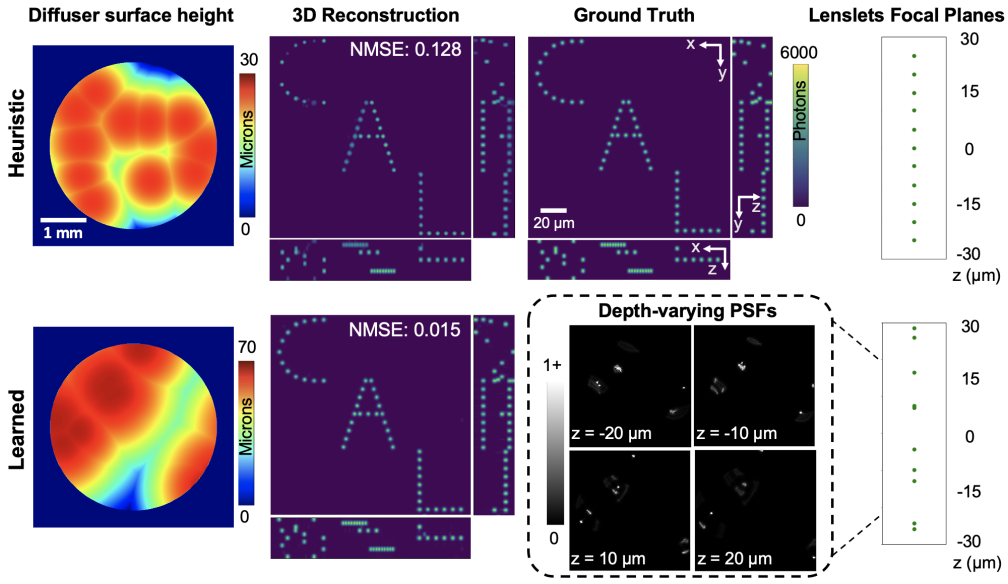


Figure 5.4: **a)** 3D volume in simulation under objective **b)** Initialization of random microlenslet (RML) surface height map. Note the green dots indicate the centroid of each active lenslet **c)** Initial focal distance of lenslets **d)** Maximum intensity projections (MIPs) of initial volume reconstruction **e)** Learned RML surface height map **f)** Learned focal distance distribution of active lenslets **g)** MIPs of reconstructed volume using learned RML and learned reconstruction algorithm

Next, we compare our learned RML and algorithm to a heuristically-designed RML with unlearned algorithms in Fig. 5.4. In our training data, we use beads of 1 µm and 2 µm diameter, randomly spaced in a 3D volume of $500 \times 500 \times 50\,\mu m^3$ with 5 µm axial steps. To achieve 1 µm lateral resolution over a 50 µm depth range, from first principles derivation [44], we need 3.2 lenslets in each direction, the square of which rounds up to 11 lenslets. Hence, the heuristic designed diffuser contains 11 randomly-located lenslets of varying focal length, focusing at uniformly-spaced depth planes. For the learned design, since we cannot take derivatives of the loss function with respect to the number of lenslets, we instead optimize the number of lenslets by starting with more than necessary and allowing them to merge and exit the pupil during the learning process. In experiments, the number of lenslets in the learned RML is consistent for each imaging scenario regardless of the initialized number, demonstrating robustness of the pipeline. The learned surface in Fig. 5.4 contains 8 lenslets; the lower number of lenslets increases the numerical aperture of each lenslet, collecting more photons per focus point and boosting SNR under our Gaussian-approximated shot noise model. For the case of the unlearned algorithm, we used an L1 proximal step with ISTA reconstructions ran until convergence. This runs ∼30x slower than our proposed

reconstruction network while achieving worse normalized mean square error (NMSE). Note that to reduce memory requirements and training time, the scale of this simulated system is 1/4 of our experimental system.

5.4 Conclusion

We demonstrated improved reconstruction speed and improved image reconstruction quality for Fourier DiffuserScope single-shot 3D microscopy by designing an optical element (a diffuser) that jointly optimizes the experimental setup and reconstruction algorithm via end-to-end learning. This data-driven approach directly optimizes the reconstruction loss and provides better insights into design for a non-traditional optical system where the first principles are limited.

Chapter 6

Conclusion

In this dissertation, I explored the design of Fourier DiffuserScope from optical hardware and algorithm perspectives, and demonstrated it is advantageous to jointly optimize the experimental setup and the reconstruction algorithm. This data-driven approach directly optimizes the reconstruction loss and provides better insights into the design of non-traditional optical systems where the first principles are limited.

In this chapter, I reflect on some challenges in the current system, provide possible solutions and propose directions for future works.

6.1 Challenges

There are still some challenges preventing the Fourier DiffuserScope from being more practical.

Scattering Effect

In the forward model, I didn't take into account tissue scattering which might impede the measurements. The scattering potential can be incorporated to enable deeper tissue imaging with higher-fidelity reconstructions, further suppressing noise and enhancing resolution. Scattering models including multi-slice [130, 131], SEAGLE [132], etc., can be combined with light field imaging [7, 133] as well as the diffuser measurements.

Sensor Model

The sensor settings will drastically affect the final measurement, but I didn't fully model the sensor, except for downsizing the intensity measurement to simulate the smear effect of the sensor pixels. However, phenomenons like the angular-dependent response, anti-aliasing filtering, and pixel non-uniformity are not considered. Moreover, if the output image is not in RAW format, the underlying image signal processor (ISP) pipeline is crucial to the final

image. The ISP pipeline for a certain camera varies case by case, but typically includes white balancing, denoising, and demosaicing for color sensors. An open-source ISP model, for example ISET [134], can be attached to the wave-optical module.

Generalizability of Priors

The outputs of sparsity-constrained algorithms vary with the choice of priors, which can be very subjective. Moreover, the hyperparameters controlling the weight of each regularizer are often hand-tuned. The criteria for a good set of parameters is based on past experience. Moreover, a carefully tuned algorithm might overfit a specific experimental system or a certain type of data, and cannot be generalized well to other models. Iterative algorithms with no tuning parameters or regularizers, e.g. the vanilla Richardson-Lucy algorithm, will provide a baseline and should be used to characterize the system metrics. As for pre-trained neural networks, we need to feed in with diversified training data, and stop early to prevent overfitting.

Memory Constraints

The raw diffuser measurements need reconstruction algorithms to reveal the object. Iterative algorithms cannot display the object in real time. In fact, for the video shown in Fig. 2.6, each frame took $\sim$ 25 minutes to reconstruct on a GPU. For a 25 fps video, that is 625 minutes of reconstruction time for every second of measurement. Pre-trained neural networks are able to give the reconstruction after 75ms latency for a small image patch (480×270 pixels) [30]. However, the size of the neural network scales with the size of the object. Due to memory limitations, with a single 12GB GPU, current applications are limited to 2D reconstruction with a small image patch. Using a single 24GB GPU, we are able to fit 10 unrolled ISTA to reconstruct a 3D volume containing $750 \times 750 \times 11$ voxels in [135]. The solutions to the memory limitations might lie in the following two categories, or a combination of both. (1) Benefit from the development of the semiconductor industry, GPUs, or cloud TPUs, will keep increasing their memories. So far, GPUs have reached 64 GB memory and cloud TPUs have reached 32 TiB memory, and yet more to come. (2) Use parallelization algorithms. For instance, JAX [136], a recent machine learning library developed by DeepMind, has a built-in function named *pmap* to execute codes in parallel across multiple GPU/TPU devices.

6.2 Future Work

Besides addressing those current challenges, there are quite a few new fields in which we can broaden the applications of diffuser imaging, and new techniques that we can utilize to empower the design of the next-generation diffuser imaging systems.

New Applications

Single-Shot Panoramic Imaging

A panorama is an image capturing a 360-degree view around the camera and is essential to robotic vision and autonomous driving. Current single-shot panorama systems use a camera ring or a wide-angle fisheye lens, which are bulky and expensive. Alternatively, diffusers are thin, flexible and can be easily bent. If using a single sensor, we can bend the diffuser sheet into a dome shape to capture light coming from all angles. If using a ring of cameras, we can roll the diffuser sheet into a cylinder circling the cameras. All in all, diffusers have the potential to provide a lightweight, small form-factor, and low-cost solution to panoramic imaging.

New Photodiode Sensors

The sensor technologies keep evolving. There are some new types of photodiode sensors available, expanding the boundary of computational imaging.

Sony Polarsens is a CMOS sensor that has on-chip polarizers with four different angles. Different from placing a polarizer in front of a traditional CMOS sensor, Polarsens has the polarizer grid between the on-chip lens and the photodiode, which not only accurately aligns the polarizers with the pixel but also yields a higher extinction ratio. Polarization is useful in material strain detection, screen defect detection, surface angle detection, etc. This polarization camera could be combined with diffuser for 3D polarization imaging.

In late 2021, Sony released its first high-speed high-resolution event-based sensors with the smallest pixel size of 4.86 µm. While traditional cameras capture complete scenes at each frame, the event-based cameras only react to the change of light intensity, reducing the data bandwidth and power consumption. The maximum frame rate can reach 1.06 Giga events per second. This type of sensor is especially suitable for dynamic scene imaging including neuron imaging and particle tracking, and is possible to be integrated into a light field imaging or a diffuser imaging system.

Thermal Imaging

Thermal detectors are a type of sensor detecting infrared light from black-body radiation and generating temperature information without contact. It is found useful in human detection, gas detection, failure analysis, etc. Different from photodiode-based cameras

that convert light directly into electrons through the photoelectric effect, thermal detectors instead first convert photons to heat using absorbers, and the heat is then converted into electrical signals. There are three types of thermal detectors: thermopiles detect the change in electromotive force, pyroelectric sensors detect the change in electrical polarization, bolometers detect the change in resistance. Although a single photodiode-based sensor has already reached 20-mega pixels, the number of pixels on the thermal detectors is quite limited. For example, the micro-bolometer sensor from *Teledyne FLIR* has up to 640×480 pixels; the thermopile array manufactured by *Heimann Sensor* has up to 120×84 pixels; the pyroelectric sensor has up to 4 detection zones. Since thermal sensors are often pixel-limited, we can use compressed sensing to enhance the sensor resolution. For example, we can place a random microlens diffuser made of Si or Ge in a thermal detection system and reconstruct a scene with more pixels than that on the sensor.

New Designs

Accurate and Lightweight Optical Simulation

From geometric optics to wave optics, there are different methods to simulate the PSFs of an optical system [137].

In this thesis, I used wave-optics to simulate the system PSFs in [44] and [135]. The wave-optical model propagates the complex electrical field from surface to surface until the sensor plane. It fully captures the phase information to simulate the PSFs with diffraction. However, it needs a high sampling rate to avoid aliasing, namely at least Nyquist sampling the smallest resolvable spot and zero padding the field. Furthermore, the whole sampling grid is applied to every surface, increasing the memory requirements.

The geometric PSF traces many rays and plots a spot diagram to show the geometric aberrations of the system. The rays don't contain any phase information, and thus there is no interference between those rays. This method is memory-efficient because as few as one ray is stored in the memory each time while computing multiple rays in parallel can speed up the simulation. Monte Carlo sampling can further speed up the procedure by reducing the number of rays traced. Unfortunately, geometric PSF is only accurate when the PSF is much larger than the diffraction limit, where geometric aberration is the main contribution to the PSF distribution. An end-to-end neural network with differentiable ray tracing has been implemented for lens design [138]. However, since the pure geometric ray tracing ignores diffraction, its application is limited to the macro photography scale and not suitable for microscopic imaging purposes.

In between the pure geometrical and the pure wave-optical PSFs, there exist another method termed *Huygens PSF*. The Huygens PSF treats rays as local plane waves containing both amplitude and phase. For each ray, the amplitude, coordinates, direction cosines, and optical path difference are logged. At the sensor plane, the algorithm sums up all the plane waves coherently. Similar to the wave-optical PSF, the Huygens PSF contains phase information and accounts for diffraction; similar to the geometric PSF, the Huygens PSF

can trace only one ray each time to save memory, and Monte Carlo sampling can be applied too. Moreover, the Huygens PSF is accurate in situations where the diffraction needs to be considered, or the pupil is hugely aberrated. Zemax OpticStudio contains a Huygens PSF function and can be integrated into Matlab/Python codes using the ZOS-API. It is also possible to implement a differentiable Huygens PSF to enable end-to-end learning.

Neural Network Architecture

Empowered by physical understandings of a optical system, emerging neural network architectures show better results in some imaging tasks. In Chap. 4.1 of this thesis, I showed an example of how using a simple 3-layer CNN with large kernels can provide a significantly better deconvolution result than a multi-layer UNet, by knowing that the PSF to be deconvolved has a large size. There are several other new neural network designs we may dig more into in the future. For instance, neural representations with coordinate-based multi-layer perceptrons (MLP), which map a coordinate to its corresponding value, have been demonstrated for 3D view synthesis [139] and optical refractive index tomography [140]. The image formation model of the coordinate-based MLP can be easily replaced to reconstruct 3D objects from other optical systems, say diffuser images. Furthermore, the concept of coordinate-based MLP is not limited to 2D or 3D space coordinates but can be generalized to represent time points, spectral wavelengths, polarization states, and so on. Another powerful tool for imaging-based neural networks is the transformer. Transformers are invented for natural language processing originally and are later proved to be useful in handling the coordinate relationships in image pixels. Positional encoding is a popular transformer that enables the representation of high-frequency details. Recent works show that Fourier features [141] and sinusoidal activation functions [142] can also broaden the frequency coverage of neural networks. These transformers can serve as an add-on module to existing neural network architectures.

Bibliography

[1] Gabriel Lippmann. "Épreuves réversibles donnant la sensation du relief". In: *Journal de Physique Théorique et Appliquée* 7.1 (1908), pp. 821–825.

[2] Fumio Okano, Haruo Hoshino, Jun Arai, and Ichiro Yuyama. "Real-time pickup method for a three-dimensional image based on integral photography". In: *Applied optics* 36.7 (1997), pp. 1598–1603.

[3] Ju-Seog Jang and Bahram Javidi. "Three-dimensional integral imaging of micro-objects". In: *Optics letters* 29.11 (2004), pp. 1230–1232.

[4] Bahram Javidi, Inkyu Moon, and Seokwon Yeom. "Three-dimensional identification of biological microorganism using integral imaging". In: *Optics express* 14.25 (2006), pp. 12096–12108.

[5] Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz. "Light field microscopy". In: *ACM Trans. Graph.* 25.3 (2006), pp. 924–934.

[6] Robert Prevedel, Young-Gyu Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrödel, Ramesh Raskar, Manuel Zimmer, Edward S Boyden, and Alipasha Vaziri. "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy". In: *Nature methods* 11.7 (2014), pp. 727–730.

[7] Nicolas C Pégard, Hsiou-Yuan Liu, Nick Antipa, Maximillian Gerlock, Hillel Adesnik, and Laura Waller. "Compressive light-field microscopy for 3D neural activity recording". In: *Optica* 3.5 (2016), pp. 517–524.

[8] Manuel Martínez-Corral and Bahram Javidi. "Fundamentals of 3D imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems". In: *Advances in Optics and Photonics* 10.3 (2018), pp. 512–566.

[9] Prashant Prabhat, Sripad Ram, E Sally Ward, and Raimund J Ober. "Simultaneous imaging of different focal planes in fluorescence microscopy for the study of cellular dynamics in three dimensions". In: *IEEE transactions on nanobioscience* 3.4 (2004), pp. 237–242.

[10] Paul M Blanchard and Alan H Greenaway. "Simultaneous multiplane imaging with a distorted diffraction grating". In: *Applied optics* 38.32 (1999), pp. 6692–6699.

[11] Sara Abrahamsson, Jiji Chen, Bassam Hajj, Sjoerd Stallinga, Alexander Y Katsov, Jan Wisniewski, Gaku Mizuguchi, Pierre Soule, Florian Mueller, Claire Dugast Darzacq, Xavier Darzacq, Carl Wu, Cornelia I Bargmann, David A Agard, Maxime Dahan, and Mats G L Gustafsson. "Fast multicolor 3D imaging using aberration-corrected multifocus microscopy". In: *Nature methods* 10.1 (2013), p. 60.

[12] Kuan He, Zihao Wang, Xiang Huang, Xiaolei Wang, Seunghwan Yoo, Pablo Ruiz, Itay Gdor, Alan Selewa, Nicola J Ferrier, Norbert Scherer, Mark Hereld, Aggelos K Katsaggelos, and Oliver Cossairt. "Computational multifocal microscopy". In: *Biomedical optics express* 9.12 (2018), pp. 6477–6496.

[13] Christian Maurer, Saranjam Khan, Stephanie Fassl, Stefan Bernet, and Monika Ritsch-Marte. "Depth of field multiplexing in microscopy". In: *Optics express* 18.3 (2010), pp. 3023–3034.

[14] Qi Guo, Zhujun Shi, Yao-Wei Huang, Emma Alexander, Cheng-Wei Qiu, Federico Capasso, and Todd Zickler. "Compact single-shot metalens depth sensors inspired by eyes of jumping spiders". In: *Proceedings of the National Academy of Sciences* 116.46 (2019), pp. 22959–22965.

[15] Yuan Luo, Se Baek Oh, and George Barbastathis. "Wavelength-coded multifocal microscopy". In: *Optics letters* 35.5 (2010), pp. 781–783.

[16] H Pin Kao and AS Verkman. "Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position". In: *Biophysical journal* 67.3 (1994), pp. 1291–1300.

[17] Bo Huang, Wenqin Wang, Mark Bates, and Xiaowei Zhuang. "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy". In: *Science* 319.5864 (2008), pp. 810–813.

[18] Adam Greengard, Yoav Y Schechner, and Rafael Piestun. "Depth from diffracted rotation". In: *Optics letters* 31.2 (2006), pp. 181–183.

[19] Sri Rama Prasanna Pavani, Michael A Thompson, Julie S Biteen, Samuel J Lord, Na Liu, Robert J Twieg, Rafael Piestun, and WE Moerner. "Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function". In: *Proceedings of the National Academy of Sciences* 106.9 (2009), pp. 2995–2999.

[20] René Berlich, Andreas Bräuer, and Sjoerd Stallinga. "Single shot three-dimensional imaging using an engineered point spread function". In: *Optics express* 24.6 (2016), pp. 5946–5960.

[21] Yoav Shechtman, Lucien E Weiss, Adam S Backer, Steffen J Sahl, and WE Moerner. "Precise three-dimensional scan-free multiple-particle tracking over large axial ranges with tetrapod point spread functions". In: *Nano letters* 15.6 (2015), pp. 4194–4199.

[22] Emmanuel J Candes and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215.

[23] M Salman Asif, Ali Ayremlou, Ashok Veeraraghavan, Richard Baraniuk, and Aswin Sankaranarayanan. "Flatcam: Replacing lenses with masks and computation". In: *2015 IEEE international conference on computer vision workshop (ICCVW)*. IEEE. 2015, pp. 663–666.

[24] Jesse K Adams, Vivek Boominathan, Benjamin W Avants, Daniel G Vercosa, Fan Ye, Richard G Baraniuk, Jacob T Robinson, and Ashok Veeraraghavan. "Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope". In: *Science advances* 3.12 (2017), e1701548.

[25] RH Dicke. "Scatter-hole cameras for x-rays and gamma rays". In: *The astrophysical journal* 153 (1968), p. L101.

[26] Edward E Fenimore and Thomas M Cannon. "Coded aperture imaging with uniformly redundant arrays". In: *Applied optics* 17.3 (1978), pp. 337–347.

[27] V Boominathan, J Adams, J Robinson, and A Veeraraghavan. "PhlatCam: Designed phase-mask based thin lensless camera". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[28] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. "DiffuserCam: lensless single-exposure 3D imaging". In: *Optica* 5.1 (2018), pp. 1–9.

[29] Grace Kuo, Fanglin Linda Liu, Irene Grossrubatscher, Ren Ng, and Laura Waller. "On-chip fluorescence microscopy with a random microlens diffuser". In: *Optics Express* 28.6 (2020), pp. 8384–8399.

[30] Kristina Monakhova, Joshua Yurtsever, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. "Learned reconstructions for practical mask-based lensless imaging". In: *Optics express* 27.20 (2019), pp. 28075–28090.

[31] Fanglin Linda Liu, Vaishnavi Madhavan, Nick Antipa, Grace Kuo, Saul Kato, and Laura Waller. "Single-shot 3D fluorescence microscopy with Fourier DiffuserCam". In: *Novel Techniques in Microscopy*. Optical Society of America. 2019, NS2B–3.

[32] Kyrollos Yanny, Nick Antipa, William Liberti, Sam Dehaeck, Kristina Monakhova, Fanglin Linda Liu, Konlin Shen, Ren Ng, and Laura Waller. "Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy". In: *Light: Science and Applications* 9.1 (2020). ISSN: 20477538.

[33] Nicholas Antipa, Sylvia Necula, Ren Ng, and Laura Waller. "Single-shot diffuser-encoded light field imaging". In: *2016 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2016, pp. 1–11.

[34] G Ajay Kumar, Jin Hee Lee, Jongrak Hwang, Jaehyeong Park, Sung Hoon Youn, and Soon Kwon. "LiDAR and Camera Fusion Approach for Object Distance Estimation in Self-Driving Vehicles". In: *Symmetry* 12.2 (2020). ISSN: 2073-8994.

[35] Clement Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. "Digging Into Self-Supervised Monocular Depth Estimation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2019.

[36] Nikolai Smolyanskiy, Alexey Kamenev, and Stan Birchfield. "On the Importance of Stereo for Accurate Depth Estimation: An Efficient Semi-Supervised Deep Neural Network Approach". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2018.

[37] Yan Wang, Wei-Lun Chao, Divyansh Garg, Bharath Hariharan, Mark Campbell, and Kilian Q. Weinberger. "Pseudo-LiDAR From Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.

[38] Vitor Guizilini, Rares Ambrus, Sudeep Pillai, Allan Raventos, and Adrien Gaidon. "3D Packing for Self-Supervised Monocular Depth Estimation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2020.

[39] Ruth R Sims, Sohaib Abdul Rehman, Martin O Lenz, Sarah I Benaissa, Ezra Bruggeman, Adam Clark, Edward W Sanders, Aleks Ponjavic, Leila Muresan, Steven F Lee, et al. "Single molecule light field microscopy". In: *Optica* 7.9 (2020), pp. 1065–1072.

[40] Raimund J Ober, Sripad Ram, and E Sally Ward. "Localization accuracy in single-molecule microscopy". In: *Biophysical journal* 86.2 (2004), pp. 1185–1200.

[41] Yoav Shechtman, Steffen J. Sahl, Adam S. Backer, and W. E. Moerner. "Optimal Point Spread Function Design for 3D Imaging". In: *Phys. Rev. Lett.* 113 (13 Sept. 2014), p. 133902.

[42] Nicholas Boyd, Eric Jonas, Hazen Babcock, and Benjamin Recht. "DeepLoco: Fast 3D localization microscopy using neural networks". In: *bioRxiv* (2018), pp. 1–16. ISSN: 2692-8205.

[43] Elias Nehme, Daniel Freedman, Racheli Gordon, Boris Ferdman, Lucien E. Weiss, Onit Alalouf, Tal Naor, Reut Orange, Tomer Michaeli, and Yoav Shechtman. "DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning". In: *Nature Methods* 17.7 (June 2020), pp. 734–740. ISSN: 15487105.

[44] Fanglin Linda Liu, K. U.O. Grace, Nick Antipa, Kyrollos Yanny, and Laura Waller. "Fourier diffuserscope: Single-shot 3d fourier light field microscopy with a diffuser". In: *Optics Express* (June 2020). ISSN: 23318422.

[45] M. A. A. Neil, R. Juškaitis, T. Wilson, Z. J. Laczik, and V. Sarafis. "Optimized pupil-plane filters for confocal microscope point-spread function engineering". In: *Optics Letters* 25.4 (2000), p. 245. ISSN: 0146-9592.

[46] Diptodip Deb, Zhenfei Jiao, Alex B. Chen, Misha B. Ahrens, Kaspar Podgorski, and Srinivas C. Turaga. "Programmable 3D snapshot microscopy with Fourier convolutional networks". In: i (2021), pp. 1–27.

[47] Harel Haim, Shay Elmalem, Raja Giryes, Alex M. Bronstein, and Emanuel Marom. "Depth Estimation From a Single Image Using Deep Learned Phase Coded Mask". In: *IEEE Transactions on Computational Imaging* 4.3 (June 2018), pp. 298–310. ISSN: 2573-0436.

[48] Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein. "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification". In: *Scientific Reports* 8.1 (Dec. 2018). ISSN: 20452322.

[49] Roarke Horstmeyer, Richard Y. Chen, Barbara Kappes, and Benjamin Judkewitz. "Convolutional neural networks that teach microscopes how to image". In: *arXiv* (Sept. 2017).

[50] Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. "Learning Rank-1 Diffractive Optics for Single-Shot High Dynamic Range Imaging". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2020, pp. 1383–1393.

[51] Christopher A. Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. "Deep Optics for Single-Shot High-Dynamic-Range Imaging". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2020), pp. 1372–1382. ISSN: 10636919.

[52] Ayan Chakrabarti. "Learning sensor multiplexing design through back-propagation". In: *Advances in Neural Information Processing Systems* Nips (2016), pp. 3089–3097. ISSN: 10495258.

[53] Julie Chang and Gordon Wetzstein. "Deep Optics for Monocular Depth Estimation and 3D Object Detection". In: *Proceedings of the IEEE International Conference on Computer Vision* 2019-Octob (Apr. 2019), pp. 10192–10201.

[54] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. "PhaseCam3D - Learning Phase Masks for Passive Single View Depth Estimation". In: *2019 IEEE International Conference on Computational Photography, ICCP 2019* (2019).

[55] Ilya Chugunov, Seung-Hwan Baek, Qiang Fu, Wolfgang Heidrich, and Felix Heide. "Mask-ToF: Learning Microlens Masks for Flying Pixel Correction in Time-of-Flight Imaging". In: (2021).

[56] Michael Kellman, Emrah Bostan, Michael Chen, and Laura Waller. "Data-Driven Design for Fourier Ptychographic Microscopy". In: *2019 IEEE International Conference on Computational Photography, ICCP 2019*. Institute of Electrical and Electronics Engineers Inc., May 2019. ISBN: 9781728132631.

[57] Michael R. Kellman, Emrah Bostan, Nicole Repina, and Laura Waller. "Physics-based learned design: Optimized coded-illumination for quantitative phase imaging". In: *IEEE Transactions on Computational Imaging* 5.3 (2018), pp. 344–353. ISSN: 23318422.

[58] Eran Hershko, Lucien E. Weiss, Tomer Michaeli, and Yoav Shechtman. "Multicolor localization microscopy by deep learning". In: *arXiv* 27.5 (2018), pp. 6158–6183.

[59] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging". In: *ACM Transactions on Graphics* 37.4 (2018). ISSN: 15577368.

[60] Adrian Stern and Bahram Javidi. "Three-dimensional image sensing and reconstruction with time-division multiplexed computational integral imaging". In: *Applied optics* 42.35 (2003), pp. 7036–7042.

[61] Michael Broxton, Logan Grosenick, Samuel Yang, Noy Cohen, Aaron Andalman, Karl Deisseroth, and Marc Levoy. "Wave optics theory and 3-D deconvolution for the light field microscope". In: *Optics express* 21.21 (2013), pp. 25418–25439.

[62] Noy Cohen, Samuel Yang, Aaron Andalman, Michael Broxton, Logan Grosenick, Karl Deisseroth, Mark Horowitz, and Marc Levoy. "Enhancing the performance of the light field microscope using wavefront coding". In: *Optics express* 22.20 (2014), pp. 24817–24839.

[63] Andrew Lumsdaine and Todor Georgiev. "The focused plenoptic camera". In: *2009 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2009, pp. 1–8.

[64] Haoyu Li, Changliang Guo, Deborah Kim-Holzapfel, Weiyi Li, Yelena Altshuller, Bryce Schroeder, Wenhao Liu, Yizhi Meng, Jarrod B French, Ken-Ichi Takamaru, Michael A Frohman, and Shu Jia. "Fast, volumetric live-cell imaging using high-resolution light-field microscopy". In: *Biomedical optics express* 10.1 (2019), pp. 29–49.

[65] Yanqin Chen, Bo Xiong, Yujia Xue, Xin Jin, Joseph Greene, and Lei Tian. "Design of a high-resolution light field miniscope for volumetric imaging in scattering tissue". In: *Biomedical Optics Express* 11.3 (2020), pp. 1662–1678.

[66] A Llavador, J Sola-Pikabea, G Saavedra, B Javidi, and M Martınez-Corral. "Resolution improvements in integral microscopy with Fourier plane recording". In: *Optics express* 24.18 (2016), pp. 20792–20798.

[67] G Scrofani, Jorge Sola-Pikabea, A Llavador, Emilio Sanchez-Ortiga, JC Barreiro, G Saavedra, J Garcia-Sucerquia, and Manuel Martınez-Corral. "FIMic: design for ultimate 3D-integral microscopy of in-vivo biological samples". In: *Biomedical optics express* 9.1 (2018), pp. 335–346.

[68] Changliang Guo, Wenhao Liu, Xuanwen Hua, Haoyu Li, and Shu Jia. "Fourier light-field microscopy". In: *Optics express* 27.18 (2019), pp. 25573–25594.

[69] Emilio Sanchez-Ortiga, Gabriele Scrofani, Genaro Saavedra, and Manuel Martınez-Corral. "Optical Sectioning Microscopy Through Single-Shot Lightfield Protocol". In: *IEEE Access* 8 (2020), pp. 14944–14952.

[70] Xing Lin, Jiamin Wu, Guoan Zheng, and Qionghai Dai. "Camera array based light field microscopy". In: *Biomedical optics express* 6.9 (2015), pp. 3179–3189.

[71] Ju-Seog Jang and Bahram Javidi. "Large depth-of-focus time-multiplexed three-dimensional integral imaging by use of lenslets with nonuniform focal lengths and aperturesizes". In: *Optics letters* 28.20 (2003), pp. 1924–1926.

[72] Christian Perwass and Lennart Wietzke. "Single lens 3D-camera with extended depth-of-field". In: *Human Vision and Electronic Imaging XVII*. Vol. 8291. International Society for Optics and Photonics. 2012, p. 829108.

[73] Todor Georgiev and Andrew Lumsdaine. "The multifocus plenoptic camera". In: *Digital Photography VIII*. Vol. 8299. International Society for Optics and Photonics. 2012, p. 829908.

[74] Lin Cong, Zeguan Wang, Yuming Chai, Wei Hang, Chunfeng Shang, Wenbin Yang, Lu Bai, Jiulin Du, Kai Wang, and Quan Wen. "Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (Danio rerio)". In: *Elife* 6 (2017), e28158.

[75] David L Donoho. "Compressed sensing". In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306.

[76] Emmanuel J Candès and Michael B Wakin. "An introduction to compressive sampling". In: *IEEE signal processing magazine* 25.2 (2008), pp. 21–30.

[77] Leonid I Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D: nonlinear phenomena* 60.1-4 (1992), pp. 259–268.

[78] Nikon Microscopy U. *Depth of Field and Depth of Focus*. www.microscopyu.com/microscopy-basics/depth-of-field-and-depth-of-focus.

[79] Thermo Fisher. *Fluoro-Max Dyed Green Aqueous Fluorescent Particles*. www.fishersci.com/shop/products/fluoro-max-dyed-green-aqueous-fluorescent-particles.

[80] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. "Distributed optimization and statistical learning via the alternating direction method of multipliers". In: *Foundations and Trends in Machine learning* 3.1 (2011), pp. 1–122.

[81] Adrian Negrean and Huibert D Mansvelder. "Optimal lens design and use in laser-scanning microscopy". In: *Biomedical optics express* 5.5 (2014), pp. 1588–1609.

[82] Thorlabs. *Scan Lenses for Laser Scanning Microscopy*. www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=2910.

[83] Edmund Optics. *Choosing the correct eyepiece.* www.edmundoptics.com/knowledge-center/application-notes/microscopy/choosing-the-correct-eyepiece.

[84] Bruce H Walker. *Optical design for visual systems.* Vol. 45. SPIE Press, 2000.

[85] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. "Light field photography with a hand-held plenoptic camera". PhD thesis. Stanford University, 2005.

[86] Tobias Nöbauer, Oliver Skocek, Alejandro J Pernıa-Andrade, Lukas Weilguny, Francisca Martinez Traub, Maxim I Molodtsov, and Alipasha Vaziri. "Video rate volumetric Ca 2+ imaging across cortex using seeded iterative demixing (SID) microscopy". In: *Nature methods* 14.8 (2017), pp. 811–818.

[87] Ori Katz, Pierre Heidmann, Mathias Fink, and Sylvain Gigan. "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations". In: *Nature photonics* 8.10 (2014), pp. 784–790.

[88] Joseph W Goodman. *Introduction to Fourier optics.* Roberts and Company Publishers, 2005.

[89] Jason Daniel Schmidt. "Numerical simulation of optical wave propagation: With examples in MATLAB". In: SPIE. 2010.

[90] David George Voelz. *Computational fourier optics: a MATLAB tutorial.* SPIE press Bellingham, Washington, 2011.

[91] Min Gu. *Advanced optical imaging theory.* Vol. 75. Springer Science & Business Media, 2000.

[92] Zemax. *Exploring Physical Optics Propagation (POP) in OpticStudio.* support.zemax.com/hc/en-us/articles/1500005488421-Exploring-Physical-Optics-Propagation-POP-in-OpticStudio.

[93] William Hadley Richardson. "Bayesian-based iterative method of image restoration". In: *JoSA* 62.1 (1972), pp. 55–59.

[94] Leon B Lucy. "An iterative technique for the rectification of observed distributions". In: *The astronomical journal* 79 (1974), p. 745.

[95] Manuel Guizar-Sicairos, Samuel T Thurman, and James R Fienup. "Efficient subpixel image registration algorithms". In: *Optics letters* 33.2 (2008), pp. 156–158.

[96] Nalux CO., LTD. *Random pitch micro lens array.* https://www.nalux.co.jp/en/pro_randompitch.html.

[97] Giuseppe A Cirino and Renato G Jasinevicius. "PMMA microlens array fabricated by indentation process". In: *Optical Components and Materials XIV.* Vol. 10100. SPIE. 2017, pp. 321–326.

[98] DL MacFarlane, V Narayan, JA Tatum, WR Cox, T Chen, and DJ Hayes. "Microjet fabrication of microlens arrays". In: *IEEE Photonics Technology Letters* 6.9 (1994), pp. 1112–1114.

[99] Norland Products. *Norland Adhesive*. http://www.norlandprod.com/adhchart.html.

[100] Simon Thiele, Kathrin Arzenbacher, Timo Gissibl, Harald Giessen, and Alois M Herkommer. "3D-printed eagle eye: Compound microlens system for foveated imaging". In: *Science advances* 3.2 (2017), e1602655.

[101] Aleksandr Ovsianikov and Boris N Chichkov. "Two-photon polymerization–high resolution 3D laser technology and its applications". In: *Nanoelectronics and Photonics*. Springer, 2008, pp. 427–446.

[102] Sam Dehaeck, B Scheid, and P Lambert. "Adaptive stitching for meso-scale printing with two-photon lithography". In: *Additive Manufacturing* 21 (2018), pp. 589–597.

[103] Sam Dehaeck. *TipSlicer*. github.com/SamDehaeck/TipSlicer.

[104] Kenneth R Castleman. *Digital image processing*. Prentice Hall Press, 1996.

[105] Daniel Sage, Lauréne Donati, Ferréol Soulez, Denis Fortun, Guillaume Schmit, Arne Seitz, Romain Guiet, Cédric Vonesch, and Michael Unser. "DeconvolutionLab2: An open-source software for deconvolution microscopy". In: *Methods* 115 (2017), pp. 28–41.

[106] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series, vol. 2*. 1949.

[107] Parthasarathy, Shreyas and Biscarrat, Camille. *Build your own DiffuserCam: Tutorial*. waller-lab.github.io/DiffuserCam/tutorial.html.

[108] Martın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems". In: *arXiv preprint arXiv:1603.04467* (2016).

[109] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. "Pytorch: An imperative style, high-performance deep learning library". In: *Advances in neural information processing systems* 32 (2019).

[110] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[111] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. "Deep convolutional neural network for inverse problems in imaging". In: *IEEE Transactions on Image Processing* 26.9 (2017), pp. 4509–4522.

[112] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012).

[113] Jian Sun, Huibin Li, Zongben Xu, et al. "Deep ADMM-Net for compressive sensing MRI". In: *Advances in neural information processing systems* 29 (2016).

[114] Jian Zhang and Bernard Ghanem. "ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2018), pp. 1828–1837. ISSN: 10636919.

[115] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[116] Oren Rippel, Jasper Snoek, and Ryan P Adams. "Spectral representations for convolutional neural networks". In: *Advances in neural information processing systems* 28 (2015).

[117] Harry Pratt, Bryan Williams, Frans Coenen, and Yalin Zheng. "Fcnn: Fourier convolutional neural networks". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2017, pp. 786–798.

[118] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Li Fei-Fei. "Novel Dataset for Fine-Grained Image Categorization". In: *First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*. Colorado Springs, CO, June 2011.

[119] Gaochang Wu, Mandan Zhao, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. "Light field reconstruction using deep convolutional network on EPI". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 6319–6327.

[120] Grace E Kuo. *Exploiting Randomness in Computational Cameras and Displays*. University of California, Berkeley, 2020.

[121] Nicolas Dey, Laure Blanc-Feraud, Christophe Zimmer, Pascal Roux, Zvi Kam, Jean-Christophe Olivo-Marin, and Josiane Zerubia. "Richardson–Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution". In: *Microscopy research and technique* 69.4 (2006), pp. 260–266.

[122] Eftychios A Pnevmatikakis, Daniel Soudry, Yuanjun Gao, Timothy A Machado, Josh Merel, David Pfau, Thomas Reardon, Yu Mu, Clay Lacefield, Weijian Yang, Misha Ahrens, Randy Bruno, Thomas M Jessell, Darcy S Peterka, Rafael Yuste, and Liam Paninski. "Simultaneous denoising, deconvolution, and demixing of calcium imaging data". In: *Neuron* 89.2 (2016), pp. 285–299.

[123] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. "Robust principal component analysis?" In: *Journal of the ACM (JACM)* 58.3 (2011), pp. 1–37.

[124] Young-Gyu Yoon, Zeguan Wang, Nikita Pak, Demian Park, Peilun Dai, Jeong Seuk Kang, Ho-Jun Suk, Panagiotis Symvoulidis, Burcu Guner-Ataman, Kai Wang, and Edward S. Boyden. "Sparse decomposition light-field microscopy for high speed imaging of neuronal activity". In: *Optica* 7 (Oct. 2020), pp. 1457–1468.

[125] Fabrice De Chaumont, Stéphane Dallongeville, Nicolas Chenouard, Nicolas Hervé, Sorin Pop, Thomas Provoost, Vannary Meas-Yedid, Praveen Pankajakshan, Timothée Lecomte, Yoann Le Montagner, et al. "Icy: an open bioimage informatics platform for extended reproducible research". In: *Nature methods* 9.7 (2012), pp. 690–696.

[126] Amir Beck and Marc Teboulle. "A Fast Iterative Shrinkage-Thresholding Algorithm". In: *Society for Industrial and Applied Mathematics Journal on Imaging Sciences* 2.1 (2009), pp. 183–202. ISSN: 1936-4954.

[127] Jens Behrmann, Will Grathwohl, Ricky T.Q. Chen, David Duvenaud, and Jörn Henrik Jacobsen. "Invertible residual networks". In: *36th International Conference on Machine Learning, ICML 2019*. Vol. 2019-June. 2019, pp. 894–910. ISBN: 9781510886988.

[128] Michael Kellman, Kevin Zhang, Eric Markley, Jon Tamir, Emrah Bostan, Michael Lustig, and Laura Waller. "Memory-Efficient Learning for Large-Scale Computational Imaging". In: *IEEE Transactions on Computational Imaging* 6 (Mar. 2020), pp. 1403–1414. ISSN: 23339403.

[129] Andreas Griewank and Andrea Walther. "Algorithm 799: Revolve: An Implementation of Checkpointing for the Reverse or Adjoint Mode of Computational Differentiation". In: *ACM Trans. Math. Softw.* 26.1 (Mar. 2000), pp. 19–45. ISSN: 0098-3500.

[130] Hsiou-Yuan Liu, Eric Jonas, Lei Tian, Jingshan Zhong, Benjamin Recht, and Laura Waller. "3D imaging in volumetric scattering media using phase-space measurements". In: *Optics express* 23.11 (2015), pp. 14461–14471.

[131] Michael Chen, David Ren, Hsiou-Yuan Liu, Shwetadwip Chowdhury, and Laura Waller. "Multi-layer Born multiple-scattering model for 3D phase microscopy". In: *Optica* 7.5 (2020), pp. 394–403.

[132] Hsiou-Yuan Liu, Dehong Liu, Hassan Mansour, Petros T Boufounos, Laura Waller, and Ulugbek S Kamilov. "SEAGLE: Sparsity-driven image reconstruction under multiple scattering". In: *IEEE Transactions on Computational Imaging* 4.1 (2017), pp. 73–86.

[133] Yi Zhang, Zhi Lu, Jiamin Wu, Xing Lin, Dong Jiang, Yeyi Cai, Jiachen Xie, Yuling Wang, Tianyi Zhu, Xiangyang Ji, et al. "Computational optical sectioning with an incoherent multiscale scattering model for light-field microscopy". In: *Nature communications* 12.1 (2021), pp. 1–11.

[134] Joyce E Farrell, Feng Xiao, Peter Bert Catrysse, and Brian A Wandell. "A simulation tool for evaluating digital camera image quality". In: *Image Quality and System Performance*. Vol. 5294. SPIE. 2003, pp. 124–131.

[135] Eric Markley*, Fanglin Linda Liu*, Michael Kellman, Nick Antipa, and Laura Waller. "Physics-Based Learned Diffuser for Single-shot 3D Imaging". In: *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*. 2021.

[136] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. *JAX: composable transformations of Python+NumPy programs*. Version 0.2.5. 2018. URL: http://github.com/google/jax.

[137] Zemax. *What is a Point Spread Function?* https://support.zemax.com/hc/en-us/articles/1500005576562.

[138] Qilin Sun, Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. "End-to-end complex lens design with differentiable ray tracing". In: *ACM Trans. Graph.* 40.4 (2021), pp. 1–13.

[139] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. "Nerf: Representing scenes as neural radiance fields for view synthesis". In: *European conference on computer vision*. Springer. 2020, pp. 405–421.

[140] Renhao Liu, Yu Sun, Jiabei Zhu, Lei Tian, and Ulugbek Kamilov. "Zero-Shot Learning of Continuous 3D Refractive Index Maps from Discrete Intensity-Only Measurements". In: *arXiv preprint arXiv:2112.00002* (2021).

[141] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. "Fourier features let networks learn high frequency functions in low dimensional domains". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 7537–7547.

[142] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. "Implicit neural representations with periodic activation functions". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 7462–7473.