Optics and Algorithms for Designing Miniature Computational Cameras and Microscopes

by

Kyrollos Yanny

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Laura Waller, Chair
Professor Bo Huang
Professor Chunlei Liu

Summer 2022

Optics and Algorithms for Designing Miniature Computational Cameras and Microscopes

Abstract

Optics and Algorithms for Designing Miniature Computational Cameras and Microscopes

by

Kyrollos Yanny

Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor Laura Waller, Chair

Traditional cameras and microscopes are often optimized to produce sharp 2D images of the object. These 2D images miss important information about the world (e.g. depth and spectrum). Access to this information can make a significant impact on fields such as neuroscience, medicine, and robotics. For example, volumetric neural imaging in freely moving animals requires compact head-mountable 3D microscopes and tumor classification in tissue benefits from access to spectral information. Modifications that enable capturing these extra dimensions often result in bulky, expensive, and complex imaging setups. In this dissertation, I focus on designing compact single-shot computational imaging systems that can capture high dimensional information (depth and spectrum) about the world. This is achieved by using a multiplexing optic as the image capture hardware and formulating image recovery as a convex optimization problem. First, I discuss designing a single-shot compact miniature 3D fluorescence microscope, termed Miniscope3D. By placing an optimized multifocal phase mask at the objective's exit pupil, 3D fluorescence intensity is encoded into a single 2D measurement and the 3D volume can be recovered by solving a sparsity constrained inverse problem. This enables a 2.76 μm lateral and 15 μm axial resolution across $900 \times 700 \times 390$ μm^3 volume at 40 volumes per second in a device smaller than a U.S. quarter. Second, I discuss designing a single-shot hyperspectral camera, termed Spectral DiffuserCam, by combining a diffuser with a tiled spectral filter array. This enables recovering a hyperspectral volume with higher spatial resolution than the spectral filter alone. The system is compact, flexible, and can be designed with contiguous or non-contiguous spectral filters tailored to a given application. Finally, the iterative reconstruction methods generally used for compressed sensing take thousands of iterations to converge and rely on hand-tuned priors. I discuss a deep learning architecture, termed MultiWienerNet, that uses multiple differentiable Wiener filters paired with a convolutional neural network to take into account the system's spatially-varying point spread functions. The result is a $625 - 1600\times$ increase in speed compared to iterative methods with spatially-varying models and better reconstruction quality than deep learning methods that assume shift invariance.

To my family

This would not be possible without you

Contents

List of Figures

List of Tables

Acknowledgments

I would like to start by thanking my Dad, Nashaat Henry Yanny. He was the biggest believer in me and looked forward to this dissertation more than I ever did. He insisted that our family immigrates from Egypt to the United States and did everything in his powers and beyond to support me during my graduate studies. Unfortunately, he passed away during my fourth year of graduate school. His memory, loving nature, and smile have been the largest driving force pushing me to continue this long journey. I would like to also thank my Mom, Salwa Rizek. Throughout my childhood, she planted the importance of education, hard work, family, and honesty in me. My brother, Mina, and my sister, Marina, have always been there for me and made the hard times a bit easier, a bit harder, and a bit easier again!

This journey started at East Los Angeles Community College (ELAC). I would like to thank the many friends and teachers that helped me there. Especially, I would like to thank Natalie Cobian and her family, and Manolo Orta. I also want to thank the MESA center at ELAC for their excellent counseling, tutoring, and for helping me land my first research opportunity at UCLA.

After ELAC, I transferred to UCLA. There, I got involved in optics research for the first time. I want to thank Professor Jacob Schmidt and Professor Aydogan Ozcan for their mentorship and support while I was doing research in their labs. I also want to thank Zoltan Gorocs for being the best post-doc mentor I ever had! We spent many nights doing research together and building microscopes and submarines! In my senior year at UCLA, I had no intentions of pursuing graduate school. However, the UCLA CARE fellows program and Prof. Tama Hasson had other plans for me. They guided me through the GREs, twice, the graduate school applications, and helped me, multiple times, in drafting my NSF Graduate Research Fellowship proposal that eventually got me funding for my PhD. I cannot thank them enough.

When I joined UC Berkeley, I was very unsure which lab to join for my PhD, I am extremely lucky and fortunate to have chosen the Waller Lab and even luckier for Laura to have chosen me as well. So I want to thank first and foremost my advisor, Laura Waller. Her guidance, mentorship, kindness, and relentless belief in my research projects were absolutely crucial in making this happen. I also would like to thank ALL my labmates whose academic advice, friendship, gossip, and social activities made this long journey very enjoyable. In particular, I want to thank Kristina Monakhova , who co-authored most of my papers, Nick Antipa, who guided me when I first joined the lab, Grace Kuo and Linda Liu, for their advice, friendship, and their many inventive ways to make diffusers. At Berkeley, I was lucky enough to have a supportive cohort and friend group. I thank them ALL for their help and support especially, Kristen, Andoni, Marc, Konlin, and all the members of the Drop-Out Squad whether honorary or actual.

I also would like to than my collaborators whose contributions were essential for the success of the projects. In particular, Willam Liberti for his constant belief, contribution, and support of the Miniscope3D, Sam Dehaeck, although we never met in person, your Nanoscribe algorithm was essential. This was an excellent example of online collaboration.

Richard, for his tireless work on the MultiWienerNet, and the many undergraduate students that helped and contributed to the projects, published and unpublished. Finally, I would like to thank all my family members and friends back in Egypt. Your love and support have always been there for me.

This work was supported, in part, by the Defense Advanced Research Projects Agency (DARPA), Contract No. N66001-17-C-4015, Gordon and Betty Moore Foundation Data Driven Discovery Initiative (Grant GBMF4562), National Institute of Health (NIH) grant 1R21EY027597-01, the National Science Foundation Grant No. 1617794, and an Alfred P. Sloan Foundation fellowship.

Chapter 1

Introduction

1.1 Dissertation Outline

In this dissertation, I demonstrate how to design and optimize single-shot volumetric and hyperspectral computational imaging systems using multiplexing optics, image recovery algorithms, and deep learning.

Chapter 1: This chapter provides the necessary background that is used throughout the rest of the dissertation. It starts by covering how optical systems form an image using linear image formation models. Then I discuss how using a multiplexing optic can help with capturing more information in a more compact device but introduces image recovery challenges. To alleviate these challenges, the multiplexing optic will need to have certain properties (e.g. randomness, incoherence) for successful image recovery. Image recovery can be performed using iterative optimization algorithms or deep learning methods which are both introduced here. Finally, previous work related to designing volumetric or hyperspectral single-shot imaging systems is discussed.

Chapter 2: This chapter demonstrates how an optimized multiplexing optic (e.g. multifocal microlens array) can replace the tube lens in a miniature fluorescence microscope to enable volumetric single-shot imaging in a device smaller than a U.S. quarter. I introduce theory and algorithms to design, optimize, and fabricate the multiplexing optic as well as a low-rank image formation model that takes into account the field-varying aberrations inherent to miniature optics. The device is experimentally validated on freely swimming fluorescent tardigrades and mouse brain tissue slices. I also discuss SNR comparisons with 2D miniature microscopes, how reconstruction quality varies with object sparsity, and provide a guide to adapt the theory to different optical systems.

Chapter 3: This chapter demonstrates another application of the co-design of a multiplexing optic with post-process algorithms. Combining an off-the-shelf diffuser with a tiled spectral filter array enables recovering a hyperspectral volume from a single-shot. The device is compact and offers higher spatial resolution than that achieved by using the spectral filter array alone. I present theory for quantifying lateral and spectral resolution as well as experimental results showing hyperspectral reconstructions with high spatio-spectral resolution.

Chapter 4: In the previous chapters, an iterative optimization algorithm is used to recover a high-dimensional object (e.g. volumetric or hyperspectral) from a 2D image. These iterative algorithms take thousands of iterations to converge and rely on hand-tuned priors to achieve good reconstruction quality. This prohibits real-time processing of images and limits the type of objects that can be used. Deep learning methods can speed-up the reconstruction process, however, they often assume shift-invariance. This assumption is not valid in miniature optical systems, where the strict size requirements usually prevent use of aberration-correcting optics. This chapter introduces a deep learning architecture that uses multiple differentiable Wiener filters with a convolutional neural network to take into account the field-varying behaviour of the optical system. This results in a $625 - 1600\times$ increase in speed compared to iterative methods with spatially-varying models and better reconstruction quality than deep learning methods that assume shift-invariance.

1.2 Image Formation Models

Traditional cameras and microscopes are often optimized to produce sharp 2D images of the object. These 2D images miss important information about the world (e.g. depth and spectrum). Access to this information can make a significant impact on fields such as neuroscience, medicine, and robotics. For example, volumetric neural imaging in freely moving animals requires compact head-mountable 3D microscopes and tumor classification in tissue benefits from access to spectral information. Modifications that enable capturing these extra dimensions often result in bulky, expensive, and complex imaging setups. In this dissertation, I focus on designing compact single-shot computational imaging systems that can capture high-dimensional information (depth and spectrum) about the world. This is achieved by using a multiplexing optic as the image capture hardware and formulating image recovery as a convex optimization problem. To understand how changing the optics affect the measurement and the image recovery, an accurate image formation model (forward model) is required. To establish this forward model, the volumetric object intensity is treated as a 3D grid of voxels, $\mathbf{v}[x, y, z]$. Each voxel produces a point spread function (PSF), $\mathbf{h}[x', y'; x, y, z]$, on the camera sensor, where $[x', y']$ are image space indices. Since the object voxels are mutually incoherent, the measurement can be expressed as a linear combination of the PSFs from each voxel in the object:

$$\begin{aligned}\mathbf{b}[x',y'] &= \sum_{z}\sum_{x,y}\mathbf{v}[x,y,z]\mathbf{h}[x',y';x,y,z] \\ &= \mathbf{A}\mathbf{v},\end{aligned} \tag{1.1}$$

where **b** is the measurement and **A** is a matrix that maps the 3D volume to the 2D measurement.

Shift-invariant model

If the optical system is well corrected for aberrations, as is the case for complex bench-top optical systems, the PSF will approximately be the same for all object points within the field of view (FOV). Such a system is said to be shift-invariant. This property significantly reduces the complexity of Eq. 1.1 as it reduces to a sum over 2D convolutions:

$$\mathbf{b}[x',y'] = \sum_{z}\sum_{x,y}\mathbf{v}[x,y,z] \overset{[x,y]}{*} \mathbf{h}[x,y,z], \tag{1.2}$$

where $\overset{[x,y]}{*}$ represents a 2D convolution. This shift-invariant model relates changes to the optical system, as described by the PSF, to changes in the measurement. This allows for a direct way to design PSFs that introduce desirable measurement properties. A modified version of this model is used in *Chapter 3* to describe how using a diffuser and a tiled spectral filter array affects the measurement. Since the diffuser can be approximated by randomly-spaced microlenses with a small aperture (~f/20), off-axis aberrations are negligible and the system is approximately shift-invariant.

Shift-varying model

While the shift-invariant model is applicable to many optical systems, miniature high NA objectives have significant off-axis aberrations that cannot be ignored. In *Chapter 2*, the 0.55 NA GRIN objective has significant coma that severely impacts the quality of reconstructions when the shift-invariant model is used (see Fig. 2.4 (c)). Thus, the field-varying behavior of the PSF needs to be included in the forward model. There are multiple ways to approximate the image formation model for shift-invariant PSFs (e.g. locally convolutional, low-rank models, etc. [5, 73, 13, 28, 109, 32, 59]). Any of these techniques is applicable to our pipeline. In particular, I introduce a low-rank forward model (discussed in more detail in *Chapter 2*), approximating the spatially-varying PSFs as a weighted sum of shift-invariant kernels:

$$\mathbf{b}[x',y'] = \sum_{z}\sum_{r=1}^{K}\left\{(\mathbf{v}[x,y,z]\mathbf{w}_r[x,y,z]) \overset{[x,y]}{*} \mathbf{g}_r[x,y,z]\right\}[x',y'], \tag{1.3}$$

where the weights $\{\mathbf{w}_r\}$ and the kernels $\{\mathbf{g}_r\}$ are computed from a singular value decomposition (SVD) of sparsely sampled PSFs from different positions in the FoV and the inner sum is over the K largest values in the SVD. Note that in the 2D imaging case, the object is a thin slice in the z-dimension, so the outer sum over z is not included in the forward model.

1.3 Multiplexing Optics and Compressed Sensing

The problem of recovering a high-dimensional (e.g. 3D) signal from a low-dimensional measurement (e.g. 2D image) can be described using a linear systems of equations as shown in Eq. 1.1, where $\mathbf{A}$ is a wide matrix forming an underdetermined linear system with fewer equations than unknowns. Generally, such a system has an infinitude of solutions. This means that there are infinitely many objects, $\mathbf{v}[x, y, z]$, that can describe the measurement $\mathbf{b}$. This makes the recovery of the object infeasible with traditional linear solvers. However, if certain conditions are imposed on the object and on the sensing matrix, *compressed sensing* theory [19] can be used to recover the object.

Object Sparsity and Matrix Coherence

Compressed sensing theory deals with recovering an object, $\mathbf{v}[x, y, z]$ that has more elements than the measurement, $\mathbf{b}$, given that the object is sparse. This means that the object has very few non-zero elements. For example, the image of stars in a dark night or a fluorescent object on a black background are considered to be natively sparse. The intuition here is that instead of recovering the full object (which has more elements than the measurement), one only has to recover a few non-zero coefficients and their location while the rest of the elements are zero. This makes for an easier optimization problem. While the requirement of object sparsity can seem to be very restrictive, compressed sensing theory requires the object to be sparse in some domain. This means that the object does not have to be natively sparse but that there is some representation of the object that has many zeros in its coefficients. Many such representations exist (e.g. wavelets, DCT). In this dissertation, I use *Total Variation* as the sparsifying transform. This requires the object to have sparse gradients which is valid for a general class of objects.

In addition to requiring the object to be sparse, compressed sensing theory requires the sensing matrix $\mathbf{A}$ to be incoherent. *Mutual coherence* is defined as the largest normalized inner product between two distinct columns of the matrix:

$$\mu(\mathbf{A}) = \max_{i \neq j} |\langle \frac{a_i}{||a_i||_2}, \frac{a_j}{||a_j||_2} \rangle|. \tag{1.4}$$

This quantity captures how close matrix $\mathbf{A}$ is to an orthogonal matrix ($\mu(\mathbf{A}) = 0$). A "good" sensing matrix will have a small $\mu(\mathbf{A})$. Intuitively speaking, since the measurement is a linear combination of the columns of matrix $\mathbf{A}$. It should be easier to determine which columns participate in the measurement if the columns are very distinct from one another.

In fact, compressed sensing guarantees the recovery of the object under the following condition:

$$\mu(\mathbf{A}) \leq \frac{1}{2||\mathbf{v}||_0} \tag{1.5}$$

where $\|\cdot\|_0$ is the l^0 norm, the number of non-zero elements in a vector. While this condition is not met in practice, the intuition behind it is very useful. As the matrix coherence decreases, a more dense object can be recovered. This promotes designing optical systems whose sensing matrices are as incoherent as possible.

A traditional lens produces axial PSFs that are not suitable for compressed sensing. This is because the sensing matrix produced by such PSFs is very coherent (i.e. high $\mu(\mathbf{A})$). In contrast, a diffuser produces a PSF consisting of a unique pseudorandom pattern with high-frequency features. This forms a much more incoherent matrix than a traditional lens. *Chapter 3* uses a diffuser as the imaging optic to multiplex spectral information into a 2D measurement. Combining the diffuser with a tiled spectral filter array allows for recovering 64 spectral channels with high spatial resolution from a single 2D measurement. While the diffuser is a great optic for matrix incoherence, it results in a measurement with less frequency content and reduced SNR (as light in the dark areas of the PSF is not completely zero). For applications that are signal starved and require high spatial resolution (e.g. fluorescence imaging), a multifocal, randomly spaced microlens array, produces PSFs with more high-frequency content and better light concentration (see Fig. 1.1). *Chapter 2* uses a multifocal microlens array to multiplex volumetric information into a 2D measurement allowing for single-shot 3D imaging. In addition, *Chapter 2* builds on the mutual coherence metric and introduces a merit function that introduces astigmatism and optimizes the microlenses' positions to produce axial PSFs that are very distinct from each other (i.e. incoherent).

Image Recovery

Provided the object is sparse in some domain, we can reconstruct the volume by solving the following sparsity-constrained inverse problem:

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}\geq 0} \|A\mathbf{v} - \mathbf{b}\|_2^2 + \tau\|\Psi\mathbf{v}\|_1, \tag{1.6}$$

with Ψ being a sparsifying transform (e.g. 3D gradient, corresponding to TV regularization) and τ being a tuning parameter.

Equation 1.6 can be solved using a variety of iterative methods; we use Fast Iterative Shrinkage Thresholding (FISTA) [11]. This requires repeatedly applying A and its adjoint. To make this computationally feasible for high megavoxel systems like ours, we need an efficient representation for A. A shift-invariant forward model is extremely computationally efficient because A becomes a convolution matrix [3, 4, 52]. It also requires only a single PSF calibration image, from which the PSFs at all other positions can be inferred. For systems that are shift-varying (e.g. miniature microscopes), the low-rank forward model provides an efficient representation (see *Chapter 2* for more detail).

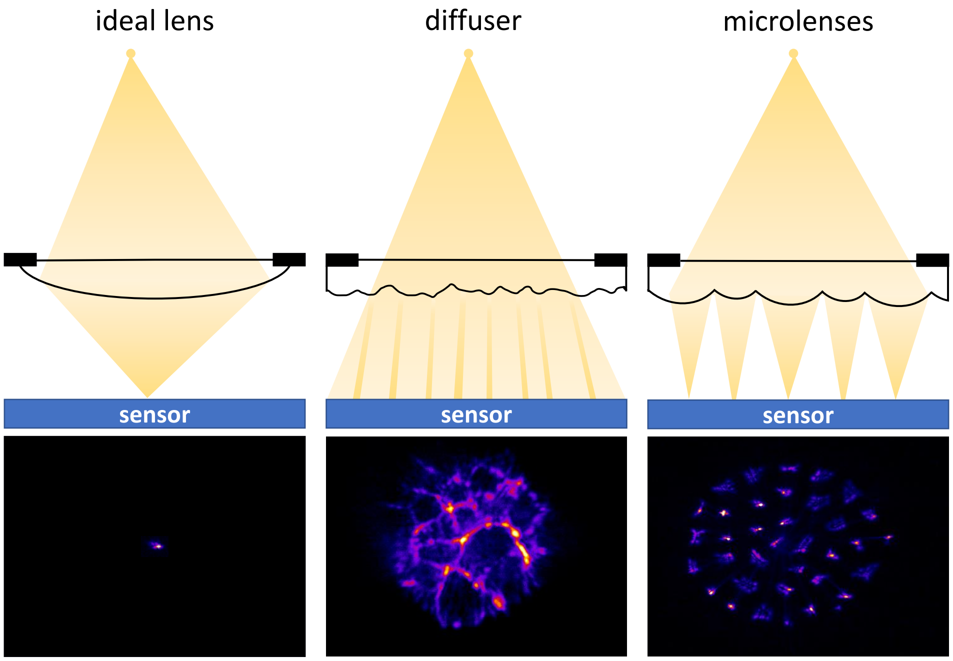

Figure 1.1: **Multiplexing vs one-to-one optics.** (left) an ideal lens focuses a point source to a point. This is ideal in terms of high frequency content and measurement SNR. However, axial PFSs from an ideal lens produce a coherent sensing matrix that is incompatible with compressed sensing based recovery. (middle) a diffuser projects a point to a pseudorandom pattern with some high frequency content. Axial PSFs from a diffuser produce a much more incoherent sensing matrix at the cost of lower measurement SNR and less high-frequency content. (right) multifocal randomly-spaced microlenses project a point to a pattern with high frequency content and higher measurement SNR than the diffuser. Axial PSFs from the microlens array can also produce an incoherent sensing matrix.

1.4 Iterative and Deep Methods to Solving the Inverse Problem

A variety of algorithms have been utilized to solve Eq. 1.6 over the years. Classical methods use iterative optimization approaches, such as the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) and the Alternating Direction Method of Multipliers (ADMM) [11, 16]. Such methods incorporate hand-picked priors, such as Total Variation (TV) and native sparsity, to improve image quality. In *Chapter 2* & *Chapter 3*, I use FISTA with a TV prior to solve the inverse problem. A typical reconstruction of size $512 \times 512 \times 20$ takes 1-3k iterations, and runs in 8-24 minutes on a GPU RTX 2080-Ti using MATLAB. While classical methods provide great flexibility in incorporating different forward models and trying out different priors for a given measurement, these methods are prohibitively slow to run in real-time.

Deep learning approaches can be used to speed up the image reconstruction [76, 98, 85, 55]. These methods provide good image quality and can run in real-time. However, they rely on a shift-invariant PSF approximation and do not generalize well to optical systems with field-varying aberrations. In *Chapter 4*, I introduce a deep learning architecture for fast, spatially-varying deconvolution. The architecture, termed MultiWienerNet, consists of multiple Wiener deconvolution layers followed by a convolutional neural network (CNN). Each Wiener deconvolution layer is initialized with a different PSF sampled from a different object point in the FoV. The result of the multiple Wiener deconvolution layers is a set of intermediate images that have sharp features in different regions depending on where the PSF is sampled from. These intermediate images are then fed to a refinement CNN to blend them together and produce the final output. The learnable Wiener deconvolution filters are initialized with PSFs captured at several locations in the FoV, but then allowed to update throughout training to learn the best filters and noise regularization parameters. This allows us to incorporate knowledge of the field-varying aberrations into the network, providing a physically-informed initialization that is further refined throughout training. The end result is a fast spatially-varying deconvolution that is $625-1600\times$ faster than the baseline iterative method (Spatially-Varying FISTA [109]), enabling real-time image reconstruction. In addition, incorporating the field-varying PSFs allows our network to have better image quality near the edges of the FoV than is achieved by existing deep learning based methods which assume shift-invariance.

1.5 Related Works

Related Works in Miniature Microscopes

For neural imaging in freely-moving mice, high-resolution two-photon microscopes have been implemented in a small form factor [112, 45]. However, they require expensive hardware (e.g. femtosecond laser), introduce non-linear motion artifacts [83] and, as a scanning method, must trade-off FoV for temporal resolution. Selective-plane illumination (light sheet) microscopes can also be built in miniaturized versions, achieving faster 3D capture [30] at a cost of adding an external illumination source which increases the size of the implanting hardware.

Unlike scanning approaches, single-shot methods project information from the entire volume onto a 2D image, then computationally reconstruct the 3D volume. This enables capturing a large 3D FoV with temporal resolution limited only by the camera exposure time. Single-shot 3D fluorescence capture has been demonstrated using a lensless architecture [1], but lacked the integrated illumination that is required for *in-vivo* imaging. Other recent work combines coding elements with multi-fiber endoscopes to achieve single-shot non-fluorescent 3D, with relatively low resolution [93]. The miniature light field microscope [96], using a conventional unifocal microlens array, demonstrated a fully-integrated 3D fluorescence system. However, like conventional LFMs, it suffers from degraded lateral resolution outside a narrow axial operating range. Additionally, the device is larger and heavier than the 2D Miniscope due to the added microlens array and propagation distance. Lastly, their algorithm [78], while computationally-efficient, relies on temporal video processing that requires multiple frames of capture and static structure in the sample; thus, it does not generalize to single-shot 3D imaging in all applications. Our system, in comparison, uses a more compact, lightweight design that achieves higher lateral resolution over a larger 3D FoV and is optimized to enable 3D recovery from a single frame. This allows it to be used on a more general class of samples and enables easier motion correction, as no temporally-consistent sample structure is required.

Related Works in single-shot Hyperspectral Imaging

Many single-shot hyperspectral imaging techniques have been proposed and evaluated over the past few years. Most approaches can be categorized into the following groups: spectral filter-based methods, coded aperture methods, speckle-based methods, and dispersion-based methods. *Spectral filter array methods* use tiled spectral filter arrays on the sensor to recover the spectral channels of interest [61]. These methods can be viewed as an extension of Bayer pattern based color imaging, but with more than three filters. As the number of filters increases (increasing the spectral resolution), the spatial resolution decreases. For instance, with an 8 × 8 filter array (64 spectral channels), the spatial resolution is 8× worse in each direction than that of the camera sensor. Demosaicing methods have been proposed to improve upon this, however they rely on intelligently guessing information that is not recorded by the sensor [75]. In contrast, our system combines a spectral filter array with a randomizing diffuser, allowing us to recover close to the full spatial resolution of the sensor, which is not possible with traditional lens-based spectral filter array methods. *Coded aperture methods* use a dispersive optical element, such as a prism or diffractive grating, along with a coded aperture in order to modulate the light [36, 65, 103, 20]. These systems are able to capture hyperspectral images and videos but tend to be large table-top systems (1 meter long) consisting of multiple lenses and optical components, often much larger than a conventional camera. In contrast, our system has a much smaller form factor, requiring only a camera sensor with an attached spectral filter array and a thin diffuser placed close to the sensor. *Speckle-based methods* use the wavelength dependence of speckle from a random media to achieve hyperspectral imaging [87, 34]. These systems can be compact, since they require only a sensor and scattering media as their optic, however their spectral resolution is limited by the speckle correlation through wavelengths. This is challenging to design for a given application and has worse spectral resolution than our proposed system. *Dispersive methods* utilize the dispersion from a prism or diffractive optic to encode spectral information on the sensor, without the use of a coded aperture. As demonstrated in [10], this can be accomplished opportunistically using a prism and standard DSLR camera. Such a system can have high spatial resolution, equal to that of the camera sensor, however the spectral information is encoded only at the edges of objects in the scene, resulting in a highly ill-conditioned problem and lower spectral accuracy. Other methods use a dispersive diffuser as opposed to a prism as the dispersive element [39]. This can be more compact than prism-based systems, however there is a trade-off between spatial and spectral resolution depending on the amount of dispersion and has only been demonstrated for up to 33 spectral bands. The spatial resolution can be improved by including an additional RGB camera to form a dual-camera system at the cost of additional hardware and space [44]. While these dispersive methods are more compact than coded-aperture methods, they still require a lens assembly in addition to a prism or diffractive element. This compactness challenge is addressed by [104, 49], in which a single diffractive optic is designed to act both as the lens and the dispersive element, uniquely encoding spectral information in a spectrally-rotating PSF. In contrast to other dispersive methods, our system is more compact and has a similar size as [49]. Our system differs from [49] in that our spectral and spatial resolutions are decoupled, enabling custom sensors tailored to specific spectral filter bands that do not need to be contiguous, enabling more flexibility in the imager design.

Related Works in Spatially Varying Deconvolution

A variety of approaches have been proposed for deconvolution over the years, ranging from Wiener filtering to iterative optimization approaches, such as Richardson-Lucy or FISTA, along with a number of hand-crafted priors, such as TV, sparsity, etc. The simplest models assume that the PSF is shift-invariant, however real systems often have a spatially-varying PSF that are field-varying. Several methods have been proposed to deal with spatially-varying deconvolutions, such as low-rank models [109] or local-convolutional models [59], however they are typically slow and computationally intensive, making them unsuitable for real-time image reconstruction. Recently, deep-learning based deconvolution methods have been demonstrated to improve image quality and reconstruction speed for deconvolutions, providing a promising improvement over traditional approaches [76, 54] [98, 85]. However, to date, these methods rely on a shift-invariant PSF approximation and are not well-suited for most optical systems with spatially-varying PSFs. Our network architecture consists of multiple Wiener deconvolution layers paired with a convolutional neural network. This allows the network to take into account the field-varying behaviour of the optical system resulting in faster reconstructions than classical iterative approaches and better reconstructions than deep learning methods that assume shift-invariance.

Chapter 2

Optimized Single-shot Miniature 3D Fluorescence Microscopy

This chapter is based on [109] and is joint work with Nick Antipa, William Liberti, Sam Dehaeck, Kristina Monakhova, Fanglin Linda Liu, Konlin Shen, Ren Ng, and Laura Waller.

2.1 Abstract

Miniature fluorescence microscopes are a standard tool in systems biology. However, widefield miniature microscopes only capture 2D information, and modifications that enable 3D capabilities increase size and weight, and have poor resolution outside a narrow depth range. Here, we achieve 3D capability by replacing the tube lens of a conventional 2D Miniscope with an optimized multifocal phase mask at the objective's aperture stop. Placing the phase mask at the aperture stop significantly reduces the size of the device and varying the focal lengths enables uniform resolution across a wide depth range. The phase mask encodes 3D fluorescence intensity into a single 2D measurement and the 3D volume is recovered by solving a sparsity-constrained inverse problem. We provide methods for designing and fabricating the phase mask and an efficient forward model that accounts for the field-varying aberrations in miniature objectives. We demonstrate a prototype that is 17 mm tall and weighs 2.5 grams, achieving 2.76 μm lateral and 15 μm axial resolution across most of the 900 × 700 × 390 μm^3 volume at 40 volumes per second. The performance is validated experimentally on resolution targets, dynamic biological samples, and mouse brain tissue. Compared to existing miniature single-shot volume-capture implementations, our system is smaller, lighter, and achieves more than 2× better lateral and axial resolution throughout a 10× larger usable depth range. Our microscope design provides single-shot 3D imaging for applications where a compact platform matters, such as volumetric neural imaging in freely-moving animals and 3D motion studies of dynamic samples in incubators and lab-on-a-chip devices.

2.2 Introduction

Miniature widefield fluorescence microscopes enable important applications in systems biology - for example, optical recording of neural activity in freely-moving animals [37, 64, 48, 42], and long-term *in situ* imaging within incubators and lab-on-a-chip devices. These miniature microscopes, commonly called 'Miniscopes', are developed by a vibrant open-source community [102] and made of 3D printed parts and off-the-shelf components. While the Miniscope is designed for 2D fluorescence imaging only, many applications can benefit from imaging 3D structure.

Volumetric microscopy methods aim to capture 3D structure; however, they often rely on scanning (e.g. two-photon, light sheet) which is difficult to miniaturize and must trade off temporal resolution and field-of-view (FoV). Two-photon microscopes have been implemented in small form factors [112, 45], giving high resolution at a cost of motion artifacts [83], limited FoV, and expensive hardware. Miniaturized light sheet microscopes achieve faster capture [30], but also depend on scanning which causes motion artifacts and increases the complexity and size of the hardware.

Unlike scanning approaches, single-shot methods [108, 67, 3, 59, 1, 7, 63, 17] offer faster capture speeds, with temporal resolution limited only by the camera frame rate. These methods encode information from the entire volume into a 2D measurement, then computationally reconstruct the 3D information. Single-shot 3D fluorescence capture has been demonstrated using a lensless architecture [1, 59], but lacked the integrated illumination that is required for in-vivo imaging. In addition, such mask-only systems have no magnifying optics, and so are limited to low effective numerical aperture (NA) resulting in poor lateral and axial resolutions. Other recent work combines coding elements with multi-fiber endoscopes to achieve single-shot non-fluorescence 3D, with relatively low resolution [93]. Recently, the miniature light field microscope (MiniLFM) [96] demonstrated an integrated 3D fluorescence system with computationally-efficient temporal video processing for neural activity tracking [78]. This system adds a standard microlens array (regularly-spaced, unifocal) to the image plane of the Miniscope, giving it single-shot 3D capabilities at the cost of degraded lateral resolution and a larger and heavier device. The MiniLFM algorithm [78] is optimized for neural activity tracking applications, so uses temporal video processing that requires sparsity, multiple frames of capture and static structure in the sample.

Here, we present a new single-shot 3D miniature fluorescence microscope, termed *Miniscope3D*, that is not only smaller and lighter weight than MiniLFM, but also achieves better resolution over a larger volume. It is designed as a simple hardware modification to the widely-used UCLA Miniscope [102], replacing the tube lens with an optimized phase mask placed directly at the aperture stop (Fourier plane) of the objective lens (Fig. 2.1). The phase mask consists of a set of multifocal nonuniformly-spaced microlenses, optimized such that each point within a 3D sample generates a unique high-frequency pattern on the sensor, encoding volumetric information in a single 2D measurement. The 3D volume is recovered by solving a sparsity-constrained compressed sensing inverse problem, enabling us to recover 24.5 million voxels from a 0.3 megapixel measurement. Our algorithm assumes the sample to be sparse in some domain, which is valid for a general class of fluorescent samples. We demonstrate the capabilities of our microscope by imaging fluorescent resolution targets, freely swimming biological samples, scattering mouse brain tissue, and optically cleared mouse brain tissue. We also validate the accuracy of our reconstructions against two-photon microscopy and discuss the limitations of our method.

To achieve high-quality imaging in a small, low-weight device, a number of technical innovations were developed. Placing the phase mask in Fourier space (instead of image space) significantly improves compactness, and also reduces computational burden [70, 92, 43]. Varying the focal lengths of the microlenses enhances the uniformity of resolution across depth, as compared to implementations like MiniLFM. Because we use an optimized forward model and calibration scheme to account for the field-varying aberrations inherent to miniature objectives, we are able to add 3D capabilities to the 2D Miniscope, at a cost of only a small loss of lateral resolution, and lower signal-to-noise ratio (SNR). Our algorithm unites optical theory with compressed sensing in a general way that can allow others to design and fabricate optimized phase masks for their applications. The main contributions of this work are:

- A new miniature 3D microscope architecture that improves upon MiniLFM, achieving significantly better resolution across a 10× larger depth range, while reducing overall device size.
- A prototype, based on easily available parts, 3D printing, and open-source designs, that weighs 2.5 grams and achieves 2.76 μm lateral and 15 μm axial resolution across most of the $900 \times 700 \times 390$ μm^3 volume at 40 volumes per second.
- Design principles for optimizing phase masks for 3D imaging and a high-quality fabrication method using two-photon polymerization in a Nanoscribe 3D printer.
- An efficient calibration scheme and reconstruction algorithm that accounts for the field-varying aberrations inherent in miniaturized objective lenses.

2.3 Results

We characterize the performance of our computational microscope with samples of increasing complexity, capturing dynamic 3D recordings at frame rates of up to 40 volumes per second.

Resolution Characterization: Lateral resolution is measured at different depths by imaging a fluorescent resolution target every 10 μm axially and determining the smallest resolved group by eye. Figure 2.2(a) demonstrates 2.76 μm uniform lateral resolution over 270 μm in depth. The resolution degrades to 3.9 μm over the next 120 μm in depth, for a total usable depth range of 390 μm. This relatively uniform resolution through a wide depth range is due to our multifocal design. Axial resolution is determined by imaging a thin layer of 4.8 μm fluorescent beads at different depths and using Rayleigh criterion (at

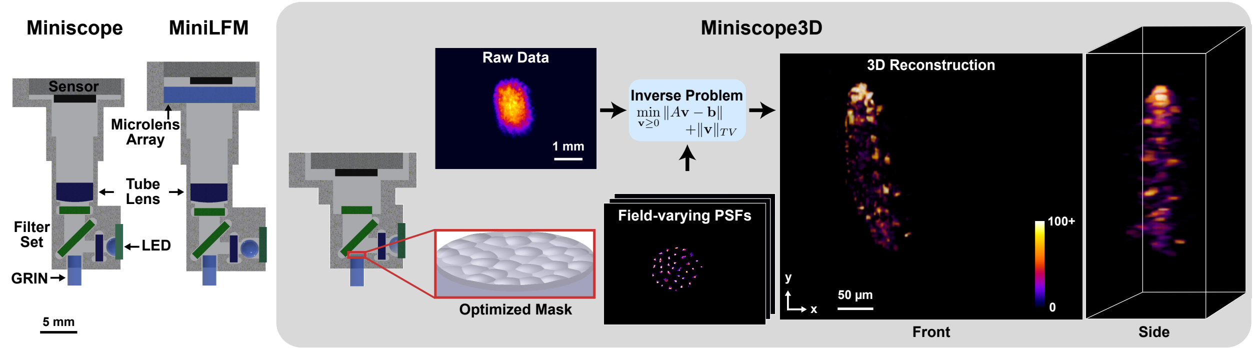

Figure 2.1: **Miniscope3D system overview.** As compared to previous Miniscope and MiniLFM designs, our Miniscope3D is lighter weight and more compact. We remove the Miniscope's tube lens and place a 55 μm thick optimized phase mask at the aperture stop (Fourier plane) of the GRIN objective lens. A sparse set (64 per depth) of calibration point spread functions (PSFs) is captured by scanning a 2.5 μm green fluorescent bead throughout the volume. We use this dataset to pre-compute an efficient forward model that accurately captures field-varying aberrations. The forward model is then used to iteratively solve an inverse problem to reconstruct 3D volumes from single-shot 2D measurements. The 3D reconstruction here is of a freely-swimming fluorescently-tagged tardigrade.

least a 20% dip between the peaks of the two reconstructed points) to determine resolution. Raw data from multiple depths are added to synthesize a measurement of two layers of beads with varying separations (see Fig. 2.10). We achieve 15 μm axial resolution across the entire 390 μm depth range, which matches the axial full-width-half-maximum (FWHM) in the reconstructions of the 3D fluorescent beads sample in Fig. 2.2(b).

Two-Photon Verification: To validate the accuracy of our results, we compare against two-photon microscopy, which is considered ground truth. Figure 2.2(b) shows results for a 160 μm thick sample of 4.8 μm green fluorescent beads. Miniscope3D accurately recovers all the beads in the volume, after visually adjusting for tip/tilt misalignment in post-processing.

Mouse Brain Tissue: Next, we show feasibility for neuro-biological samples by imaging post-fixed mouse brain slices where GFP is expressed in a sparse population of neurons throughout the sample. Figure 2.3(a) shows reconstructions from two 100 μm thick scattering samples from different parts of the hippocampus, and Fig. 2.3(b) shows results from a 300 μm thick optically cleared section. In the 300 μm slice, dendrites can be seen running across the reconstruction axially (~1 μm features), and individual cell bodies appear at distinct depths (see **Video 1**).

Dynamic Biological Samples: Finally, we image dynamic samples of freely-swimming SYBR-green stained tardigrades at a maximum of 40 frames per second. Figure 2.3(c) shows maximum intensity projections of the reconstructed videos at different time points from two different recordings. The reconstructions show that Miniscope3D can track freely-moving

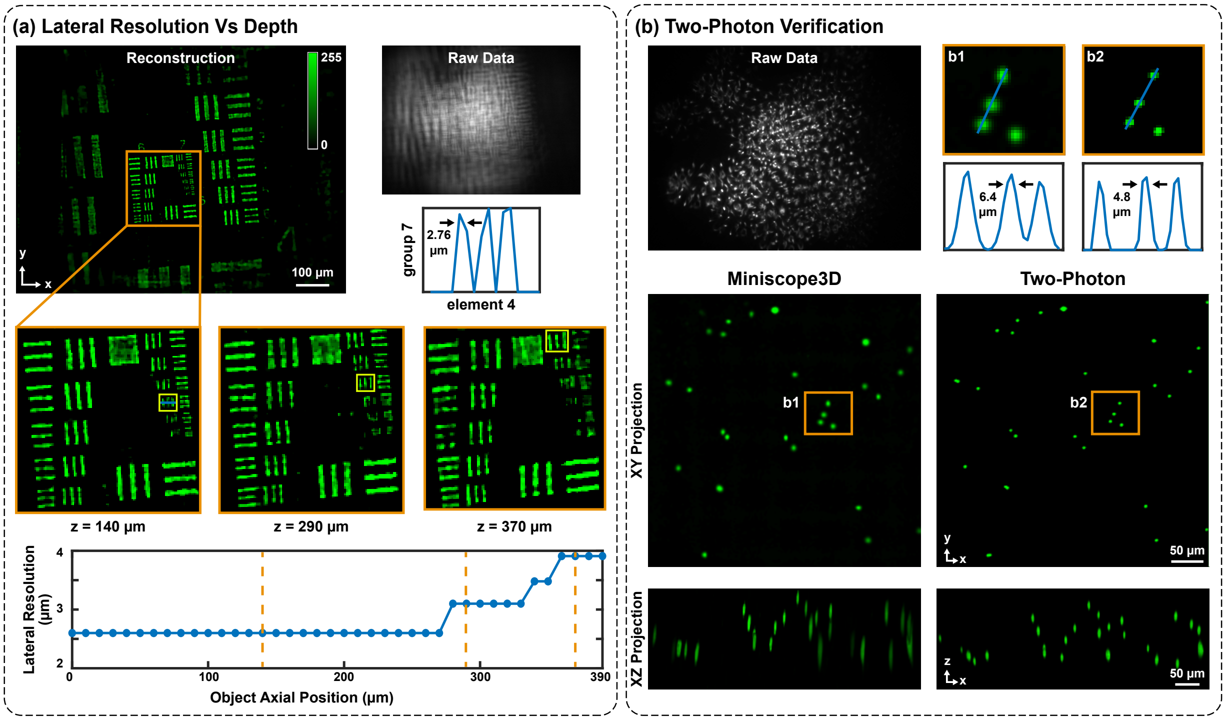

Figure 2.2: **Experimental characterization:** (a) Reconstructions of a fluorescent USAF target at different axial positions to determine depth-dependent lateral resolution. We recover 2.76 μm resolution across most of the 390 μm range of depths, with a worst case of 3.9 μm (dashed orange lines mark inset locations and yellow boxes on insets indicate smallest resolved groups). Note that the resolution target has discrete levels of resolution that result in jumps in the data and resolution refers to the gap between bars, not the line-pair width. (b) Reconstruction of a 160 μm thick sample of 4.8 μm fluorescent beads, as compared to a two-photon 3D scanning image (maximum intensity projections in yx and zx are shown). Our system detects the same features, with a slightly larger lateral spot size.

biological samples at high spatial and temporal resolution (see **Videos 2-6**).

2.4 Theory and Methods

System Theory

Miniscope3D encodes volumetric information via a thin phase mask placed at the aperture stop of the gradient index (GRIN) objective lens (see Fig. 2.1). The goal of our design is to optimize the microscope optics for compressed sensing, enabling capture of a large number of voxels from a small number of sensor pixels. To achieve this, the phase mask comprises

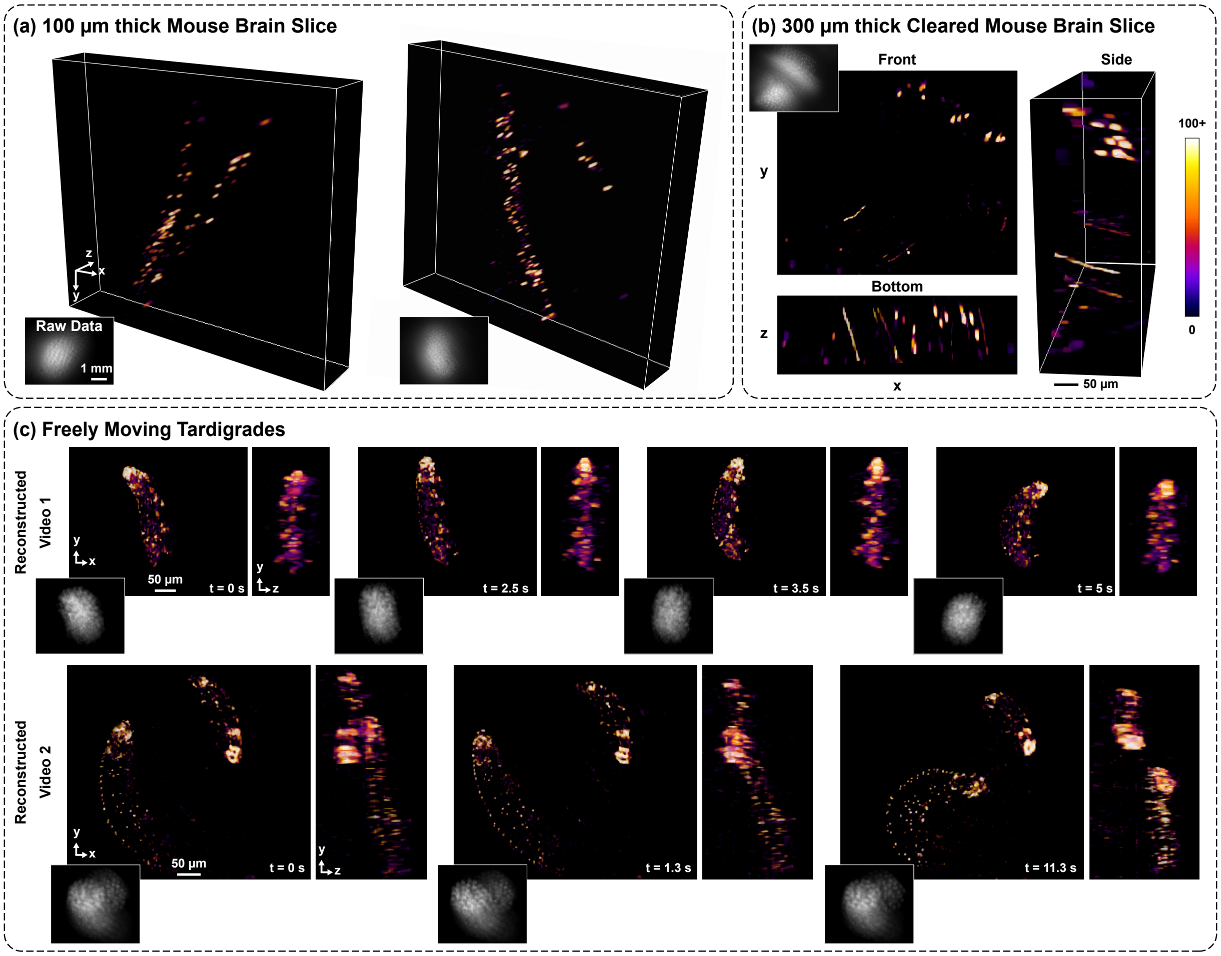

Figure 2.3: **Experimental 3D reconstructions of** (a) GFP-tagged neurons in two different samples of 100 μm thick fixed mouse brain tissue, and (b) 300 μm thick optically cleared mouse brain slice. We clearly resolve dendrites running across the volume axially (see **Video 1**). All mouse brain volume reconstructions are $790 \times 617 \times 210$ μm^3. (c) Maximum intensity projections from several frames of the reconstructed 3D videos of two different samples of freely moving tardigrades captured at a maximum of 40 frames per second (see **Video 2 & 3**).

an engineered pattern of multifocal microlenses, designed such that each fluorescent point source in the scene produces a unique high-frequency pattern of focal spots at the sensor plane, thus encoding its 3D position. The structure and spatial frequencies present in this pattern, termed the *point spread function* (PSF), determine our reconstruction resolution at that position; theory for these limits is presented in the *Lateral Resolution* section below.

Figure 2.4 shows how our PSF changes with the lateral and axial position of a point source in the object space. As the point source moves laterally, the PSF translates (Fig. 2.4(b)). In an idealized microscope with the phase mask in Fourier space, the system would be shift-invariant [70, 92]; however, because of the inherent aberrations in the GRIN lens, the pattern also slightly changes structure as it shifts. As the point source moves axially, the overall PSF changes size and different spots come into focus (Fig. 2.4(a)), because we use a diversity of microlens focal lengths in our phase mask. As discussed in the section on *Multifocal Design*, this ensures that the PSFs at a wide range of depths all contain sharp focal spots, unlike unifocal microlenses. To maximize the performance of our system, we optimize the spacing and focal lengths of the microlenses, as described in the *Phase Mask Optimization* section.

Our distributed, unique PSFs satisfy the multiplexing requirement of compressed sensing. Hence, we utilize sparsity-constrained inverse methods to recover the voxelized sparse 3D fluorescence emission, $\mathbf{v}$, from a single 2D sensor measurement, $\mathbf{b}$. To do this, we model $\mathbf{b}$ as a linear function of $\mathbf{v}$, denoting the measurement process as $\mathbf{b} = A\mathbf{v}$. Here, A is the measurement matrix, a linear operator that captures how each voxel maps to the sensor. Provided the sample is sparse in some domain, we reconstruct the volume by solving the sparsity-constrained inverse problem:

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}\geq 0} \|A\mathbf{v} - \mathbf{b}\|_2^2 + \tau\|\Psi\mathbf{v}\|_1, \tag{2.1}$$

with Ψ being a sparsifying transform (e.g. 3D gradient, corresponding to TV regularization) and τ being a tuning parameter.

Equation 2.1 can be solved using a variety of iterative methods; we use Fast Iterative Shrinkage Thresholding (FISTA) [11]. This requires repeatedly applying A and its adjoint. To make this computationally feasible for high megavoxel systems like ours, we need an efficient representation for A. A shift-invariant forward model is extremely computationally efficient because A becomes a convolution matrix [3, 4, 52]. It also requires only a single PSF calibration image, from which the PSFs at all other positions can be inferred. Unfortunately, miniature integrated systems like ours are not shift invariant, due to the off-axis aberrations inherent to compact objectives. To account for this, in the following sections we develop a field-varying forward model and a practical calibration scheme that account for aberrations with minimal added computational cost.

Field-varying Forward Model

Because aberrations in the GRIN lens of the Miniscope render the shift-invariant model invalid, we need to both measure and model how the PSF changes across the FoV. Explicitly measuring the PSF at each position within the volume is infeasible, both in terms of amount of calibration data and computational burden of reconstruction. It is also unnecessary since the PSF structure changes slowly across the FoV. Instead, our calibration scheme samples the PSF sparsely across the field and uses a weighted convolution model to estimate the PSF at other positions [33]. We capture 64 PSF measurements at each depth, then use them to predict the full set of over 300,000 PSFs. Our forward model thus only requires computing a limited number of convolutions (typically 10-20) and achieves 2.2× better resolution and better quality than the shift-invariant model (see Fig. 2.4(c)).

Our field-varying forward model approximates A using a weighted sum of shift-invariant (convolution) kernels. We treat the volumetric intensity as a 3D grid of voxels, denoted $\mathbf{v}[x,y,z]$. A voxel at location $[x,y,z]$ produces a PSF on the sensor, $\mathbf{h}[u,v;x,y,z]$, where $[u,v]$ indexes sensor rows and columns. For ease of notation we will assume the system has magnification $M=1$ and apply appropriate scaling to the solution after 3D image recovery. We also assume $\mathbf{v}$ has finite axial and lateral support. By treating the voxels as mutually incoherent, the measurement will be a linear combination of PSFs:

$$\begin{aligned}\mathbf{b}[u,v] &= \sum_{z}\sum_{x,y}\mathbf{v}[x,y;z]\mathbf{h}[u,v;x,y,z] \\ &= A\mathbf{v},\end{aligned} \tag{2.2}$$

where the bounds of the sums implicitly contain the sample. To capture field-varying behavior, we seek to model the PSF from each voxel as a weighted sum of K shift-invariant kernels [33]. The kernels, $\mathbf{g}_r[u,v;z]$, and weights, $\mathbf{w}_r[x,y,z]$, which will be described below, should be chosen to represent all PSFs accurately with the smallest possible K. Mathematically, the forward model can be written as:

$$\mathbf{h}[u,v;x,y,z] = \Lambda[u,v]\sum_{r=1}^{K}\mathbf{w}_r[x,y,z]\mathbf{g}_r[u-x,v-y;z], \tag{2.3}$$

where $\Lambda[u,v]$ is an indicator function that selects only the values that fall within the sensor pixel grid. In other words, the PSF from position $[x,y,z]$ is modeled by shifting the kernels, $\{\mathbf{g}_r[u,v;z]\}$ $r=1\ldots K$, associated with depth z, to be centered at the PSF location on the sensor, $[u,v]=[x,y]$. Then, each kernel is assigned a field-dependent weight, $\mathbf{w}_r[x,y,z]$, and the weighted kernels are summed over r. Note that this motivates the placement of the phase mask in the aperture stop. By ensuring that all field points fully illuminate the mask, the system will be close to shift-invariant, which will keep the necessary number of kernels low.

To find the kernels and weights that best represent all of the PSFs, first consider each PSF in a coordinate space relative to the chief ray. We do this by centering each measured PSF on-axis:

$$\mathbf{h}[u+x,v+y;x,y,z] = \sum_{r=1}^{K}\mathbf{w}_r[x,y,z]\mathbf{g}_r[u,v], \tag{2.4}$$

where $[x,y]$ is the chief ray spatial coordinate at the sensor. We assume that the calibration procedure will capture N PSFs across the field, $\{\mathbf{h}[u,v;x_i,y_i,z]\}$ $i=1\ldots N$, for each depth z. We estimate the chief ray coordinate $[x,y]$ of off-axis PSFs by cross-correlating each with the on-axis PSF. The off-axis measurements are then shifted on-axis, vectorized, and combined into a registered PSF matrix, denoted H. For smoothly varying systems, H will be low rank and can be well approximated by solving

$$\hat{G}, \hat{W} = \underset{G,W}{\arg\min} \left\|GW - H\right\|_2^2 , \tag{2.5}$$

where $G \in \mathbb{R}^{M_p \times K}$ and $W \in \mathbb{R}^{K \times N}$ for a sensor with M_p pixels. The optimal rank-K solution can be found by computing the the K largest values of the singular value decomposition (SVD) of H. The r-th column of the left singular vector matrix, $\hat{G}$, contains the kernel $\mathbf{g}_r[x, y; z]$ in vectorized form. Similarly, combining the singular values with the right singular vector matrix produces $\hat{W}$, of which the r-th row contains the optimal weights $\mathbf{w}_r[x_i, y_i, z]$ for voxel $[x_i, y_i, z]$. Empirically, we find that the weights vary smoothly across the field, so we use natural neighbor interpolation to estimate the weights between sampled points. After testing the number of sample points per depth (N) empirically, we find 64 to be sufficient for our system.

The computational-efficiency of this model can be analyzed by substituting Eq. 2.3 into Eq. 2.2, yielding:

$$\begin{aligned}\mathbf{b}[u, v] &= \sum_{z} \sum_{x,y} \mathbf{v}[x, y, z]\Lambda[u, v] \sum_{r=1}^{K} \mathbf{w}_r[x, y, z]\mathbf{g}_r[u - x, v - y; z] \\ &= \Lambda[u, v] \sum_{z} \sum_{r=1}^{K} \left\{ (\mathbf{v}[x, y, z]\mathbf{w}_r[x, y, z]) \overset{[x,y]}{*} \mathbf{g}_r[x, y; z] \right\} [u, v] ,\end{aligned} \tag{2.6}$$

where $\overset{[x,y]}{*}$ denotes discrete linear convolution over the lateral variables. In practice, each convolution can be implemented using a combination of padding and FFT-convolution, while $\Lambda[u, v]$ represents a crop [3]. Note that the summation over z assumes no voxel is partially occluded. Because this model comprises K point-wise multiplications and K 2D convolutions per depth, it is approximately $K-$times slower than a shift-invariant model. Hence minimizing K via choice of weights and kernels, or by reducing aberrations in the hardware, improves computational efficiency.

Calibration

Experimentally, our calibration procedure captures PSF images of a 2.5 μm green fluorescent bead at 64 equally-spaced points across the FoV, for each depth. Empirically, we find that the singular values decay quickly and a model with rank between $K = 10$ and $K = 20$ is sufficient for our system. Note that we can trade-off the speed and accuracy of our model by varying K, but the decomposition need only be performed once. This method allows characterization of an extremely large matrix by only capturing a relatively small number of images. For example, our typical calibration requires 80 depths. Densely sampling every PSF using a 0.3 megapixel sensor would require 24 million calibration images (300,000 per depth) and terabytes of storage. In contrast, our method enables calibrating this entire volume using only 80 depths $\times$ 64 images/depth $= 5,120$ images, which takes 2 hours to capture using automated stages and requires a few gigabytes to store.

Reconstruction Algorithm

In solving Eq. 2.1 we use sparsifying transform $\Psi = [\nabla_x \nabla_y \nabla_z]^\intercal$, which corresponds to 3D anisotropic TV regularization, promoting sparse 3D gradients in the reconstruction. The regularization parameter, τ, controls the balance between the data fidelity and the sparse 3D gradients prior. In practice, we hand-tune τ on a range of test data, then leave it fixed for subsequent captures (see Fig. 2.13). We solve Eq. 2.1 using FISTA [11], with the fast, subiteration-free parallel proximal method [50]. Computationally, our method has similarities to light-field deconvolution [17], but because our PSF is not periodic and our focal lengths are not all the same, we are able to remove the need for aperture matching and achieve higher resolution across a larger volume. In order to solve Eq. 2.1, we compute the linear forward and adjoint matrix-vector multiplies using FFT-convolution. A typical reconstruction takes 1-3k iterations, and runs in 8-24 minutes on a GPU RTX 2080-Ti using MATLAB.

Phase Mask Design

In this section, we present theory for designing and optimizing a phase mask that achieves a target resolution uniformly across a specified 3D volume. We assume that the phase mask will be placed in the aperture stop of the objective with the sensor at a fixed distance, since this architecture reduces the size and weight of our device, makes the system close to shift-invariant and enables multiplexing, which is necessary for compressed sensing. We aim for all PSFs produced by the mask to have high spatial-frequency content and be mutually incoherent (i.e. all as dissimilar as possible). Toward this goal, we propose a multifocal array of nonuniformly-spaced microlenses as our phase mask.

We choose to use a phase mask made of microlenses because it provides good light throughput, while balancing the trade-offs between SNR and compressive sensing capabilities. Our previous work employed off-the-shelf diffusers with a pseudorandom Gaussian surface profile [3]. These generate a caustic PSF that has poor SNR due to the spreading of the light by the concave bumps of the diffuser surface. In contrast, microlenses concentrate the light into a small number of sharp spots, giving better performance in low-light applications like fluorescence microscopy (see Sec. 2.5). By parameterizing our design as a set of microlenses, we can also derive simple design rules from first-principles (sections *Lateral Resolution* & *Multifocal Design*), then use those to formulate an optimization problem that locally optimizes the placement and aberrations of each microlens.

We space our microlenses nonuniformly to ensure that the PSFs from all field points are dissimilar. Regularly-spaced arrays will produce highly similar PSFs when shifted by one microlens period, causing certain spatial frequencies to be poorly measured. Previous work avoided this ambiguity by introducing a field stop [70, 92, 43] that prevents the PSFs from overlapping, but this restricts the FoV significantly. Our design yields a larger FoV by using nonuniform spacing and computationally disambiguating the overlapping PSFs. In Fig. 2.5 we compare PSFs and reconstructions from regularly-spaced and nonuniform phase mask designs. Looking at Fig. 2.5(c), the PSF of the regular array causes unwanted peaks at low frequencies in its radially-averaged *inverse power spectral density* (IPSD), a metric related to deconvolution performance [24] (lower is better). This manifests as artifacts in the simulated reconstruction, which are significantly reduced in reconstructions from both of the nonuniform designs.

Using multiple microlens focal lengths extends the depth range across which we obtain good resolution, as described in the section on *Multifocal Design*. Multifocal designs have sharp focal spots across a wider desired depth range than can be achieved with unifocal designs, trading SNR in-focus for better performance off-focus. Figure 2.5(c,d) compares the PSFs and reconstruction quality of our approach versus unifocal designs in-focus and 200 μm away from the native focus of the unifocal arrays. The blurry features in the out-of-focus PSFs for both unifocal designs cause poor performance, as shown in the reconstructions and high inverse power spectra. To capture the performance across depth, Fig.2.5(b) shows the integrated IPSD (up to the cutoff frequency) of each design versus depth. As expected, our multifocal design is slightly worse than a unifocal design in focus, but achieves far better (lower) values across the full depth range.

In the compact system architecture we propose, it is clear that our nonuniform multifocal microlenses are a good choice of phase mask. This motivates the next sections which provide guidance on optimizing the nonuniform spacing, as well as the focal lengths and aberrations of the microlenses for achieving a target resolution and depth range. For our prototype, we aim for 3.5 μm lateral resolution, and show that this can be achieved over a depth range up to 360 μm, which agrees with our experimental characterization.

Lateral Resolution

Lateral resolution will be primarily determined by the diffraction-limited aperture size of the microlenses, which also determines the number of microlenses that fit across the objective's full aperture, and thus, the depth range we can target. We design for lateral resolution that does not require the full pupil, so that we can fit multiple microlenses in the aperture for better depth coding. The example in Fig. 2.5 targets 3.5 μm resolution (cutoff frequency of 0.35 cycles/μm) using 36 microlenses with average NA=0.09. Because each design has the same number of microlenses, each has a similar resolution limit.

To quantify, we perform a diffraction analysis to find the clear aperture a single microlens needs to support a δx lateral resolution at the sample. Note that this assumes we will recover resolution no better than the band-limit of the measurement, neglecting any resolution gained from the non-linear solver. We start by calculating the magnification for our system:

$$M \approx \frac{-t}{f_G}, \tag{2.7}$$

where f_G is the GRIN focal length and t is the mask-to-sensor distance (derivation in Sec. 2.7). Note that M is approximately independent of the microlens focal length. For our system, $f_G = 1.67$ mm and $t = 8.7$ mm, so $M \approx -5.2$. Using Eq. 2.7 and the Rayleigh criterion, the microlens clear aperture, Δ_M, needed for a target object resolution δx at wavelength λ is:

$$\Delta_{ML} = \frac{1.22\lambda t}{|M|\delta x} \approx \frac{1.22\lambda f_G}{\delta x}. \tag{2.8}$$

This expression is also independent of the microlens focal length because we have assumed the microlens is focused. Equation 2.8 allows us to select the appropriate average microlens spacing for a desired resolution. Our system is designed for 3.5 μm lateral resolution (though experimentally we achieve 2.76 μm, due to the non-linear solver), which gives an average microlens diameter of 300 μm. Given that the GRIN clear aperture has diameter 1.8 mm, this results in 36 microlenses that can fit in the phase mask. Note that since the GRIN is aberration limited, the 2D Miniscope does not achieve the diffraction-limited resolution predicted by its full aperture size. Hence, our experimentally-measured resolution is not much worse than the 2D Miniscope (lateral resolution of 2 μm), despite dividing the GRIN pupil into 36 regions to add depth sensing capabilities.

Multifocal Design for Extended Depth Range

Focal length diversity in the microlens array results in an extended depth range, a key advantage of our architecture over conventional LFM. To maintain a uniform lateral resolution across all depths in the volume of interest, the PSF should have sharp, high-frequency focal spots for each axial position. This requires at least one microlens to be in focus for each object axial plane, with planes spaced by the microlens depth-of-field (DoF). The DoF of a single microlens, d_{ML}, is inversely proportional to the microlens clear aperture, Δ_{ML}, giving $d_{ML} = \pm 20$ μm in our system (see Sec. 2.8 for details).

Our design aims to have a minimum of 4 microlenses in focus within each DoF. Given that our lateral resolution criterion allows 36 microlenses, this means we should have 9 different focal lengths and a depth range of 360 μm, nearly 10× what a single focal length achieves. Note that there is a trade-off between the imaging depth range and lateral resolution. We can increase the depth range by including more microlenses in the mask; however, that decreases their clear aperture (Eq. 2.8) and thus the lateral resolution. Conversely, for imaging thin samples where only a narrow range of focal lengths is required, better lateral resolution is possible.

To determine the focal length distribution, we find the focal length needed to focus at the beginning of the depth range (f_{min} = 7mm) and at the end of the depth range (f_{max} = 25 mm). Then, we dioptrically space the focal lengths across the target range because this leads to microlenses that come into focus at linearly-spaced depth planes in the sample space.

Phase Mask Parameterization

The previous sections outlined first-order design principles, considering only a single microlens. In the next section, we will optimize the ensemble of microlenses (their positions and added aberrations) with metrics based on compressed sensing theory. Here, we first build our representation of the microlens phase mask by parameterizing the i^{th} microlens by its lateral vertex location, $(\rho^i_{xc}, \rho^i_{yc}) := \boldsymbol{\rho^i_c}$ and radius of curvature, R_i. The spherical sag of the microlens is:

$$s_i = d_i + R_i \sqrt{1 - \left(\frac{\boldsymbol{\rho} - \boldsymbol{\rho}^i_c}{R_i}\right)^2}, \tag{2.9}$$

where d_i is an offset constant added to each microlens to control its clear aperture. We parameterize aspheric terms in the microlenses by using Zernike polynomials. The j^{th} Zernike coefficient for microlens i is denoted α_{ij}, so the total aspheric component at that microlens is $\sum_j \alpha_{ij} Z_j(\boldsymbol{\rho} - \boldsymbol{\rho}^i_c)$ with Z_j being the j^{th} Zernike polynomial. As long as the microlenses are all convex ($R_i > 0$), a phase mask with full fill-factor can be constructed by taking the point-wise maximum thickness (see Fig. 2.6). The phase mask surface is thus:

$$T(\rho_x, \rho_y; \boldsymbol{\theta}) = \max_i \left[s_i + \sum_j \alpha_{ij} Z_j(\boldsymbol{\rho} - \boldsymbol{\rho}^i_c) \right], \tag{2.10}$$

where $\boldsymbol{\theta}$ denotes the collection of parameters that define the phase mask: vertex locations $\{\boldsymbol{\rho}^i_c\}$, radii $\{R_i\}$, offsets $\{d_i\}$, and Zernike coefficients $\{\alpha_{ij}\}$. The resulting surface is guaranteed to be continuous and will have a well-defined local focal length given by $f_i = \frac{n-1}{R_i}$ within the region belonging to the i^{th} microlens, provided the power Zernike $j = 4$ is excluded. In practice, we optimize the Zernike coefficients for tilt ($j = 1, 2$) and astigmatism ($j = 3, 5$).

With the microlens array defined, the on-axis PSF at a given sample depth z can be modeled by Fresnel propagation of the pupil wavefront from a point source at depth z, denoted $W(\rho_x, \rho_y; z)$, multiplied by the phase of the designed mask, $\phi(\rho_x, \rho_y; \boldsymbol{\theta}) = \frac{2\pi(n-1)}{\lambda} T(\rho_x, \rho_y; \boldsymbol{\theta})$:

$$\mathbf{h}(u, v; z, \boldsymbol{\theta}) = \left| F_t \left\{ P(\rho_x, \rho_y) \exp\left[i\phi(\rho_x, \rho_y; \boldsymbol{\theta}) \right] W(\rho_x, \rho_y; z) \right\} \right|^2, \tag{2.11}$$

where $P(\rho_x, \rho_y)$ is the GRIN pupil amplitude, n is the microlens substrate index of refraction, and F_t denotes Fresnel propagation to the sensor a distance t away. Importantly, the on-axis PSFs are differentiable with respect to the microlens parameters, $\boldsymbol{\theta}$, enabling us to optimize the design using gradient methods, as discussed in the next section.

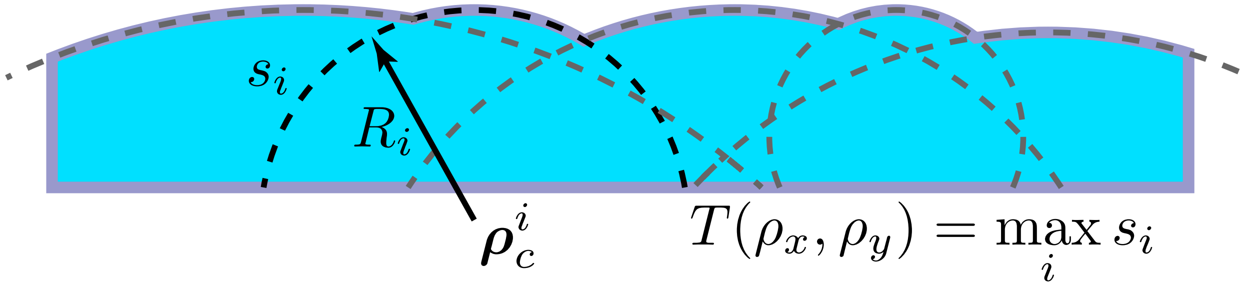

Figure 2.6: **Phase mask parameterized by point-wise maximum of convex spheres.** Each sphere is outlined by a dashed line, and the final optic is shaded blue (not to scale).

Phase mask Optimization Using Matrix Coherence

Given the first-principles guidance in the above sections, we set the number of microlenses, their characteristic aperture size and their focal length distribution; next, we aim to optimize the microlens positions and aberrations to maximize performance. In order to make the optimization computationally feasible, we ignore the field-varying changes in the PSF and assume that the system is shift invariant for the purposes of design.

To optimize the microlens parameters, $\boldsymbol{\theta}$, in terms of the on-axis PSFs at each depth, we set up a loss function to be optimized that consists of two terms. The first term, a cross-coherence loss, promotes good axial resolution by ensuring that the PSFs at different depths are as dissimilar as possible. Cross-coherence between any two depths is defined as $\|\mathbf{h}(u,v;z_n) \star \mathbf{h}(u,v;z_m)\|_\infty := \max\left[\mathbf{h}(u,v;z_n) \star \mathbf{h}(u,v;z_m)\right]$, where $\star$ represents 2D correlation and $\max \cdot$ is the element-wise maximum. Intuitively, we want the cross-coherence to be small, since it represents the worst-case ambiguity that would arise by placing two point sources adversarially at depths spaced according to the separation of their PSF's cross-correlation peaks. By computing this quantity for all pairs of z-depths, we can produce a differentiable figure-of-merit that optimizes the matrix coherence [19] between depths. In practice, rather than optimizing the cross-coherence, we smoothly approximate the max [25] using $\|x\|_\infty \approx \sigma \ln \sum \exp\left(x^2/\sigma\right)$. Here, $\sigma > 0$ is a tuning parameter that trades accuracy of the approximation against smoothness. For our purposes, this has the advantage of penalizing all large cross correlation values, not just the single largest. We will denote this $\|\cdot\|_{\overline{\infty}}$.

The total cross-coherence loss is then:

$$q(\boldsymbol{\theta}) = \sum_n \sum_{m>n} \|\mathbf{h}(u,v;\boldsymbol{\theta},z_n) \star \mathbf{h}(u,v;\boldsymbol{\theta},z_m)\|_{\overline{\infty}}. \tag{2.12}$$

The second term in the optimization ensures that lateral resolution is maintained. To do so, we optimize the autocorrelation of the PSF at each depth using the frequency domain least-squares method. The analysis in the *Lateral Resolution* section above only applies to a single microlens; building a phase mask of multiple lenses generally degrades resolution by introducing dips in the spectrum that reduce contrast at certain spatial frequencies. Hence, we treat the single-lens case as an upper limit that defines the bandlimit of the multi-lens PSF. To reduce spectral ripple, we penalize the ℓ_2 distance between the MTFs of the PSF and a diffraction-limited single microlens, $|H|$. We include a weighting term, denoted D, that ignores spatial frequencies beyond the bandlimit, as well as low spatial frequencies which are less critical and difficult to optimize due to out-of-focus microlenses. The autocorrelation design term is then

$$p(\boldsymbol{\theta}) = \sum_n \left\| D\left[\mathbb{F}\left\{\mathbf{h}(u, v; \boldsymbol{\theta}, z_n) \star \mathbf{h}(u, v; \boldsymbol{\theta}, z_n)\right\} - |H|^2\right]\right\|_2^2, \tag{2.13}$$

where $\mathbb{F}\{\cdot\}$ is the 2D discrete Fourier transform.

The total loss is the weighted sum of the two terms:

$$f(\boldsymbol{\theta}) = p(\boldsymbol{\theta}) + \tau_0 q(\boldsymbol{\theta}), \tag{2.14}$$

where τ_0 is a tuning parameter to control their relative importance. To initialize, we randomly generate 5,000 heuristically-designed candidate phase masks, each with 36 microlenses spaced according to Poisson disc sampling across the GRIN aperture stop. The focal lengths are distributed dioptrically between the minimum and maximum values computed in the *Multifocal Design* section. The best candidate from these 5,000 is then optimized using gradient descent applied to $f(\boldsymbol{\theta})$. This is implemented in Tensorflow Eager to enable GPU-accelerated automatic differentiation.

The results of our optimized design are shown in Fig. 2.7, where we compare our optimized mask to the random multifocal design that scored worst during initialization, and a regular unifocal array. The optimized design has the best axial cross-coherence (Fig. 2.7(b)), with the random array having worse off-diagonal terms. Hence, in the 3D reconstructions (Fig. 2.7(c)) the optimized design performs slightly better than the random design. The regular microlenses produce large off-diagonal peaks in the cross-coherence which manifests as poor 3D reconstruction performance off-focus.

Phase Mask Fabrication

Since our phase mask designs can be tailored to specific applications with different resolution requirements and volumes-of-interest, the ability to rapidly generate phase mask prototypes is very useful. Recently, the Nanoscribe two-photon polymerization 3D printer has been shown to print free-form microscale optics on-demand [101]. However, in its current implementation, Nanoscribe uses planar galvanometric scanning to polymerize the resist, resulting in a limited FoV (diameter of approximately 350 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. Stitching artifacts from this process can seriously impact the produced object [26], usually by causing rectangular or hexagonal blocking artifacts. As can be seen in Fig. 2.8(a), rectangular seams going trough the center of the microlenses can be very detrimental to our design.

One solution to this is an adaptive stitching algorithm that has been demonstrated for slender objects and a non-overlapping microlens array [26]. Here, we propose a new height-based segmentation algorithm capable of placing the stitching seams in the overlapping region between the overlapping microlenses (Fig. 2.8(a)). This is based on the local height functions for each microlens, described in the *Phase Mask Parameterization* section. Each of these functions has a region where they result in the largest values and this region is precisely the printing block that will be printed from that microlens center location (see Sec. 2.14). Once the adaptive stitching mask is obtained, the writing instructions per block can be generated using TipSlicer [88]. Figure 2.8(b) compares the designed and experimental PSFs at three depth planes, showing a good match with some degradation at the end of the volume.

Device Assembly

Our prototype Miniscope3D system consists of a custom phase mask, a CMOS sensor (Ximea MU9PM-MH), fluorescent filter set (Chroma ET525/50m, T495lpxr, ET470/40x), GRIN lens (Edmund Optics 64-520), and half-ball lens (Edmund 47-269), with a 3D-printed optomechanical housing. The 55 μm thick phase mask is glued to the back surface of the GRIN lens using optical epoxy. Note that our experimental PSF calibration accounts for slight misalignment in the phase mask. The final device is 17 mm tall and weighs 2.5 grams.

2.5 Microlenses vs Gaussian Diffuser

For our phase mask, we choose a microlens array instead of the Gaussian diffuser used in our previous work[3]. This is because the microlenses can achieve point spread functions (PSFs) with higher SNR and frequency content than the diffuser (see Fig. 2.9), due to their better concentration of light in focus. Microlenses focus light into small focus spots, with dark areas between them, as opposed to the diffuser, which has some light spread between the caustics, generating unwanted low frequencies in the PSFs. Sharper focus spots in the microlens PSF mean that the SNR of the measurements is better and the inverse problem better posed. While using fewer focal spots would improve 2D measurement SNR and resolution, using a small number of microlenses does not provide enough multiplexing to gain 3D capability over a large depth range.

2.6 Axial Resolution

We determined the axial resolution by imaging a thin layer of 4.8 μm fluorescent beads. Because it is difficult to controllably place two beads at specific axial separation distances, raw data from a single bead at different depths are digitally added in order to synthesize a measurement of two layers of beads with varying separations. Figure 2.10 shows that we achieve a uniform 15μm axial resolution across our depth range of 360 μm. This closely matches with the axial full-width-half-maximum (FWHM) we observe in the 3D fluorescent beads sample in Fig. 2.2.

2.7 Lateral Resolution

Examining a single microlens, the Rayleigh criterion defines the minimum resolvable separation of two diffraction-limited spots on the sensor, $\delta x'$, in terms of the wavelength, λ, the microlens clear aperture, Δ_{ML}, and the distance from the mask to the sensor, t:

$$\delta x' = \frac{1.22\lambda t}{\Delta_{ML}} = M\delta x \tag{2.15}$$

Here we have used the fact that two points in object space separated by δ_x will appear as a separation of $M\delta_x$ on the sensor. Thus, we need to calculate the magnification of our system.

We use ray transfer matrices (with a paraxial approximation) to evaluate the magnification of the system. The system ABCD matrix is:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1/f_\mu & 1 \end{bmatrix} \begin{bmatrix} A_G & B_G \\ C_G & D_G \end{bmatrix} \begin{bmatrix} 1 & Q \\ 0 & 1 \end{bmatrix} \tag{2.16}$$

and the system magnification, which is used in the lateral resolution derivation, is:

$$M = A = \left(1 - \frac{t}{f_\mu}\right) A_G + tC_G \tag{2.17}$$

where A_G, B_G, C_G, & D_G are elements for the GRIN's ray transfer matrix ($A_G = 0.0725$, $B_G = 1.6931$, $C_G = -0.599$, and $D_G = 0.124$) and t is the distance from the phase mask to the sensor. Given that f_μ, the microlens focal length, ranges from 7 mm to 25 mm, combined with the small value for A_G, this results in the first term, $(1 - t/f_\mu)A_G$, being negligible and the magnification can be approximated simply as tC_G. This shows that for our system, the magnification is given by:

$$M \approx tC_G \tag{2.18}$$

Substituting Eq. 2.18 into Eq. 2.15 and solving for Δ_{ML}, we get an expression for the microlens clear aperture needed for a target object resolution:

$$\Delta_{ML} = \frac{1.22\lambda t}{M\delta x} \approx \frac{1.22\lambda}{C_G \delta x} \tag{2.19}$$

2.8 Depth of Focus

We aim to determine the microlens depth-of-focus (DoF), defined as the distance that a point source in-focus can move axially before the blur spot on the camera sensor is bigger than a target circle of confusion radius, γ_c. To do so, we examine a single microlens' image in the GRIN entrance pupil for an object at distance z from the first principal plane of the GRIN. As the object moves axially by a distance d_{ML}, we can use similar triangles to derive (see Fig. 2.12 for variable definitions):

$$\frac{y}{d_{ML}} = \frac{\Delta_{EP}}{d_{ML} + z + L} \approx \frac{\Delta_{EP}}{L} \tag{2.20}$$

where Δ_{EP} is the radius of the microlens' clear aperture in the entrance pupil (i.e. object side) of the GRIN and L is the distance from the first principal plane to the entrance pupil. Given that $L = 13\ mm$ is much larger than z, d_{ML}, which are on the order of 0.2 mm, we drop both z and d_{ML}. By substituting $y = \gamma_c/M$ into Eq. 2.20, we can solve for the microlens DoF as a function of our system parameters:

$$d_{ML} = \frac{\gamma_c L}{\Delta_{EP} M} \tag{2.21}$$

Since the entrance pupil of the GRIN is very far from the object (it is approximately telecentric in object space), the object axial position is negligible in determining the microlens DoF. Designing for $\gamma_c = 12\ \mu m$, a circle-of-confusion smaller than the diffraction-limited spot size, $|M|\delta_x$, and using $\Delta_{EP} = 4\ mm$ (calculated using Zemax for a microlens with a clear aperture of 300 μm), we determine the DoF to be ±20 μm.

2.9 Choice of Reconstruction Grid

To successfully reconstruct $\mathbf{v}$, we should define the reconstruction grid with sufficient sampling to realize the best resolution possible, but without oversampling, which increases computation and memory requirements. The theory above defines a band-limit for the measurements, so our goal is to use a sensor with a matching effective pixel size. In our architecture, increasing the sensor pixel size directly corresponds to increased lateral reconstruction voxel size and lower final resolution. Because of complicated interactions between nonlinear reconstructions and grid size, we determine our choice of lateral sampling empirically by binning the raw data from the resolution tests in Fig. 2.2 by 2×, 4×, and 8× and evaluating the final resolution. We find that the resolution begins to degrade between 4× and 8× binning, so we operate at 4× binning. This results in our sensor's effective object-space pixel size being 1.7 μm, which is sufficiently below the 2.76 μm minimum feature size that we observe experimentally. Note that the ability to use on-chip binning allows our approach to read data faster than a conventional LFM, which cannot use conventional on-chip binning without resolution loss. This allows us to achieve a 40 volume-per-second measurement rate using a low-cost USB 2.0 camera.

The choice of axial sampling informs our sampling interval during calibration (*Calibration* subsection). We measure every 5 μm, and perform axial binning (summing of consecutive PSFs) at 1×, 2×, and 4×. We find 1× yields the best results. The resulting 5 μm axial sampling is reasonable given the empirically observed 15 μm axial resolution. Hence our choice of grid balances fast frame rates and efficient reconstruction with image quality and resolution.

2.10 Choice of Regularization Parameter

One important parameter in our optimization problem is the regularization parameter τ. The regularization parameter sets the trade-off between the data fidelity term and our sparsity prior. In practice, this parameter sets the balance between preserving image details and noise reduction. Very small values of τ will preserve sharp details in our object; however, the reconstructions can be noisy. Very large values will suppress noise, but also suppress the object's details with it.

To test the reconstruction quality as a function of the regularization parameter, we ran our 3D reconstruction algorithm on the experimental resolution target data at $z = 270$ μm with values of τ ranging from 10^{-14} to 10^{-1}. Figure 2.13(a) shows that the reconstructions and the data fidelity term are stable for a wide range of τ values. As expected, for very large values of τ, the Total Variation (TV) prior over-regularizes the image, resulting in smoothed out details.

Since the experimental data lacks ground truth to compare against, we simulate a raw measurement by running our 3D shift-varying forward model on a two-photon microscopy zebra fish 3D dataset with our measured PSFs and adding realistic additive white Gaussian noise. The measurement is then processed with values of τ ranging from 10^{-14} to 10^{-1}. Figure 2.13(b) shows a trend similar to experimental results - the mean-squared error is stable for a large range of τ values, with over-smoothed reconstructions as τ gets very large. We note that all the data shown was processed using the same value of τ, which further show that once a good value for τ is found, it can be used to process different classes of objects. While it may be possible to fine-tune τ for each measurement to achieve better performance, it is, however, more practical for users to use the default value. If the user is to fine-tune τ, we recommend using the largest value of τ that still preserves the object's fine details.

2.11 2D Miniscope PSNR Comparison

Our Miniscope3D design is aimed at 3D imaging, but because it is smaller and lighter weight than 2D Miniscope, it might be useful in applications that only require 2D imaging. Because of the inherent aberrations in the GRIN lens, the 2D Miniscope does not achieve its full-aperture diffraction-limited resolution and our Miniscope3D resolution is only marginally worse than the 2D. However, we do suffer from reduced SNR as compared to the 2D Miniscope, because our PSFs spread the light over a larger area than a focused 2D Miniscope. To quantify this loss of SNR, we simulate measurements using on-axis PSFs from both our device and the 2D Miniscope (single lens with 2 μm blur). The simulation is performed at 3 light levels (100, 1000, and 10,000 photocounts) using a shift-invariant model with Poisson and read noise added. We use our reconstruction algorithm with an optimized τ value and display the results in Fig. 2.14. For a fair comparison, we show both the 2D Miniscope raw image and one reconstructed from an image deconvolution process. Our Miniscope3D system has better PNSR than the unprocessed 2D Miniscope data, but the deconvolved 2D Miniscope result performs the best, as expected. This is because our algorithm is denoising and deblurring. For a scene that does not fit our denoising priors, the processed results would perform worse. Also, note that the loss of PSNR in our system for 2D imaging is a neceessary sacrifice for gaining single-shot 3D imaging capability.

2.12 Sparsity Comparison

Our approach assumes the object to have a sparse representation in some domain. In this paper, we use a general TV sparsity prior to promote gradient sparsity. This is a commonly-used prior for fluorescent imaging for a number of reasons: (1) fluorescent samples are generally sparsely labeled. (2) Even if a 2D slice of the sample is not spatially sparse, it will be sparse when considered with respect to our full 3D volume. (3) If native sparsity does not hold, images are generally sparse in gradient or wavelet domain. (4) Time-priors can further render a volume sparse by only considering temporally-varying information (i.e. neural firings). While it is an NP hard problem to generate a phase transition curve for our system as it requires running a large number of reconstructions of many different classes of objects at each sparsity level, we give an example of how our system performs at different sparsity levels by thresholding a 3D volume to generate different sparsity levels and reporting mean-squared error (MSE) and PSNR. The simulated volume is of a 3D zebrafish dataset. The simulations are done using our 3D shift-varying model and the experimental PSFs from our system. Figure 2.15 shows MSE and PSNR for the reconstructed volume at different sparsity levels (33%, original volume, to 0.2%, thresholded volume). As expected, our system performs better for sparser volumes. For denser volumes, our system recovers a lower-resolution version of the object and does not fail catastrophically.

2.13 Guide to Different Designs Using Our Theory

Our theory is general and enables other users to design their own optimized 3D microscope targeting different resolutions or volumes-of-interest. To do so, users should implement the following design process:

- For a target lateral resolution, determine the microlens' average clear aperture needed to support that resolution (Sec. *Lateral Resolution*). This also determines the number of microlenses in the phase mask.
- For a target depth range, distribute the focal lengths dioptrically across the depth range.
- Using our optimization criterion, optimize the microlenses positions and aberrations to further enhance the 3D performance.
- Fabricate the phase mask using our adaptive stitching algorithm with a Nanoscribe 3D printer.

2.14 Adaptive Stitching

The Nanoscribe 3D printer can only print across a field-of-view (FoV) of 350 μm, and so the 1.8 mm sized phase mask must be printed in multiple stitched blocks, with the mask translating between them. Due to the optical requirements on the microlenses, care needs to be taken when dividing the microlens array into blocks for printing with Nanoscribe. Our adaptive stitching approach aims to print each lens with minimal stitching artifacts. As the clear aperture for each lens is of the same order of magnitude as the maximum printing block size of Nanoscribe, each stitching block will correspond approximately to a single microlens. The center location of each microlens is known, so the problem reduces to dividing the plane in a number of regions, with each region attributed to one of the microlens centres. Preferably, the stitching lines should then fall in the overlapping region of two (or more) microlenses. We assume that such a division will result in the best possible optical quality.

This problem definition is quite similar to the basic Voronoi segmentation, where we are given a set of points in a plane and the task is to attribute each location in the plane to one of the given points. That problem is solved as follows. For each location in the plane, the distance to all centres is calculated. Attribution to one centre is then decided by it being the closest one (minimum search). As a result, a dividing line is defined by the fact that the distance to two or more centres is equal. The question now is, how can this be adapted to take into account finite shapes?

Rephrasing, we need to define a smooth function in the plane for each microlens followed by attributing locations to microlenses based on a (minimum) search over these different functions. To this end, we will use the height function for each microlens individually and then do a maximum search for the attribution. As a result, segmentation lines would fall exactly at those locations where the height of two or more microlenses are equal (see Fig 2.6). This is precisely what we want to achieve.

The resulting height-based segmentation is shown in Fig. 2.16. Here, different slices are shown (50 to 53 μm height). Colored regions need to be printed by Nanoscribe as a single FoV. The different colors correspond to different stitching blocks.

2.15 Discussion

Our device is designed with compressed 3D imaging and miniaturization in mind. For some 2D imaging applications where the loss of SNR (see Fig. 2.14) and lateral resolution (2.76 μm vs 2 μm) are acceptable, our device may have advantages over 2D Miniscope, due to its smaller size (17 mm vs. 23.5 mm tall) and weight (2.5 grams vs. 3 grams), or the ability to digitally refocus via 3D reconstruction. However, we expect that most applications of Miniscope3D will be for true 3D microscopy, so we mainly compare our specifications to MiniLFM, which is considered the gold-standard for single-shot miniature 3D fluorescence imaging.

Miniscope3D offers multiple improvements over MiniLFM. First, using multifocal microlenses (as opposed to unifocal in LFM) allows us to achieve better lateral resolution ($2.76 - 3.9$ μm) across a larger depth range (390 μm^3). In contrast, MiniLFM [96] demonstrated *best-case* lateral resolution of 6 μm at a particular depth and, while their resolution at other depths was not reported, we predict that their unifocal microlens design will result in lateral resolution that degrades significantly beyond 40 μm depth, based on previous analysis [17] and that in the *Multifocal Design* section below. We estimate that our Miniscope3D provides approximately 10× increase in the usable measurement volume over MiniLFM, with 2.2× better peak lateral resolution. Taken together, our Miniscope3D reconstructs approximately 50× more usable voxels than MiniLFM, significantly improving the utility of the device. This improved performance comes in a hardware package that is smaller than MiniLFM (17 mm tall vs. 26 mm tall) and lighter weight (2.5 grams vs. 4.7 grams), because we replace the heavy doublet tube lens and the microlens array assembly with a thin phase mask. This will be particularly valuable in head-mounted experiments with freely-moving animals.

Both our method and MiniLFM make sparsity assumptions on the sample in order to solve the inverse problem to recover a 3D volume from a 2D image. We require the sample to be sparse in some domain, meaning that there is some representation of the sample that has many zeros in its coefficients [19, 3]. Fluorescence imaging is a good candidate for such priors, since most biological samples are sparsely labelled. Because we optimize the microscope optics explicitly for single-shot 3D imaging, typical sparsity priors such as native sparsity, sparse 3D gradients (*Total Variation (TV)*, as used in this paper), or sparse wavelets work well in our system. The MiniLFM is designed specifically for neural activity tracking and so makes further structural and temporal sparsity assumptions, which improves their axial resolution from 30 μm (single-shot performance) to 15μm (temporal video processing performance). In contrast, our Miniscope3D achieves 15 μm single-shot axial resolution, across a large depth range, and could presumably improve upon that by incorporating temporal application-specific priors. In this paper, however, we aim to record highly dynamic samples (see supplementary videos) and so only impose sample sparsity. We demonstrate the generality of our approach experimentally with samples that exhibit different levels of sparsity (Fig. 2.2,2.3), achieving resolution sufficient for single-neuron imaging. As sparsity decreases, image quality and resolution degrade smoothly (see Fig. 2.15), roughly following previous theoretical analyses [19, 18, 3].

Scattering is a limitation for all single-photon microscopes, including ours. For applications such as neural imaging and studying the 3D motion of freely-swimming samples like C. elegans or tardigrades, the small amount of scattering should not hinder resolution. However, as the imaging depth within the scattering medium increases, we expect the resolution to degrade in a way similar to other single-photon microscopes. We show experimental reconstructions with and without scattering for the 100 μm thick scattering mouse brain tissue, and the 300 μm thick cleared brain tissue. Both reconstructions achieve single-neuron resolution.

Another limitation of our model is that it assumes no partial occlusions. This is a common limitation of 3D recovery methods in fluorescence microscopy (e.g. double helix [81], light field deconvolution microscopy [17], 3D localization microscopy) and generally works well in non-absorbing fluorescent samples. Modeling occlusions would be valuable in many practical situations, but remains a challenging problem.

Accessibility was a key consideration in our Miniscope3D design. By building on the popular open-source Miniscope platform, our method can be easily adopted into existing experimental pipelines. Any of the 450 labs currently using the 2D Miniscope can upgrade to our 3D prototype with minimal effort. Also, our method for 3D printing custom phase masks can enable others to fabricate their own mask designs tailored to particular applications. Because experimental results are in good agreement with our theoretical design and analysis, we are confident that our design theory can provide a useful framework for future customization of single-shot 3D systems.

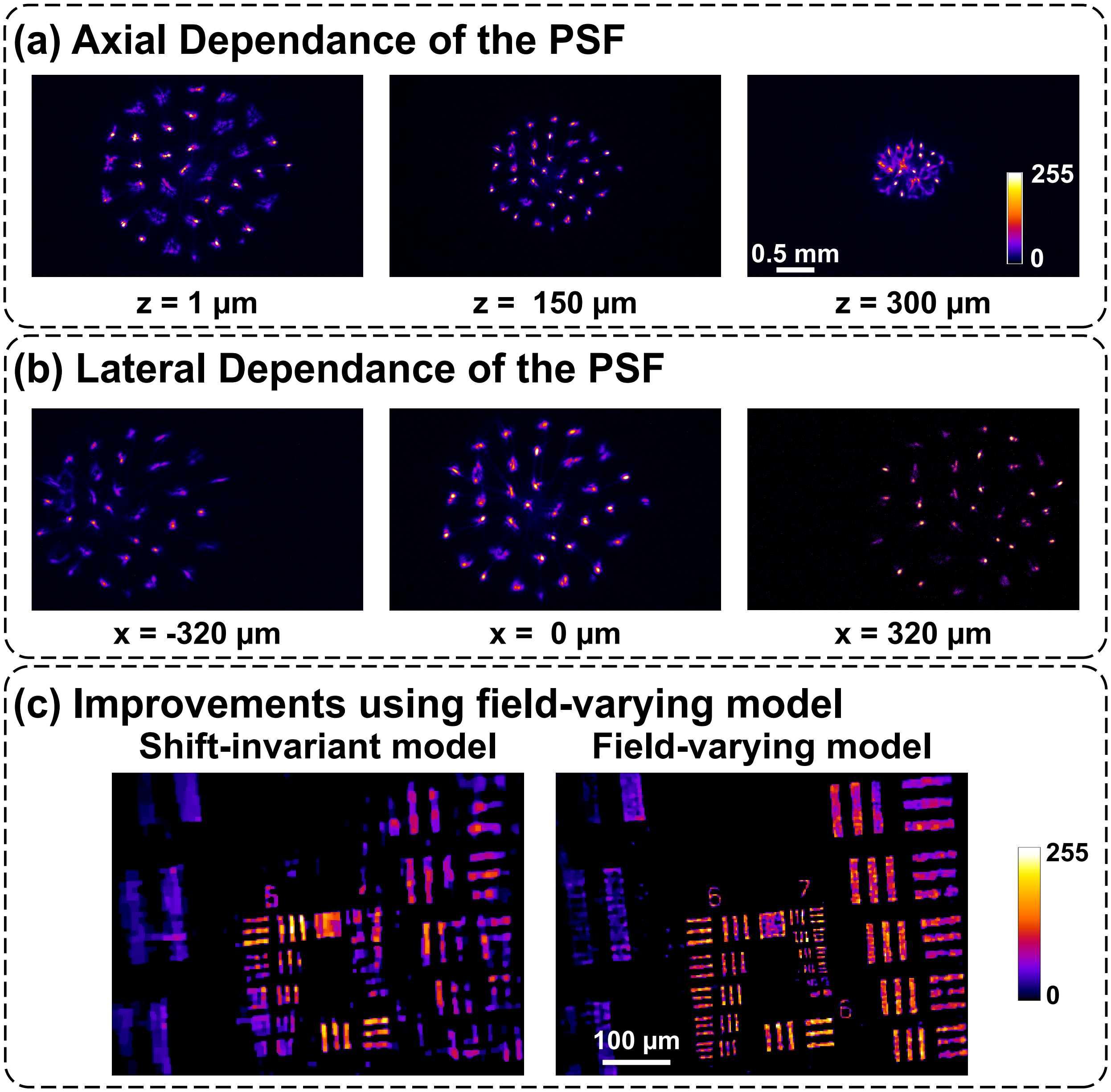

Figure 2.4: **Each 3D voxel maps to a different PSF:** (a) As a point source translates axially, the PSF scales and different spots come into focus. (b) As a point source translates laterally, the PSF shifts and incurs field-varying aberrations which destroy shift invariance. (c) When a shift-invariant approximation is made, reconstructions of a fluorescent resolution target (at $z = 250$ μm) display worse resolution (6.2 μm resolution) and more artifacts than when our field-varying model is used (2.76 μm resolution).

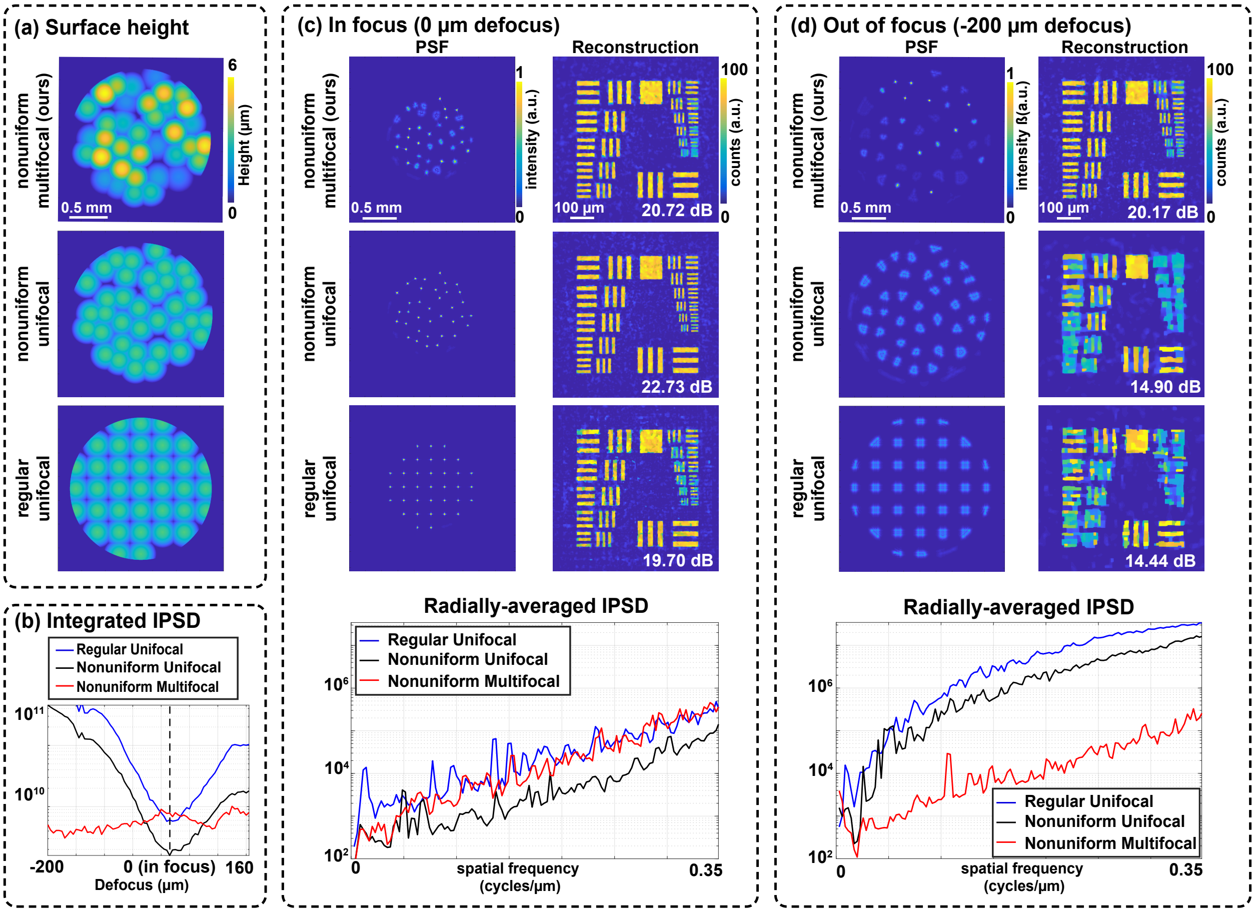

Figure 2.5: **Simulations to motivate our phase mask design, comparing our proposed nonuniform multifocal design with regular unifocal and nonuniform unifocal designs.** (a) Surface height profiles. (b) Sum of each design's PSF inverse power spectral density (IPSD) versus object depth (up to the designed cutoff frequency, lower is better). (c) PSFs and simulated reconstructions in-focus (at the unifocal arrays' native focus), with the reconstruction peak signal-to-noise ratio (PSNR) listed. The measurement is corrupted with 100 e^{-1} (peak) Poisson noise. In focus, the nonuniform unifocal design has slightly better PSNR and resolution than our design, and regular unifocal performs worse. The radially-averaged IPSD (lower is better) matches this trend. (d) Imaging 200μm off-focus, both unifocal designs produce blurry PSFs which result in significantly worse PSNR and resolution in the reconstruction, as compared to our design. This is also seen in the much higher inverse power spectra curves for unifocal designs.

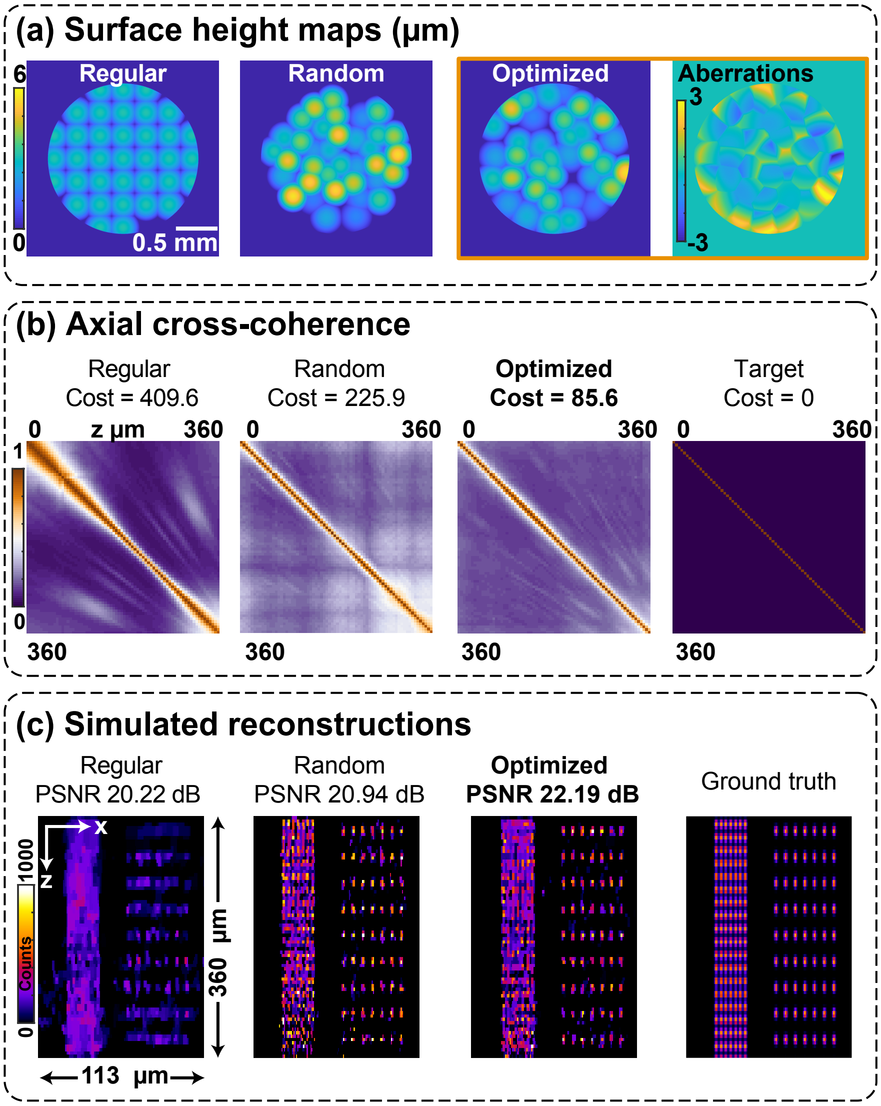

Figure 2.7: **Comparison of our optimized phase mask with random multifocal and regular microlens arrays:** (a) Phase mask surface height maps for all three cases, including the designed aberrations that were added in our optimized phase mask. (b) Axial cross-coherence matrices for all three cases: each entry is the maximum cross-correlation between the PSFs at the depths indicated by the row and column labels. The ideal system would be close to an identity matrix. (c) x-z slices from the 3D reconstructions of a test object consisting of differently-spaced point sources (x-spacings of 3.5 μm and 7 μm, z-spacings of 19.4 μm and 38 μm). We add Poisson noise with 1,000 peak counts to each measurement. Both nonuniform multifocal designs do significantly better than the regular unifocal array, and our optimized design performs slightly better than the random version, particularly near the edges of the depth range.

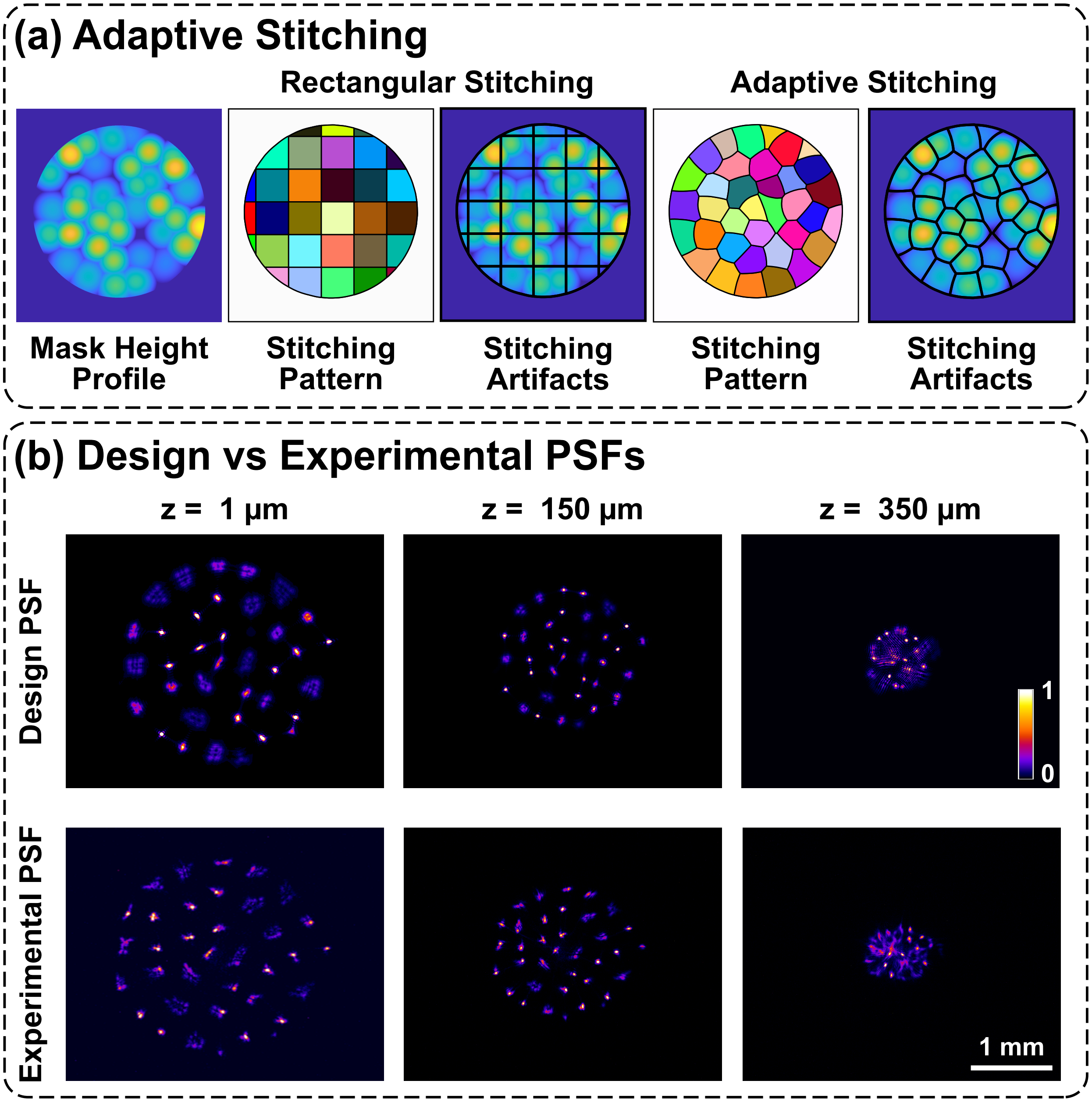

Figure 2.8: **Phase mask fabrication with Nanoscribe:** (a) Rectangular stitching leads to seams (black lines) going trough the many microlenses, while adaptive stitching puts the seams at the boundaries of the microlenses to mitigate artifacts. (b) Comparison between designed and experimental PSFs at a few sample depths, showing good agreement, with slight degradation at the edge of the volume.

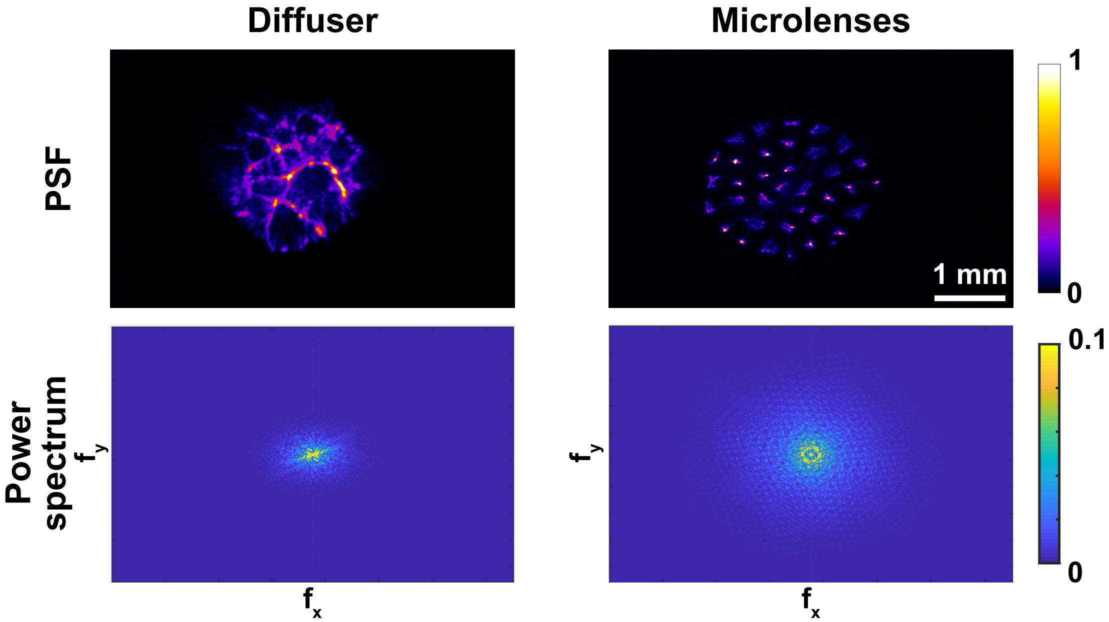

Figure 2.9: **Comparison of experimental PSFs resulting from a Gaussian diffuser and our microlens phase mask.** The microlenses generate PSFs with more high-frequency content, as seen in the power spectrum. The microlenses also have better light concentration; to achieve the same brightness as the microlenses PSF, the diffuser requires 4× the exposure time.

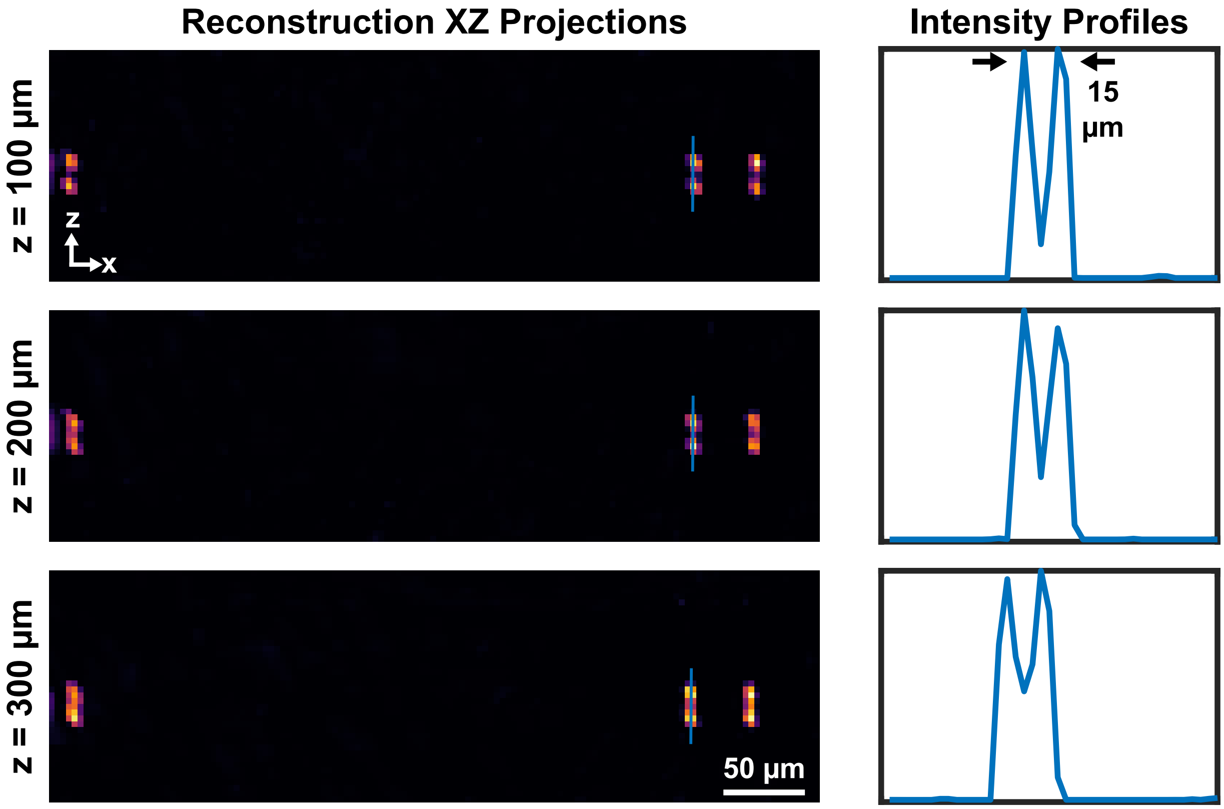

Figure 2.10: **Reconstructions results demonstrating** 15μ**m axial resolution across our depth range.** On left are x-z projections of the 3D reconstruction for the case of two layers of 3 beads each, separated by 15 μm axially. At right we show cross-cuts of the projections demonstrating clear resolving of the beads. The rows show results for placing the pairs of beads at different axial distances from the native focus plane.

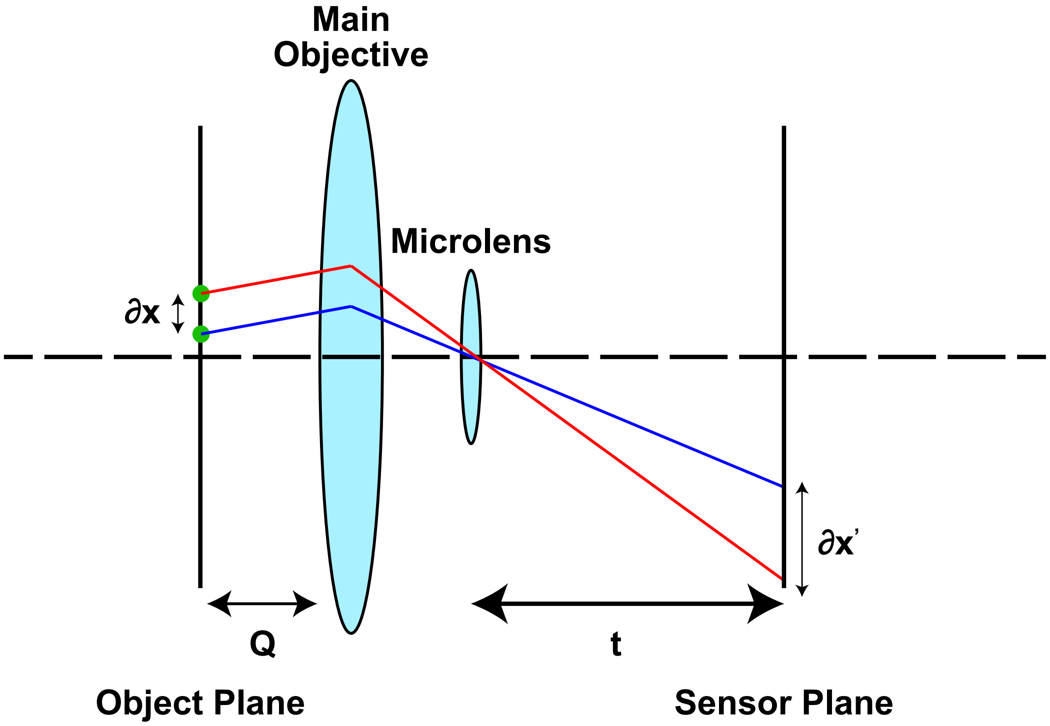

Figure 2.11: **Lateral resolution derivation.** Examining a single microlens placed immediately after the main objective.

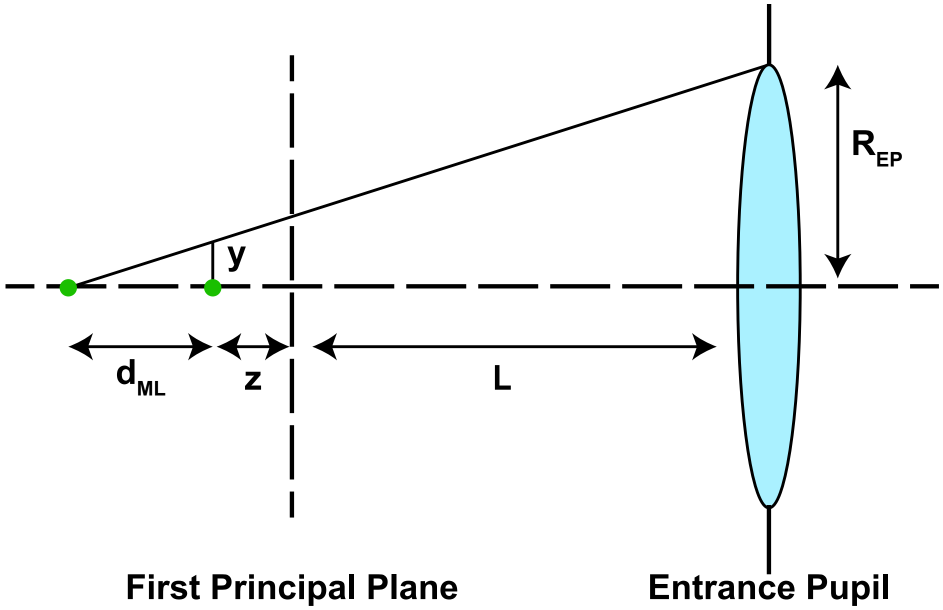

Figure 2.12: **Depth-of-focus** (DoF) derivation setup, with distance variables defined.

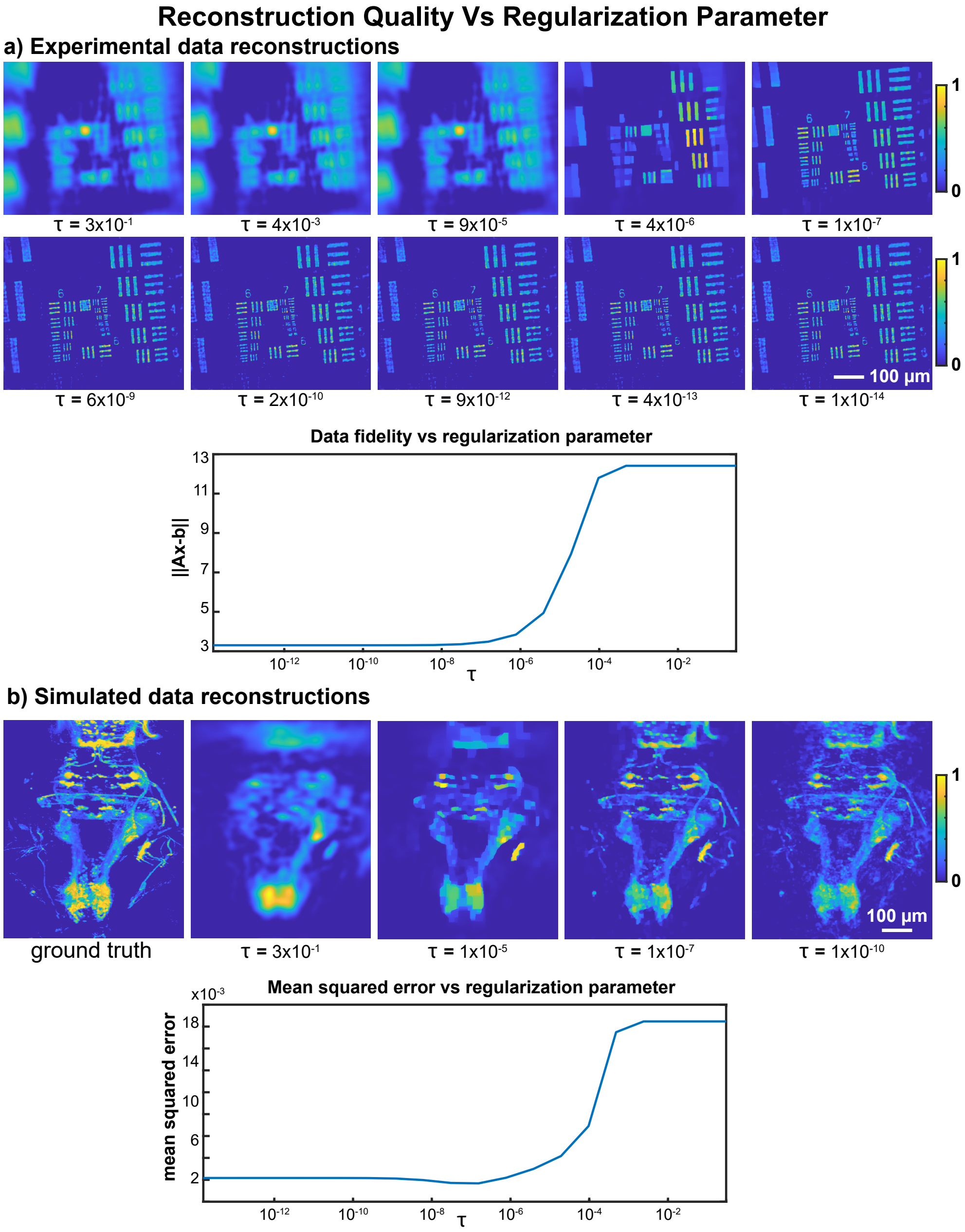

Figure 2.13: **Reconstruction quality as a function of regularization parameter, τ.** (a) Maximum intensity projections of an experimental volume reconstructed with different τ settings, along with a plot of the data fidelity term as a function of τ on a semi-log scale. (b) Maximum intensity projections of a simulated volume reconstructed with different τ settings, along with a plot of mean-squared error as a function of τ on a semi-log scale. The results demonstrate the stability of reconstructions for a large range of τ values.

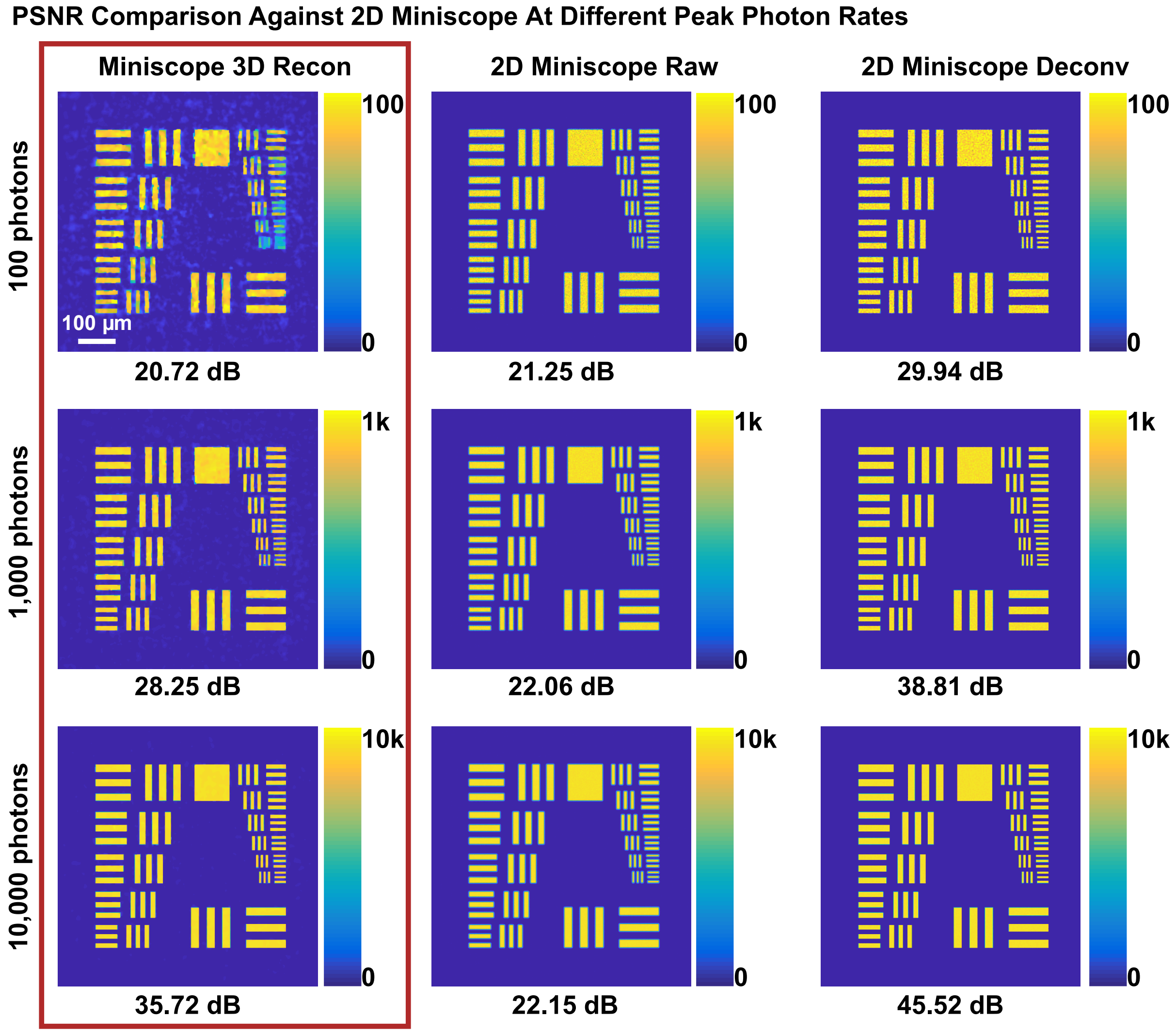

Figure 2.14: **PSNR comparison of Miniscope3D and 2D Miniscope.** (Left) Simulated reconstructions from our system at different light levels. (Middle) 2D Miniscope (simulated) raw measurement. (Right) 2D Miniscope deconvolved reconstructions. The multiplexing properties of our system that enable 3D capabilities result in a loss of PSNR.

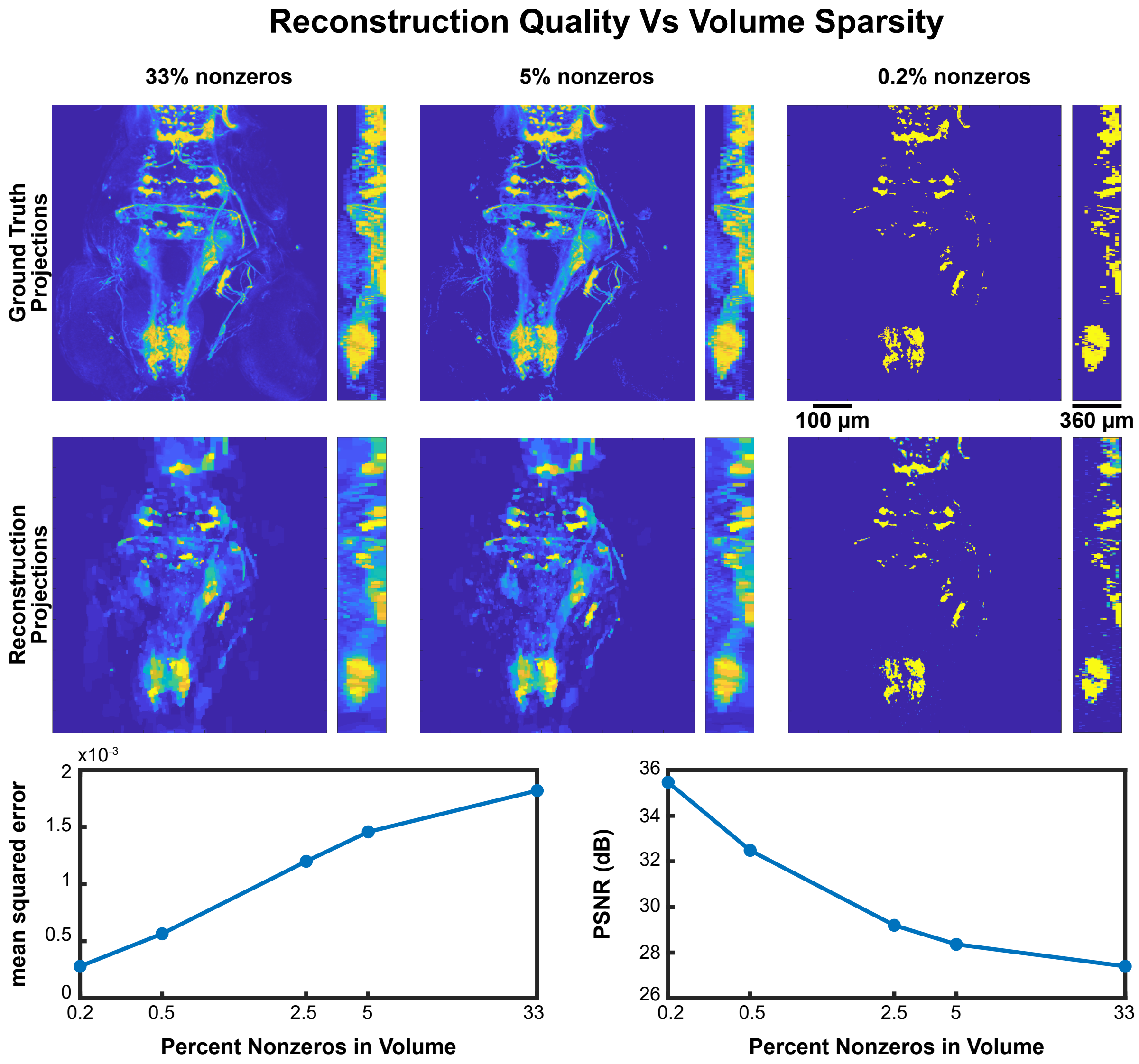

Figure 2.15: **Simulations of reconstruction quality at different sparsity levels.** Maximum intensity projections (y-x, z-x) show the quality of our reconstructions as compared to the ground truth at different sparsity levels. As the volume gets more dense, our reconstruction resolution degrades.

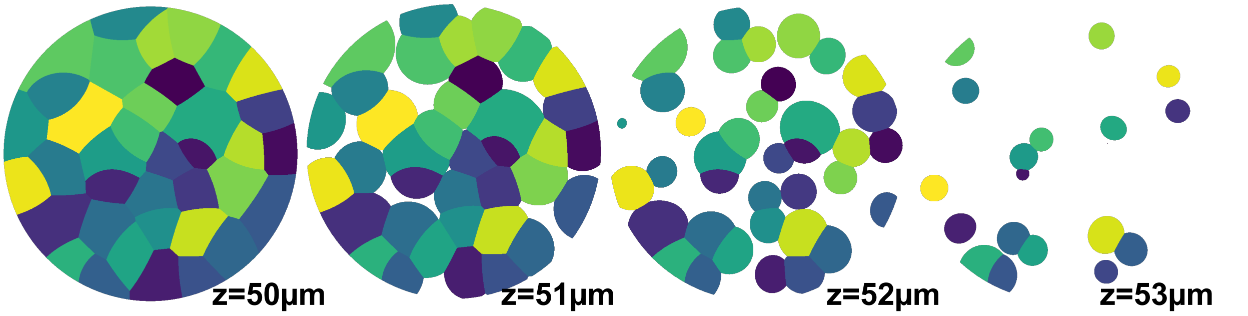

Figure 2.16: **Different slices** are shown, with different colors corresponding to different stitching blocks.

Chapter 3

Lensless snapshot hyperspectral imaging with a spectral filter array

This chapter is based on [77] and is joint work with Kristina Monakhova, Neerja Aggarwal, and Laura Waller.

3.1 Abstract

Hyperspectral imaging is useful for applications ranging from medical diagnostics to agricultural crop monitoring; however, traditional scanning hyperspectral imagers are prohibitively slow and expensive for widespread adoption. Snapshot techniques exist but are often confined to bulky benchtop setups or have low spatio-spectral resolution. In this paper, we propose a novel, compact, and inexpensive computational camera for snapshot hyperspectral imaging. Our system consists of a tiled spectral filter array placed directly on the image sensor and a diffuser placed close to the sensor. Each point in the world maps to a unique pseudorandom pattern on the spectral filter array, which encodes multiplexed spatio-spectral information. By solving a sparsity-constrained inverse problem, we recover the hyperspectral volume with sub-superpixel resolution. Our hyperspectral imaging framework is flexible and can be designed with contiguous or non-contiguous spectral filters that can be chosen for a given application. We provide theory for system design, demonstrate a prototype device, and present experimental results with high spatio-spectral resolution.

3.2 Introduction

Hyperspectral imaging systems aim to capture a 3D spatio-spectral cube containing spectral information for each spatial location. This enables the detection and classification of different material properties through spectral fingerprints, which cannot be seen with an RGB camera alone. Hyperspectral imaging has been shown to be useful for a variety of applications, from agricultural crop monitoring to medical diagnostics, microscopy, and food quality analysis [27, 53, 71, 97, 40, 2, 72, 79, 47, 9]. Despite the potential utility, commercial hyperspectral cameras range from $25,000 - $100,000 (at the time of publication of this paper). This high price point and the large size have limited the widespread use of hyperspectral imagers.

Traditional hyperspectral imagers rely on scanning either the spectral or spatial dimension of the hyperspectral cube with spectral filters or line-scanning [41, 35, 110]. These methods can be slow and generally require precise moving parts, increasing the camera complexity. More recently, snapshot techniques have emerged, enabling capture of the full hyperspectral data cube in a single shot. Some snapshot methods trade-off spatial resolution for spectral resolution by using a color filter array or splitting up the camera's field-of-view (FOV). Computational imaging approaches can circumvent this trade-off by spatio-spectrally encoding the incoming light, then solving a compressive sensing inverse problem to recover the spectral cube [103], assuming some structure in the scene. These systems are typically table-top instruments with bulky relay lenses, prisms, or diffractive elements, suitable for laboratory experiments, but not the real world. Recently, several compact snapshot hyperspectral imagers have been demonstrated that encode spatio-spectral information with a single optic, enabling a practical form factor [87, 34, 49]. Using a single optic to control both the spectral and spatial resolution, they are generally constrained to measuring contiguous spectral bins within a given spectral band.

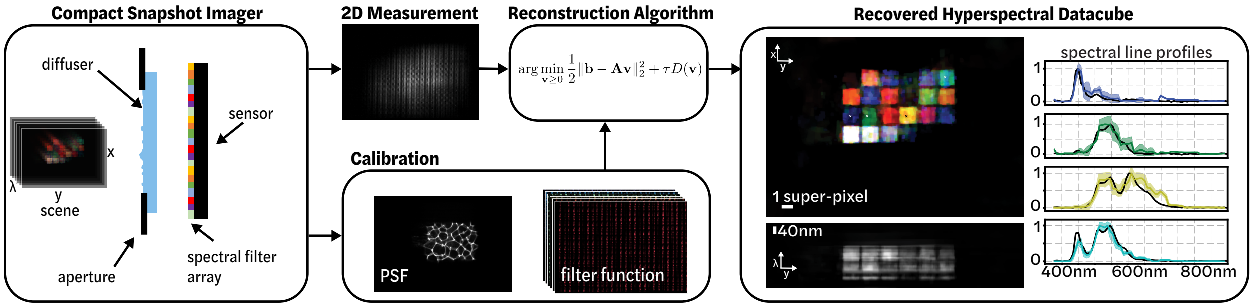

Figure 3.1: **Overview of the Spectral DiffuserCam imaging pipeline, which reconstructs a hyperspectral datacube from a single-shot 2D measurement.** The system consists of a diffuser and spectral filter array bonded to an image sensor. A one-time calibration procedure measures the point spread function (PSF) and filter function. Images are reconstructed using a non-linear inverse problem solver with a sparsity prior. The result is a 3D hyperspectral cube with 64 channels of spectral information for each of 448×320 spatial points, generated from a 2D sensor measurement that is 448×320 pixels.

Here, we propose a new encoding scheme that takes advantage of recent advances in patterned thin film spectral filters [91], and lensless imaging, to achieve high-resolution snapshot hyperspectral imaging in a small form factor. Our system consists of a tiled spectral filter array placed directly onto the sensor and a randomizing phase mask (i.e. diffuser) placed a small distance away from the sensor, as in the DiffuserCam architecture [3]. The diffuser spatially multiplexes the incoming light, such that each spatial point in the world maps to many pixels on the camera. The spectral filter array then spectrally encodes the incoming light via a structured erasure function. The multiplexing effect of the diffuser allows recovery of scene information from a subset of sensor pixels, so we are able to recover the full spatio-spectral cube without the loss in resolution that would result from using a non-multiplexing optic, such as a lens.

Our encoding scheme enables hyperspectral recovery in a compact and inexpensive form factor. The spectral filter array can be manufactured directly on the sensor, costing under $5 for both the diffuser and the filter array at scale. A key advantage of our system over previous compact snapshot hyperspectral imagers is that it decouples the spectral and spatial responses, enabling a flexible design in which either contiguous or non-contiguous spectral filters with user-selected bandwidths can be chosen. Given some conditions on scene sparsity and the diffuser randomness, the spectral sampling is determined by the spectral filters and the spatial resolution is determined by the autocorrelation of the diffuser response. This should find use in task-specific/classification applications [89, 23, 62, 46], where one may wish to tailor the spectral sampling to the application by measuring multiple non-contiguous spectral bands, or have higher-resolution spectral sampling for certain bands.

We present theory for our system, simulations to motivate the need for a diffuser, and experimental results from a prototype system. The main contributions of our paper are:

- A novel framework for snapshot hyperspectral imaging that combines compressive sensing with spectral filter arrays, enabling compact and inexpensive hyperspectral imaging.
- Theory and simulations analyzing the system's spatio-spectral resolution for objects with varying complexity.
- A prototype device demonstrating snapshot hyperspectral recovery on real data from natural scenes.

3.3 Related Work

Snapshot Hyperspectral Imaging

There have been a variety of snapshot hyperspectral imaging techniques proposed and evaluated over the past decades. Most approaches can be categorized into the following groups: spectral filter array methods, coded aperture methods, speckle-based methods, and dispersion-based methods.

Spectral filter array methods use tiled spectral filter arrays on the sensor to recover the spectral channels of interest [61]. These methods can be viewed as an extension of Bayer filters for RGB imaging, since each 'super-pixel' in the tiled array has a grid of spectral filters. As the number of filters increases, the spectral resolution increases and the spatial resolution decreases. For instance, with an 8×8 filter array (64 spectral channels), the spatial resolution is 8× worse in each direction than that of the camera sensor. Demosaicing methods have been proposed to improve upon this in post-processing; however, they rely on intelligently guessing information that is not recorded by the sensor [75]. Recently, photonic crystal slabs have been demonstrated for compact spectroscopy based on random spectral responses (as opposed to traditional passband responses) and extended to hyperspectral imaging through the tiling of the photonic crystal slab pixels [105, 106]. While these methods have high spectral accuracy, they have only been demonstrated in a 10×10 spatial pixel configuration. Our system uses a spectral filter array, but combines it with a randomizing diffuser in a lensless imaging architecture, allowing us to recover close to the full spatial resolution of the sensor, which is not possible with traditional lens-based methods. Our method uses traditional pass-band spectral filters, but could be extended to photonic crystal slabs and other spectral filter designs.

Coded aperture methods use a coded aperture, in combination with a dispersive optical element (e.g. a prism or diffractive grating), in order to modulate the light and encode spatial-spectral information [36, 65, 103, 20]. These systems are able to capture hyperspectral images and videos but tend to be large table-top systems consisting of multiple lenses and optical components. In contrast, our system has a much smaller form factor, requiring only a camera sensor with an attached spectral filter array and a thin diffuser placed close to the sensor.

Speckle-based methods use the wavelength dependence of speckle from a random media to achieve hyperspectral imaging. This has been demonstrated for compact spectrometers [84, 22] and extended to hyperspectral imaging [87, 34]. These systems can be compact, since they require only a sensor and scattering media as their optic; however their spectral resolution is limited by the speckle correlation through wavelengths. This is challenging to design for a given application, since the spatial and spectral resolutions are highly coupled. In contrast, our system uses spectral filters that can easily be adjusted for a given application and can be selected to have variable bandwidth or non-uniform spectral sampling.

Dispersive methods utilize the dispersion from a prism or diffractive optic to encode spectral information on the sensor. This can be accomplished opportunistically by a prism added to a standard DSLR camera [10]. The resulting system has high spatial resolution, equal to that of the camera sensor, but spectral information is encoded only at the edges of objects in the scene, resulting in a highly ill-conditioned problem and lower spectral accuracy. Other methods use a diffuser (as opposed to a prism) as the dispersive element [39]. This can be more compact than prism-based systems and can have improved spatial resolution when combined with an additional RGB camera [44]. To further improve compactness, [49] uses a single diffractive optic as both the lens and the dispersive element, uniquely encoding spectral information in a spectrally-rotating point spread function (PSF).

Our system uses a lensless architecture and a spectral filter array, together with sparsity assumptions, to reconstruct 3D hyperspectral information across 64 wavelengths. The design is most similar to [49] and achieves a similar compact size; however, our system achieves better spectral accuracy, and the use of the color filter array and diffuser results in more

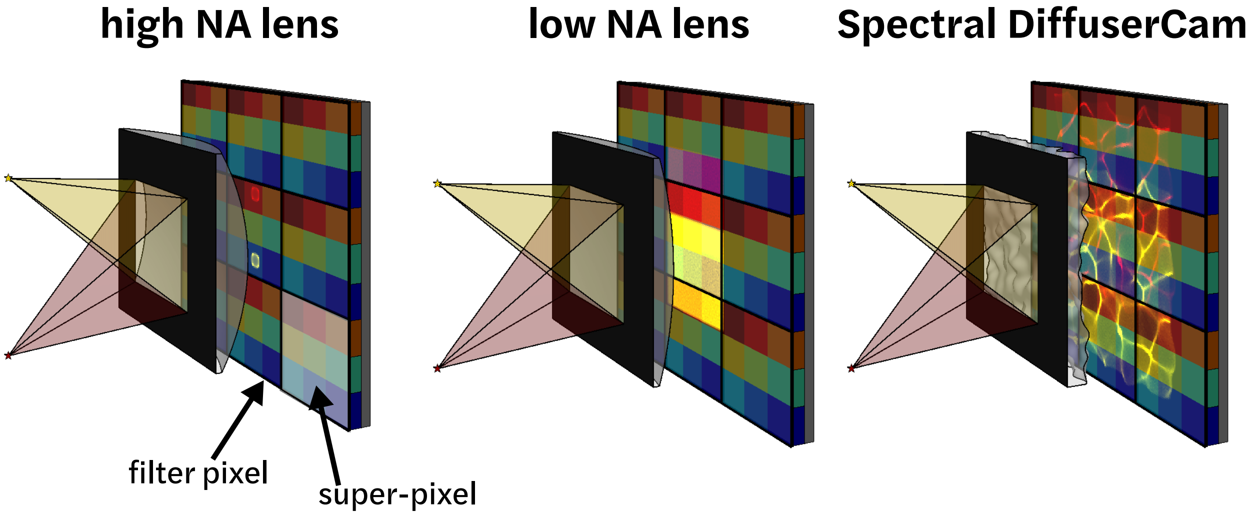

Figure 3.2: **Motivation for multiplexing:** A high-NA lens captures high-resolution spatial information, but misses the yellow point source, since it comes into focus on a spectral filter pixel designed for blue light. A low-NA lens blurs the image of each point source to be the size of the spectral filter's super-pixel, capturing accurate spectra at the cost of poor spatial resolution. Our DiffuserCam approach multiplexes the light from each point source across many super-pixels, enabling the computational recovery of both point sources and their spectra without sacrificing spatial resolution. Note that a simplified 3×3 filter array is shown here for clarity.

design flexibility, as our spectral and spatial resolutions are decoupled, enabling custom sensors tailored to specific spectral filter bands that do not need to be contiguous.

Lensless Imaging

Lensless, mask-based imaging systems do not have a main lens, but instead use an amplitude or phase mask in place of imaging optics. These systems have been demonstrated for very compact, small form factor 2D imaging [8, 58, 100, 99]. They are generally amenable to compressive imaging, due to the multiplexing nature of lensless architectures; each point in the scene maps to many pixels on the sensor, allowing a sparse scene to be completely recovered from a subset of sensor pixels [31]. Or, one can reconstruct higher-dimensional functions like 3D [3] or video [4] from a single 2D measurement. In this work, we use diffuser-based lensless imaging to spatially-multiplex light onto a repeated spectral filter array, then reconstruct 3D hyperspectral information. Because of the compressed sensing framework, our spatial resolution is better than the array super-pixel size, despite the missing information due to the array.

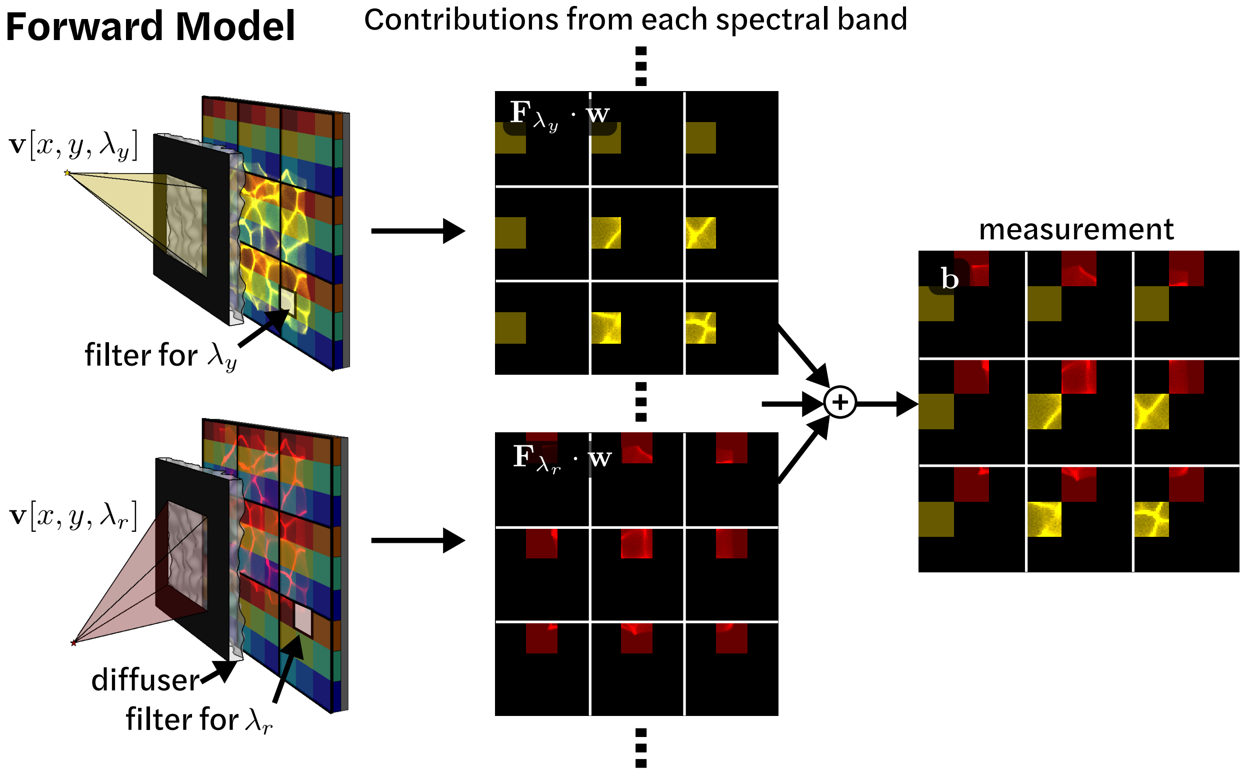
Forward Model
Contributions from each spectral band
$\mathbf{v}[x, y, \lambda_y]$
filter for λ_y
$\mathbf{F}_{\lambda_y} \cdot \mathbf{w}$
measurement
$\mathbf{b}$
$\mathbf{v}[x, y, \lambda_r]$
$\mathbf{F}_{\lambda_r} \cdot \mathbf{w}$
diffuser
filter for λ_r

Figure 3.3: **Image formation model** for a scene with two point sources of different colors, each with narrow-band irradiance centered at λ_y (yellow) and λ_r (red). The final measurement is the sum of the contributions from each individual spectral filter band in the array. Due to the spatial multiplexing of the lensless architecture, all scene points $v(x, y, \lambda)$ project information to multiple spectral filters, which is why we can recover a high-resolution hyperspectral cube from a single image, after solving an inverse problem.

3.4 System Design Overview

Our system leverages recent advances in both spectral filter array technology and compressive lensless imaging to decouple the spectral and spatial design. Furthermore, the spectral filter arrays can be deposited directly on the camera sensor. With a diffuser as our multiplexing optic, the system is compact and inexpensive at scale.

To motivate our need for a multiplexing optic instead of an imaging lens, let us consider three candidate architectures: one with a high numerical aperture (NA) lens whose diffraction-limited spot size is matched to the filter pixel size, one with a low-NA lens whose diffraction-limited spot size is matched to the super-pixel size, and finally our design with a diffuser as a multiplexing optic. Figure 3.2 illustrates these three scenarios with a simplified example of a spectral filter array consisting of 3 × 3 spectral filters (9 total) repeated horizontally and vertically. Assume that the monochrome camera sensor has square pixels of lateral size N_{pixel}, the spectral filter array has square filters of size N_{filter}, and each 3×3 block of spectral filters creates a *super-pixel* of size $N_{\text{super-pixel}}$, where $N_{\text{pixel}} < N_{\text{filter}} < N_{\text{super-pixel}}$.

In the high-NA lens case, a point source in the scene will be imaged onto a single filter pixel of the sensor, and thus will only be measured if it is within the passband of that filter; otherwise it will not be recorded, Fig. 3.2 (left). In the low-NA lens case, each point source will be imaged to an area the size of the filter array super-pixel, and thus recorded by the sensor correctly, but at the price of low spatial-resolution (matched to the the super-pixel size), Fig. 3.2 (middle). In contrast, a multiplexing optic can avoid the gaps in the measurement of the high-NA lens and achieve better resolution than the low-NA case.

A diffuser multiplexes the light from each point source such that it hits many filter pixels, covering all of the spectral bands. And the spatial resolution of the final image can be on the order of the camera pixel size, provided that conditions for compressed sensing are met, Fig. 3.2 (right). In practice, the spatial resolution of our system will be bounded by the autocorrelation of the point spread function (PSF), as detailed in Sec. 3.8, and the diffuser PSF must span multiple super-pixels to ensure that each point in the world is captured. Since compressive recovery is used to recover a 3D hyperspectral cube from a 2D measurement, the resolution is a function of the scene complexity, as described in Sec. 3.8.

3.5 Imaging Forward Model

Given our design with a diffuser placed in front of a sensor that has a spectral filter array on top of it, in this section we outline a forward model for the optical system, illustrated in Fig. 3.3. This model is a critical piece of our iterative inverse algorithm for hyperspectral reconstruction and will also be used to analyze spatial and spectral resolution.

Spectral filter model

The spectral filter array is placed on top of an imaging sensor, such that the exposure on each pixel is the sum of point-wise multiplications with the discrete filter function,

$$\mathbf{L}[x, y] = \sum_{\lambda=0}^{K-1} \mathbf{F}_{\lambda}[x, y] \cdot \mathbf{v}[x, y, \lambda], \tag{3.1}$$

where $\cdot$ denotes point-wise multiplication, $\mathbf{v}[x, y, \lambda]$ is the spectral irradiance incident on the filter array and $\mathbf{F}_{\lambda}[x, y]$ is a 3D function describing the transmittance of light through the spectral filter for K wavelength bands, which we call the *filter function.* In this model, we absorb the sensor's spectral response into the definition of $\mathbf{F}_{\lambda}[x, y]$. Our device's filter function is determined experimentally (see Sec 3.7.C) and shown in Fig. 3.4(b). This can be generalized to any arbitrary spectral filter design and does not assume alignment between the filter pixels and the sensor pixels. Here, we focus on the case of a repeating grid of spectral filters, where each 'super-pixel' consists of a set of narrow-band filters. Our device has a 8×8 grid of filters in each super-pixel; Fig. 3.3 illustrates a simplified 3×3 grid, for clarity.

Diffuser model

The diffuser (a smooth pseudorandom phase optic) in our system achieves spatial multiplexing; this results in a compact form factor and enables reconstruction with spatial resolution better than the super-pixel size via compressed sensing. The diffuser is placed a small distance away from the sensor and an aperture is placed on the diffuser to limit higher angles. The sensor plane intensity resulting from the diffuser can be modeled as a convolution of the scene, $\mathbf{v}[x, y, \lambda]$ with the on-axis PSF, $\mathbf{h}[x, y]$ [58]:

$$\mathbf{w}[x, y, \lambda] = \text{crop}\Big(\mathbf{v}[x, y, \lambda] \overset{[x,y]}{*} \mathbf{h}[x, y])\Big) \tag{3.2}$$

where $\overset{[x,y]}{*}$ represents a discrete 2D linear convolution over spatial dimensions. The crop function accounts for the finite sensor size. We assume that the PSF does not vary with wavelength and validate this experimentally in Sec. 3.7.B. However, this model can be easily extended to include a spectrally-varying PSF, $\mathbf{h}[x, y, \lambda]$ if there is more dispersion across wavelengths.

We assume that objects are placed beyond the hyperfocal distance of the imager, therefore the PSF has negligible depth-variance and a 2D convolutional model is valid [58]. If objects are placed within the hyperfocal distance, a 3D model will be needed to account for the depth-variance of the PSF.

Combined model

Combining the spectral filter model with the diffuser model, we have the following discrete forward model:

$$\begin{aligned} \mathbf{b} &= \sum_{\lambda=0}^{K-1} \mathbf{F}_\lambda[x, y] \cdot \text{crop}\Big(\mathbf{h}[x, y] \overset{[x,y]}{*} \mathbf{v}[x, y, \lambda]\Big) && (3.3) \\ &= \sum_{\lambda=0}^{K-1} \mathbf{F}_\lambda[x, y] \cdot \mathbf{w}[x, y, \lambda] && (3.4) \\ &= \mathbf{A}\mathbf{v}. && (3.5) \end{aligned}$$

The linear forward model is represented by the combined operations in matrix $\mathbf{A}$. Figure 3.3 illustrates the forward model for several point sources, showing the intermediate variable $\mathbf{w}[x, y, \lambda]$, which is the scene convolved with the PSF, before point-wise multiplication by the filter function. The final image is the sum over all wavelengths.

3.6 Hyperspectral Reconstruction

To recover the hyperspectral datacube from the 2D measurement, we must solve an underdetermined inverse problem. Since our system falls within the framework of compressive sensing due to our incoherent, multiplexed measurement, we use l_1 minimization. We use a weighted 3D total variation (3DTV) prior on the scene, as well as a non-negativity constraint, and a low-rank prior on the spectrum. This can be written as:

$$\hat{\mathbf{v}} = \underset{\mathbf{v} \geq 0}{\arg\min} \frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{v}\|_2^2 + \tau_1 \|\nabla_{xy\lambda}\mathbf{v}\|_1 + \tau_2 \|\mathbf{v}\|_*, \tag{3.6}$$

where $\nabla_{xy\lambda} = [\nabla_x \nabla_y \nabla_\lambda]^T$ is the matrix of forward finite differences in the x, y, and λ directions, $\|\cdot\|_*$ represents the nuclear norm, which is the sum of singular values. τ_1 and τ_2 are the tuning parameters for the 3DTV prior and low-rank priors, respectively. We use the fast iterative shrinkage-thresholding algorithm (FISTA) [12] with weighted anisotropic 3DTV to solve this problem according to [51].

3.7 Implementation Details

We built a prototype system using a CMOS sensor, a hyperspectral filter array provided by Viavi Solutions (Santa Rosa, CA)[91], and an off-the-shelf diffuser (Luminit 0.5°) placed 1cm away from the sensor. The sensor has 659×494 pixels (with a pixel pitch of $9.9\mu m$), which we crop down to 448×320 to match the spectral filter array size. The spectral filter array consists of a grid of 28×20 super-pixels, each with an 8×8 grid of filter pixels (64 total, spanning the range 386-898nm). Each filter pixel is $20\mu m$ in size, covering slightly more than 4 sensor pixels. The alignment between the sensor pixels and the filter pixels is unknown, requiring a calibration procedure (detailed in Sec. 3.73.7). The exposure time is adjusted for each image, ranging from 1ms-13ms, which is short enough for video-rate acquisition. The computational reconstruction typically takes 12-24 minutes (for 500-1000 iterations) on an RTX 2080-Ti GPU using MATLAB.

Filter Function Calibration

To calibrate the filter function ($\mathbf{F}_\lambda[x, y]$ in Eqn. 3.3), including the spectral sensitivity of both the sensor and the spectral filter array, we use a Cornerstone 130 1/3m motorized monochromator (Model 74004). The monochromater creates a narrow-band source of 5nm full-width at half-maximum (FWHM) and we measure the filter response (without the diffuser) while sweeping the source by 8nm increments from 386nm to 898nm. The result is shown in Fig. 3.4(b).

PSF Calibration

We also need to calibrate the diffuser response by measuring the diffuser PSF pattern without the spectral filter array. Because the diffuser is relatively smooth with large features (relative to the wavelength of light), the PSF remains relatively constant as a function of wavelength, as shown in Fig. 3.4(a). Hence, we only need to calibrate for a single wavelength by capturing a single point source calibration image [3]. However, this is not trivial because the spectral filter array is bonded to the sensor and cannot be removed easily. In our setup, we instead take advantage of the fact that our filter array is smaller than our sensor, so we can measure the PSF using the edges of the raw sensor, by shifting the point source to scan the different parts of the PSF over the raw sensor area and stitching the sub-images together. In a system where the filter size is matched to the sensor, this trick will not be possible, but an optimization-based approach could be developed to recover the PSF from measurements.

System Non-idealities

Our reconstruction quality and spectral resolution are limited by two non-idealities in our system. First, our camera development board performs an undefined and uncontrollable non-linear contrast-stretching to all images. This makes the measurement non-linear and impedes our imaging of dim objects (since the camera performs a larger contrast stretching for dimmer images). Further, our spectral calibration may have errors, since each calibration image cannot be normalized by the intensity of light hitting the sensor. This may cause certain wavelength bands to appear brighter or dimmer than they should in our spectral reconstructions. A better camera board without automatic contrast stretching should fix this problem and provide more quantitative spectral profile reconstructions in the future.

Second, we used a simplified spectral calibration in which we measured the response with uniform spectral sampling, instead of at the true wavelengths of the filters. Due to the mismatch between our calibration scheme (measured every 8nm with constant bandwidth) and the actual spectral filters (center wavelengths spaced 5-12nm apart with bandwidths between 6-23nm), sometimes our calibration wavelengths fall between two filters, resulting in an ambiguity. Given this non-ideal calibration, our effective spectral bands are limited to 49 bands, instead of 64. In our results, we show all 64 bands, but note that some will have overlapping spectral responses. In the future, we will calibrate at the design wavelengths of the filter to fix this issue. Further, the deposition of the spectral filters directly on-top of the camera pixels (requiring precise placement during the manufacturing stage) would alleviate the need for this calibration entirely.

3.8 Resolution Analysis

Here, we derive our theoretical resolution and experimentally validate it with our prototype system. First, we discuss spectral resolution, which is set by the filter bandwidths, and then we compute the expected two-point spatial resolution, based on the PSF autocorrelation.

Since our resolution is scene-dependent, we expect the resolution to degrade with scene complexity. To characterize this, we present theory for multi-point resolution based on the condition number analysis introduced in [3]. We compare our system against those with a high-NA and low-NA lens instead of a diffuser. Our results demonstrate two-point spatial resolution of ~0.19 super-pixels and multi-point spatial resolution of ~0.3 super-pixels for 64 spectral channels ranging from 386-898nm.

Spectral Resolution

Spectral resolution is determined by the spectral channels of the filter array. As such, we expect to be able to resolve the 64 spectral channels present in our spectral filter array. The filters have an average spacing of 8nm across a 386-898nm range with bandwidths between 6-23nm. To validate our spectral resolution, we scan a point source across those wavelengths using a monochrometer. Figure 3.6 shows a sampling of spectral reconstructions overlaid on top of each other, with the shaded blocks indicating the ground-truth monochrometer spectra. Our reconstructions all match the ground-truth peaks within 5nm of the true wavelength. The small red peaks around 400nm are artifacts from the monochrometer, which emitted a 2nd peak around 400nm for the longer wavelengths.

Two-point Spatial Resolution

Spatial resolution of our system, in terms of the two-point resolution, will be bounded by that of a lensless imager with the diffuser only (without the spectral filter array). The expected resolution can be defined as the autocorrelation peak half-width at 70% the maximum value [58], Fig. 3.5(a). For our system, this is ~3 sensor pixels, or 0.19 super-pixels. To experimentally measure the spatial resolution of our system, we image two point sources at three different wavelengths (618 nm, 522 nm, 466 nm). The reconstructions in Fig. 3.5 show that we can resolve two point sources that are 0.19 super-pixels apart for each wavelength and orientation, as determined by applying the Rayleigh criterion. This demonstrates that our system achieves sub-super-pixel spatial resolution, consistent with the expected resolution that would be achieved without the spectral filter array.

Multi-point resolution

Because our image reconstruction algorithm contains nonlinear regularization terms, our reconstruction resolution will be object dependent. Hence, two-point resolution measurements are not sufficient for fully characterizing the system resolution, and should be considered a *best case* scenario. To better predict real-world performance, we perform a local condition number analysis, as introduced in [3], that estimates resolution as a function of object complexity. The local condition number is a proxy for how well the forward model can be inverted, given known support, and is useful for systems such as ours in which the full **A** matrix is never explicitly calculated [18].

The local condition number theory states that given knowledge of the *a priori* support of the scene, **v**, we can form a sub-matrix consisting only of columns of **A** corresponding to the non-zero voxels. The reconstruction problem will be ill-posed if any of the sub-matrices of **A** are ill-conditioned, which can be quantified by the condition number of the sub-matrices. The worst-case condition number will be when sources are near each other, therefore we compute the condition number for a group of point sources with a separation varying by an integer number of voxels and repeat this for increasing numbers of point sources.

In Fig. 3.7, we calculate the local condition number for two cases: the 2D spatial reconstruction case, considering only a single spectral channel, and the 3D case, considering points with varying spatial and spectral positions. For comparison, we also simulate the condition number for a low-NA and high-NA lens, as introduced in Sec. 3.4. The results show that our diffuser design has a consistently lower condition number than either the low- or high-NA lens, having a condition number below 40 for separation distances of greater than ~0.3 super-pixels. The low-NA lens needs a separation distance closer to ~1 super-pixel, as expected, and the high-NA lens has an erratic condition number due to the missing information in the measurement.

From this analysis, we can see that, beyond 0.3 super-pixels separation, the condition number for the diffuser does not get arbitrarily worse for increasing scene complexity. Thus, our expected spatial resolution is approximately 0.3 super-pixels.

Simulated Resolution Target Reconstruction

Next, we validate the results of our condition number analysis through simulated reconstructions of a resolution target with different spatial locations illuminated by different sources (red, green, blue and white light), as shown in Fig. 3.8. For each simulation, we add Gaussian noise with a variance of 1×10^{-5} and run the reconstruction for 2,000 iterations of FISTA with 3DTV. Our system resolves features that are 0.3 super-pixels apart, whereas the low-NA lens can only resolve features that are roughly 1 super-pixel apart and the high-NA lens results in gaps, validating our predicted performance.

3.9 Experimental Results

We start with experimental reconstructions of simple objects with known properties - a broadband USAF resolution target displayed on a computer monitor, and a grid of RGB LEDs (Fig. 3.9). We resolve points that are ~.3 super-pixels apart, which matches our expected multi-point resolution based on the condition number analysis above. For the RGB LED scene, the ground truth spectral profiles of the LEDs are measured using a spectrometer, and our recovered spectral profile closely matches the ground truth, as shown in Fig. 3.9(b).

Next, we show reconstructions of more complex objects, either displayed on a computer monitor or illuminated with two halogen lamps (Figure 3.10). We plot the ground truth spectral line profiles, as measured by a Thorlabs CCS200 spectrometer, from four points in the scene, showing that we can accurately recover the spectra. A reference RGB scene is shown for each image, demonstrating that the reconstructions spatially match the expected scene.

3.10 Discussion

A key advantage of our design over previous work is its flexibility to choose the spectral filters in order to tailor the system to a specific application. For example, one can non-linearly sample a wide range of wavelengths (which is difficult with many previous snapshot hyperspectral imagers). In future, we plan to design implementations specific to various task-based applications, which could make hyperspectral imaging more easily adopted, especially since the price is several orders-of-magnitude lower than currently available hyperspectral cameras.

Currently, we experimentally achieve a spatial resolution of ~0.3 super-pixels, or 5 sensor pixels. In future designs, we should be able to achieve the full sensor resolution (along with better quality reconstructions) by optimizing the randomizing optic, instead of using an off-the shelf diffuser. This could be achieved by end-to-end optical design [95, 82].

Our system has two main limitations: light-throughput and scene-dependence. Due to the use of narrow-band spectral filters, much of the light is filtered out by the filters. This provides good spectral accuracy and discrimination, but at the cost of low light throughput. In addition, since the light is spread by the diffuser over many pixels, the signal-to-noise ratio (SNR) is further decreased. Hence, our imager is not currently suitable for low-light conditions. This light-throughput limitation can be mitigated in the future by the use of photonic crystal slabs instead of narrowband filters, in order to increase light-throughput while maintaining spatio-spectral resolution and accuracy [106]. In addition, end-to-end design of both the spectral filters and the phase mask should improve efficiency, since application-specific designs can use only the set of wavelengths necessary for a particular task, without sampling the in-between wavelengths. Reducing the number of spectral bands improves both light throughput (because more sensor area will be dedicated to each spectral band) and spatial resolution (because the super-pixels will be smaller).

Our second limitation is scene-dependence, as our reconstruction algorithm relies on object sparsity (e.g. sparse gradients). Because of the non-linear regularization term, it is difficult to predict performance, and one might suffer artifacts if the scene is not sufficiently sparse. Recent advances in machine learning and inverse problems seek to provide better signal representations, enabling the reconstruction of more complicated, denser scenes [68, 15]. In addition, machine learning could be useful in speeding up the reconstruction algorithm [76] as well as potentially utilizing the imager more directly for a higher-level task, such as classification [29].

3.11 Conclusion

Our work presents a new hyperspectral imaging modality that combines a color filter array and lensless imaging techniques for an ultra-compact and inexpensive hyperspectral camera. The spectral filter array encodes spectral information onto the sensor and the diffuser multiplexes the incoming light such that each point in the world maps to many spectral filters. The multiplexed nature of the measurement allows us to use compressive sensing to reconstruct high spatio-spectral resolution from a single 2D measurement. We provided an analysis for the expected resolution of our imager and experimentally characterized the two-point and multi-point resolution of the system. Finally, we built a prototype and demonstrated reconstructions of complex spatio-spectral scenes, achieving up to 0.19 super-pixel spatial resolution across 64 spectral bands.

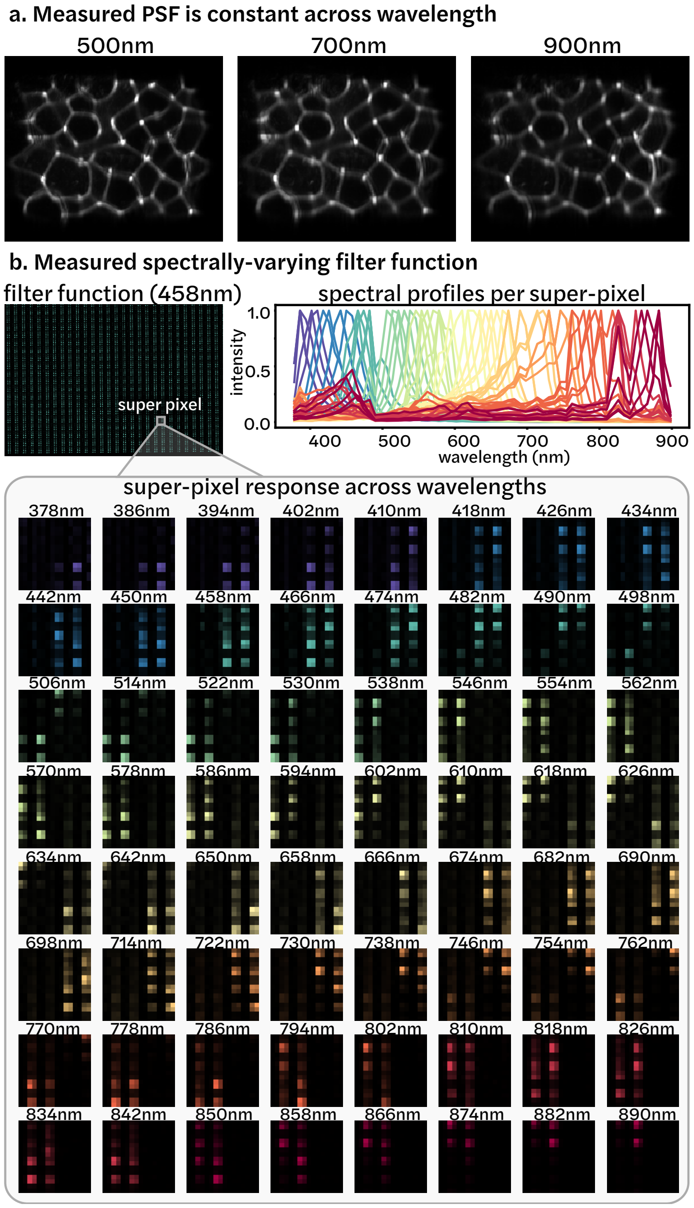

Figure 3.4: **Experimental calibration of Spectral DiffuserCam.** (a) The caustic PSF (contrast-stretched and cropped), before passing through the spectral filter array, is similar at all wavelengths. (b) The spectral response with the filter array only (no diffuser). (Top left) Full measurement with illumination by a 458nm plane wave. The filter array consists of 8×8 grids of spectral filters repeating in 28×20 super-pixels. (Top right) Spectral responses of each of the 64 color channels. (Bottom) Spectral response of a single super-pixel as illumination wavelength is varied with a monochromater.

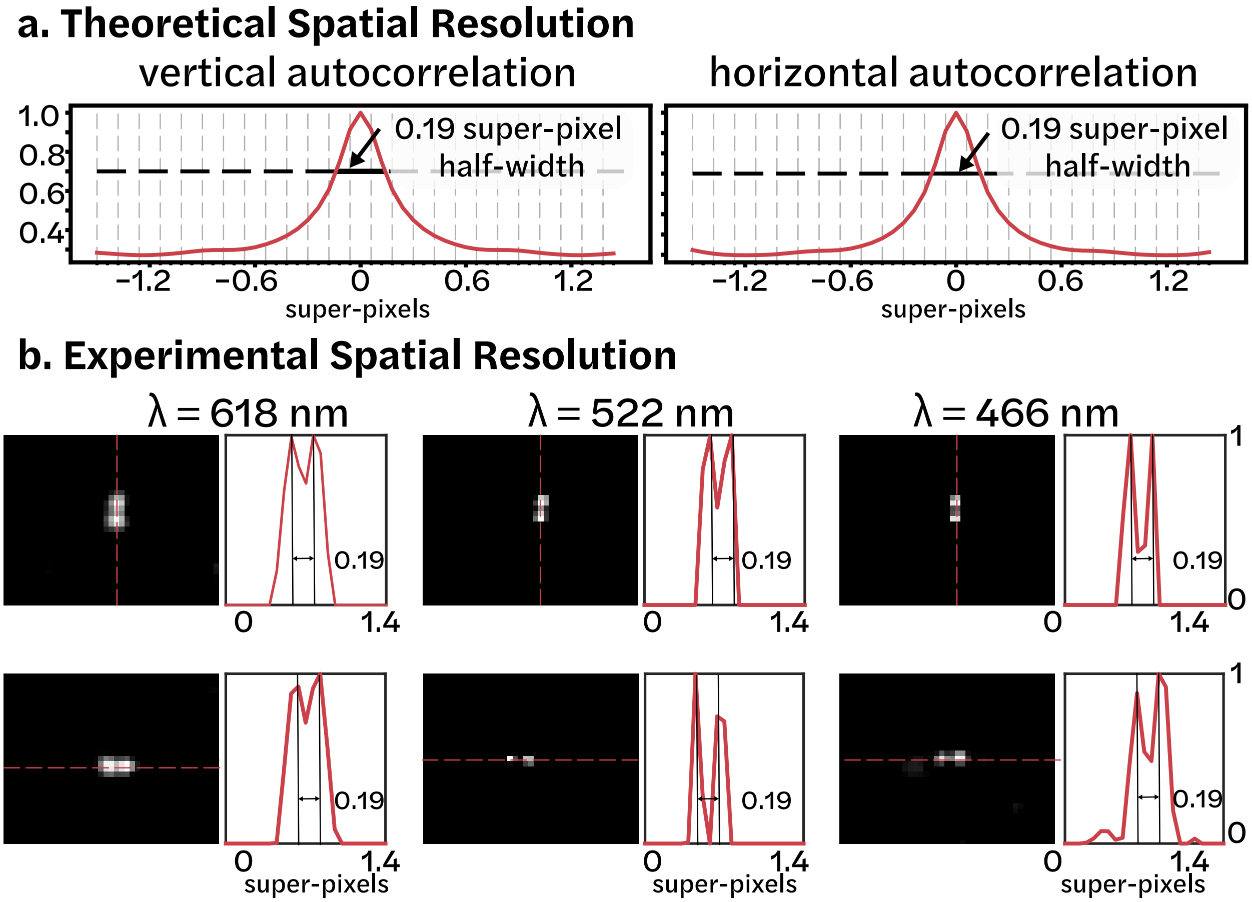

Figure 3.5: **Spatial Resolution analysis.** (a) The theoretical resolution of our system, defined as the half-width of the autocorrelation peak at 70% its maximum value, is 0.19 super-pixels. (b) Experimental two-point reconstructions demonstrate 0.19 super-pixel resolution across all wavelengths (slices of the reconstruction shown here), matching the theoretical resolution.

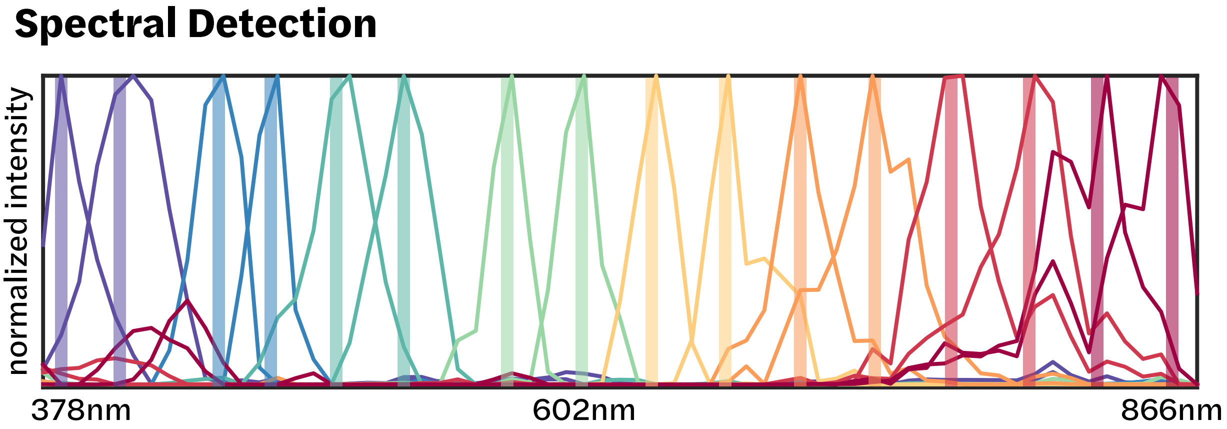

Figure 3.6: **Spectral resolution analysis.** Sample spectra from hyperspectral reconstructions of narrow-band point sources, overlaid on top of each other, with shaded lines indicating the ground-truth. For each case, the recovered spectral peak matches the true wavelength within 5nm.

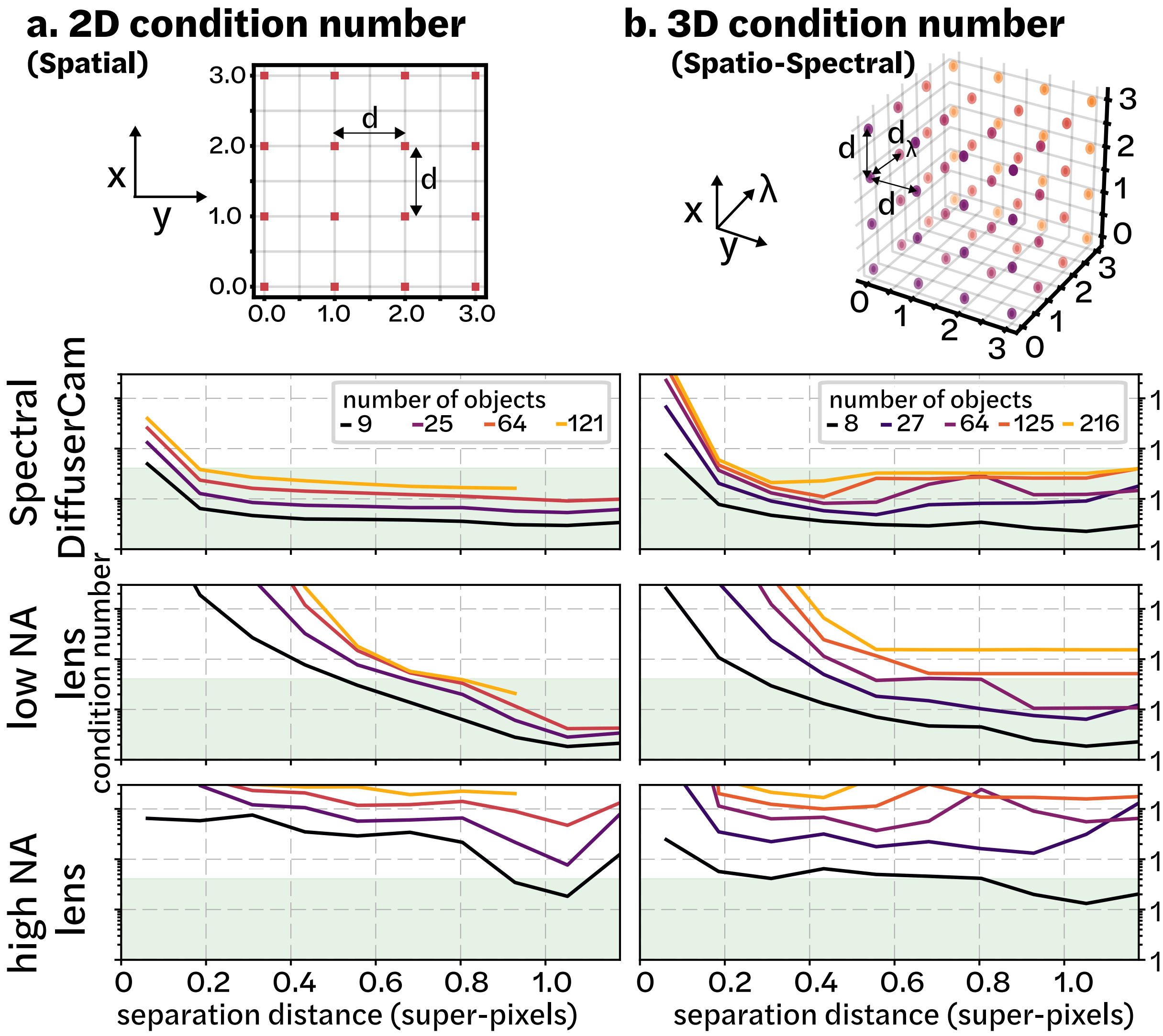

Figure 3.7: **Condition number analysis for Spectral DiffuserCam, as compared to a low-NA or high-NA lens.** (a) Condition numbers for the 2D spatial case (single spectral channel) are calculated by generating different numbers of points on a 2D grid, each with separation distance *d*. (b) Condition numbers for the full spatio-spectral case are calculated on a 3D grid. A condition number below 40 is considered to be good (shown in green). The diffuser has a consistently better performance for small separation distances than either the low-NA or the high-NA lens. The diffuser can resolve objects as low as 0.3 super-pixels apart for more complex scenes, whereas the low-NA lens requires larger separation distances and the high-NA lens suffers errors due to gaps in the measurement.

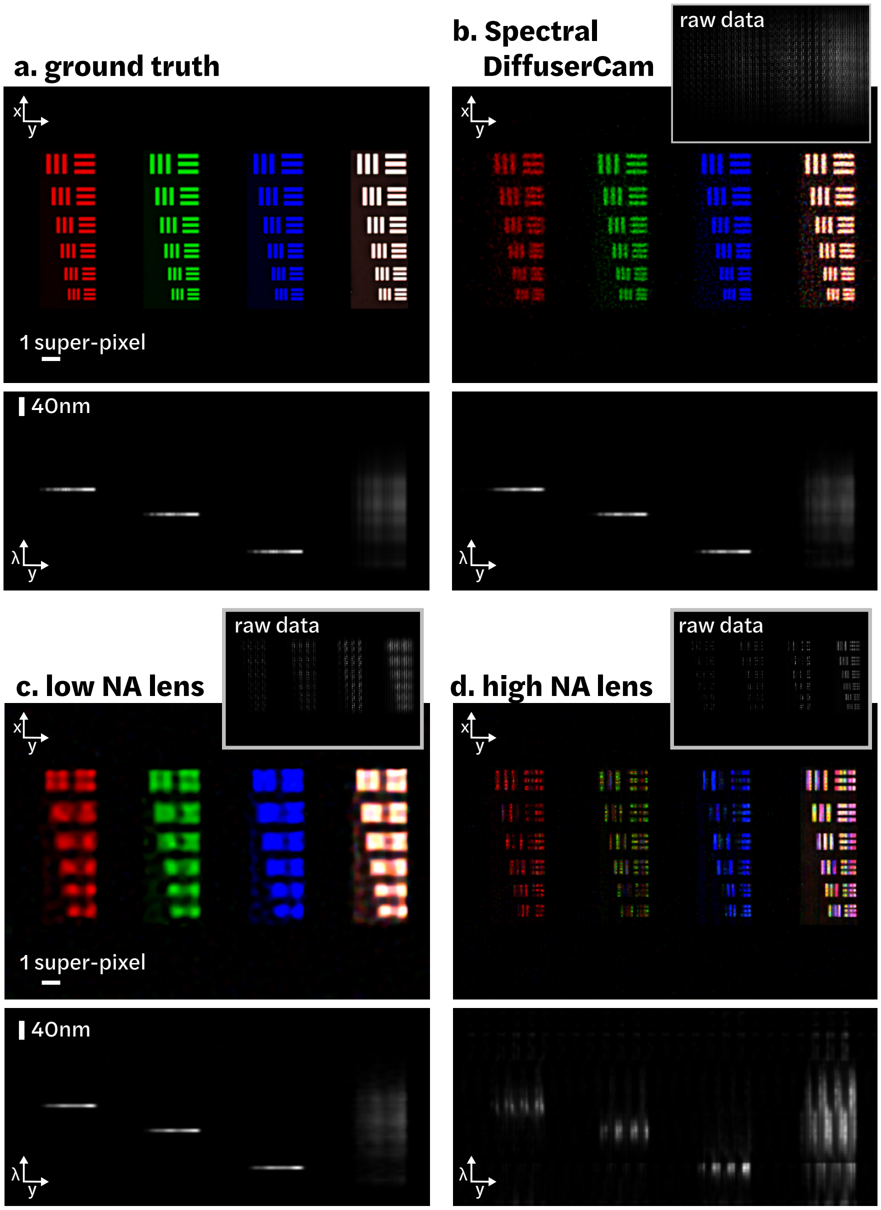

Figure 3.8: **Simulated hyperspectral reconstructions comparing our Spectral DiffuserCam result with alternative design options.** (a) Resolution target with different sections illuminated by narrow-band 634nm (red), 570nm (green), 474nm (blue), and broadband (white) sources. (b) Reconstruction of the target by Spectral DiffuserCam, (c) a low-NA lens design, and (d) a high-NA lens design, each showing the raw data, false-colored reconstruction and λy sum projection. The diffuser achieves higher spatial resolution and better accuracy than the low-NA and the high-NA lens.

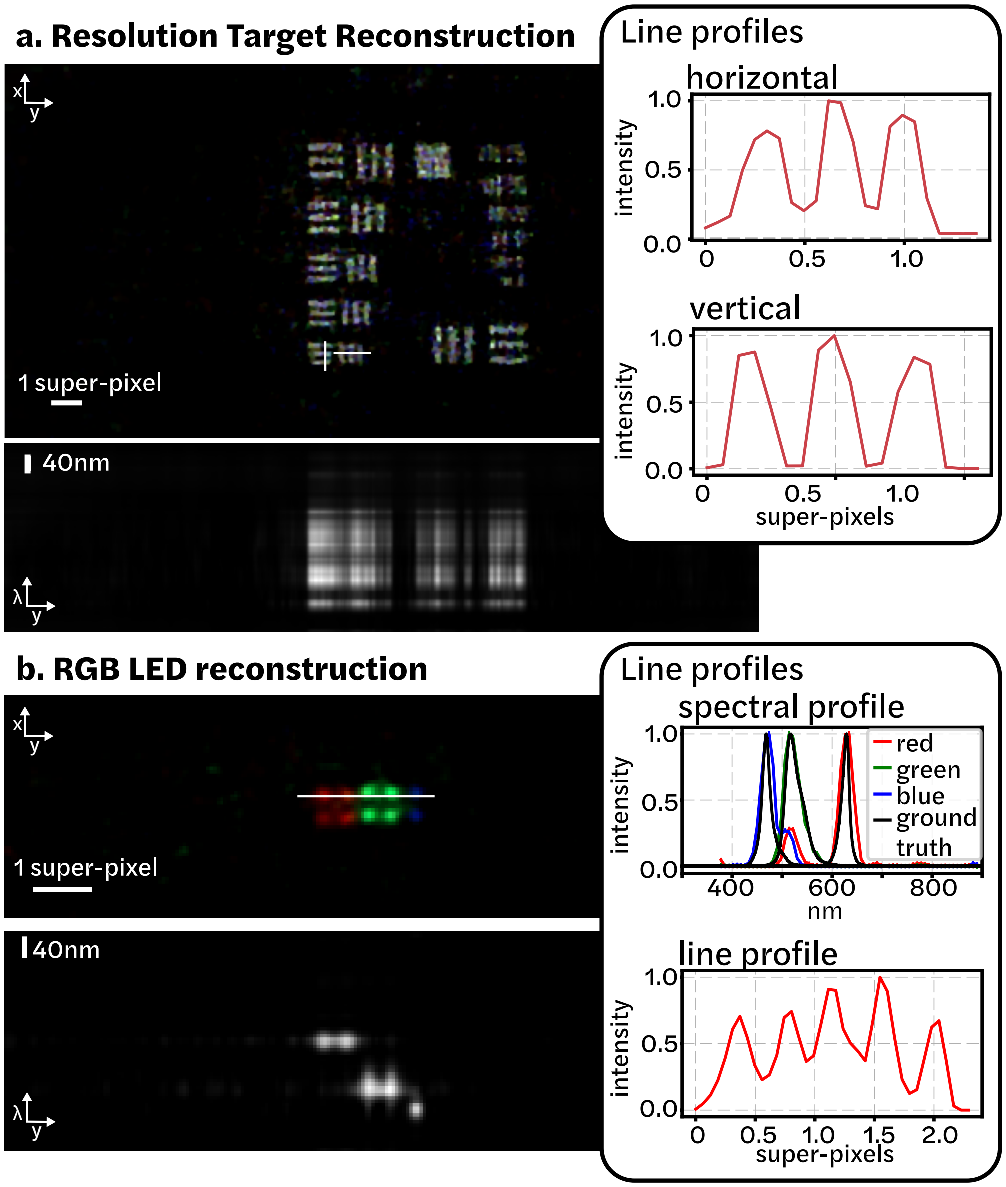

Figure 3.9: **Experimental resolution analysis:** (a) Experimental reconstruction of a broadband resolution target, showing the xy sum projection (top) and λy sum projection (bottom), demonstrating spatial resolution of 0.3 super-pixels. (b) Experimental reconstruction of 10 multi-colored LEDs in a grid with ~0.4 super-pixels spacing (four red LEDs on left, four green in middle, two blue at right). We show the xy sum projection (top) and λy sum projection (bottom). The LEDs are clearly resolved spatially and spectrally, and spectral line profiles for each color LED closely match the ground truth spectra from a spectrometer.

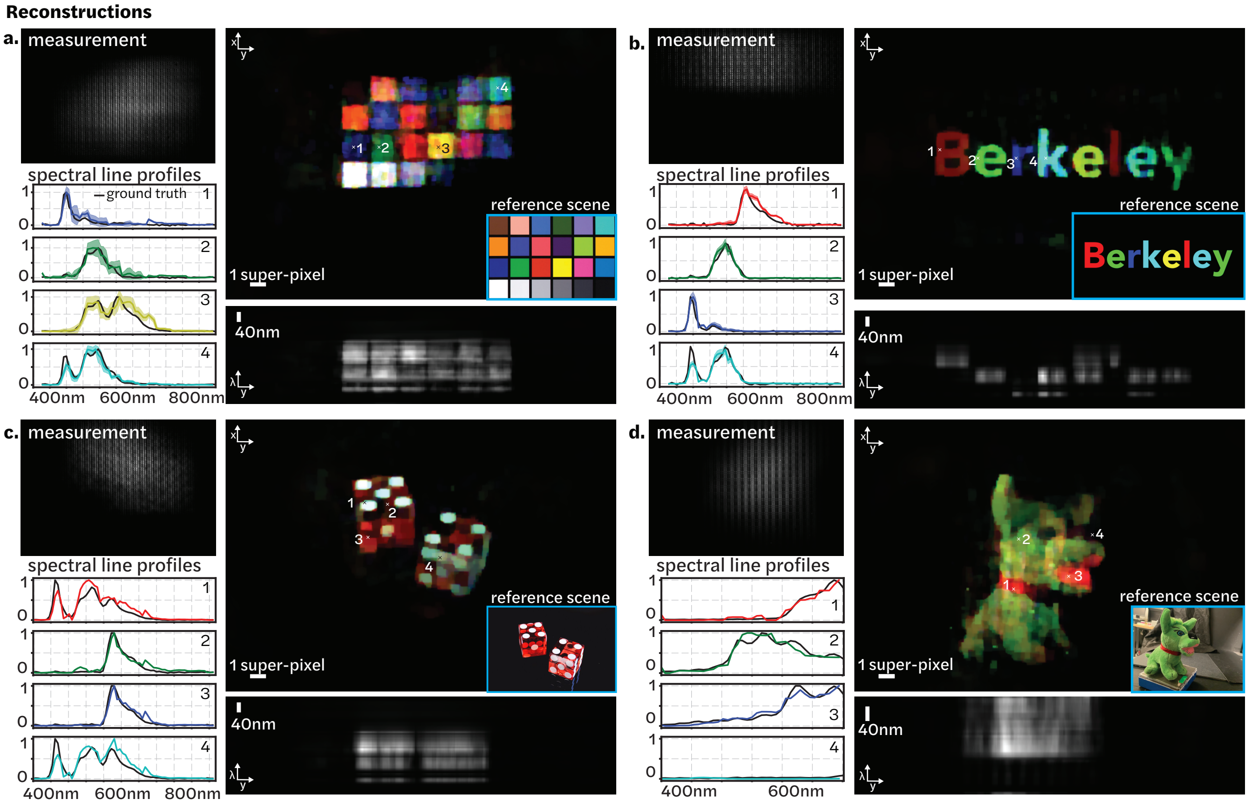

Figure 3.10: **Experimental hyperspectral reconstructions.** (a-c) Reconstructions of color images displayed on a computer monitor and (d) Thorlabs plush toy placed in front of the imager and illuminated by two Halogen lamps. The raw measurement, false color images, $x\lambda$ sum projections and spectral line profiles for four spatial points are shown for each scene. The ground truth spectral line profiles, measured using a spectrometer, are plotted in black for reference. Spectral line profiles in (a,b) show the average and standard deviation spectral profiles across the area of the box or letter in the object, whereas (c-d) show a line profile from a single spatial point in the scene.

Chapter 4

Deep learning for fast spatially-varying deconvolution

This chapter is based on [107] and is joint work with Kristina Monakhova, Richard W. Shuai, and Laura Waller.

4.1 Abstract

Deconvolution can be used to obtain sharp images or volumes from blurry or encoded measurements in imaging systems. Given knowledge of the system's point spread function (PSF) over the field-of-view, a reconstruction algorithm can be used to recover a clear image or volume. Most deconvolution algorithms assume shift-invariance; however, in realistic systems, the PSF varies laterally and axially across the field-of-view, due to aberrations or design. Shift-varying models can be used, but are often slow and computationally intensive. In this work, we propose a deep learning-based approach that leverages knowledge about the system's spatially-varying PSFs for fast 2D and 3D reconstructions. Our approach, termed MultiWienerNet, uses multiple differentiable Wiener filters paired with a convolutional neural network to incorporate spatial-variance. Trained using simulated data and tested on experimental data, our approach offers a $625-1600\times$ speed-up compared to iterative methods with a spatially-varying model, and outperforms existing deep-learning based methods that assume shift-invariance.

4.2 Introduction

Deconvolution is integral to many modern imaging systems. Imperfections in the optics may inadvertently blur the image (e.g. aberrations) and deconvolution can be used to computationally undo some of this blur [94, 86]. In microscopy, deconvolution can reduce out-of-focus fluorescence to provide sharper 3D images [74, 14, 90]. Alternatively, distributed point spread functions (PSFs) can be intentionally designed into an imaging system in order to enable new capabilities, such as single-shot 3D [109, 59, 66, 3, 8] or hyperspectral imaging [77, 49]. In this case, multiplexing optics encode 2D or 3D information by mapping each point in object space to a distributed pattern on the image sensor, then deconvolution is used to recover the encoded image or volume. In either case, a deconvolution algorithm is needed in order to recover a clear image or volume from the blurred or encoded measurement.

A variety of algorithms have been utilized for deconvolution over the years. Classical methods range from closed-form approaches such as Wiener filtering to iterative optimization approaches, such as Richardson-Lucy and the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA). Many methods incorporate hand-picked priors, such as Total Variation (TV) and native sparsity, to improve image quality. These approaches often assume that the system is shift-invariant, meaning that all parts of the image have the same blur kernel. Shift-invariance allows the forward model to be efficiently expressed as a convolution between the PSF and the object. However, most imaging systems will have a blur that varies across the field-of-view (FoV) - that is, they have spatially-varying PSFs, usually due to field-varying aberrations. This motivates the use of spatially-varying deconvolution, for which several methods have been proposed [5, 80, 73, 13, 28, 109, 59]. Unfortunately, many of these algorithms are prohibitively slow and computationally intensive, making them unsuitable for real-time image reconstruction. Furthermore, these methods can suffer from poor image quality, especially for highly multiplexed imaging systems that have PSFs with large spatial extent, or for poorly chosen priors. Recently, deep-learning based deconvolution methods have been demonstrated to improve both image quality and reconstruction speed, providing a promising improvement over iterative approaches [76, 98, 85, 55]. However, to date, these methods rely on a shift-invariant PSF approximation and do not generalize well to optical systems with field-varying aberrations.

In this work, we propose a new deep-learning based approach for fast, spatially-varying deconvolution. Our network, termed MultiWienerNet, consists of multiple learnable Wiener deconvolutions followed by a refinement convolutional neural network (CNN). The Wiener deconvolution layer performs multiple Fourier-space deconvolutions, each with a different PSF from a particular field point, yielding several intermediate images which have sharp features in different regions of the image. These intermediate images are then fed into the refinement CNN which fuses and refines them to create the final sharp deconvolved image. The learnable Wiener deconvolution filters are initialized with PSFs captured at several locations in the FoV, but then allowed to update throughout training to learn the best filters and noise regularization parameters. This allows us to incorporate knowledge of the field-varying aberrations into the network, providing a physically-informed initialization that is further refined throughout training. The end result is a fast spatially-varying deconvolution that is $625-1600\times$ faster than the baseline iterative method (Spatially-Varying FISTA [109]), enabling real-time image reconstruction. In addition, incorporating the field-varying PSFs allows our network to have better image quality near the edges of the FoV than is achieved by existing deep learning based methods which assume shift-invariance.

4.3 Theory and Results

Our approach consists of the following steps: 1) generating a simulated training dataset by applying measured PSFs to images from open-source microscopy datasets, 2) initializing and training the MultiWienerNet using the simulated data, and 3) utilizing the trained network for fast shift-varying deconvolutions, where the input to the network is a single measurement. To demonstrate, we choose the challenging example of single-shot 3D microscopy with Miniscope3D [109] as our test case. Miniscope3D uses a phase mask that consists of a random array of multi-focal microlenses to encode 3D information in a 2D image. The system maps each object point in the FoV to a unique pseudorandom pattern on the sensor, then decodes the captured images by solving a sparsity-constrained inverse problem. We select this system both for its spatially and depth-varying PSFs, Fig. 4.1, and its high degree of multiplexing, which creates a particularly challenging deconvolution problem. We demonstrate our approach on both 2D deconvolution, where the goal is to recover a 2D image from a 2D measurement, as well as 3D deconvolution, where the goal is to recover a 3D volume from a single 2D measurement.

To generate simulated datasets, we first need a forward model that can faithfully relate how a 3D object is mapped to a 2D measurement in our microscope system, taking into account the effects of spatially-varying blur introduced by the system. To establish this forward model, the volumetric object intensity is treated as a 3D grid of voxels, $\mathbf{v}[x, y, z]$. Each voxel produces a PSF, $\mathbf{h}[x', y'; x, y, z]$, on the camera sensor, where $[x', y']$ are image space indices. Since the object voxels are mutually incoherent, the measurement can be expressed as a linear combination of the PSFs from each voxel in the object:

$$\begin{aligned}\mathbf{b}[x', y'] &= \sum_z \sum_{x,y} \mathbf{v}[x, y, z]\mathbf{h}[x', y'; x, y, z] \\ &= \mathbf{A}\mathbf{v},\end{aligned} \tag{4.1}$$

where $\mathbf{b}$ is the measurement and $\mathbf{A}$ is a matrix that maps the 3D volume to the 2D measurement. For the shift-invariant case - where the PSF is the same at all points within the FoV for each depth - Eq. 4.1 reduces to a sum over 2D convolutions:

$$\mathbf{b}[x', y'] = \sum_z \sum_{x,y} \mathbf{v}[x, y, z] \overset{[x,y]}{*} \mathbf{h}[x, y, z], \tag{4.2}$$

where $\overset{[x,y]}{*}$ represents a 2D convolution. However, this shift-invariant assumption is generally not true and its corresponding convolutional forward model will lead to inaccurate deconvolution. There are multiple ways to approximate the image formation model for shift-invariant PSFs (e.g. locally convolutional, low-rank models, etc. [5, 73, 13, 28, 109, 32, 59]). Any of these techniques is applicable to our pipeline. In particular, we choose to use the low-rank model from [109], approximating the spatially-varying PSFs as a weighted sum of shift-invariant kernels:

$$\mathbf{b}[x', y'] = \sum_z \sum_{r=1}^{K} \left\{ (\mathbf{v}[x, y, z]\mathbf{w}_r[x, y, z]) \overset{[x,y]}{*} \mathbf{g}_r[x, y, z] \right\} [x', y'], \tag{4.3}$$

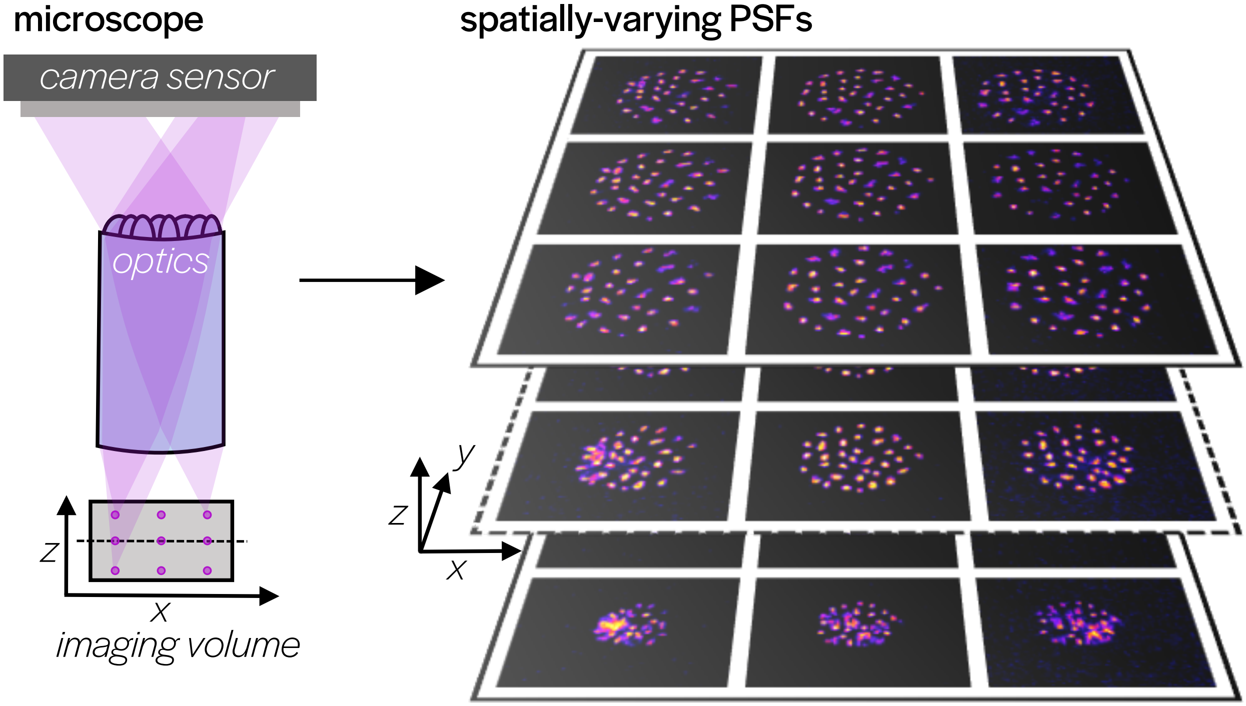

Figure 4.1: **Spatially-varying Point Spread Functions (PSFs)**. (left) Simplified diagram of Miniscope3D showing a lateral and axial scan of a point source through the volumetric field-of-view (FoV) to capture the spatially-varying PSFs. (right) Experimental images of the PSFs from different points in the volumetric FoV, which are used to initialize the Wiener deconvolution layer of our MultiWienerNet method.

where the weights $\{\mathbf{w}_r\}$ and the kernels $\{\mathbf{g}_r\}$ are computed from a singular value decomposition (SVD) of sparsely sampled PSFs from different positions in the FoV and the inner sum is over the K largest values in the SVD. Note that in the 2D imaging case, the object is a thin slice in the z-dimension, so the outer sum over z is not included in the forward model.

Using this forward model, we can simulate measurements from our microscope to use in training datasets. We run images from online microscopy datasets [57, 6, 69, 111] through the low-rank forward model, generating pairs of ground truth volumes/images and simulated measurements. Given a good system forward model (see Supplement for PSF calibration details), it is possible to generate any number of image pairs, which we can use to train our MultiWienerNet. We generate both 2D and 3D training datasets; the 2D dataset contains 2D target objects with dimensions (x,y,z) of (336,480,1), representing a FoV of 700 × 1000 μm^2, while the 3D datasets contain 3D target objects with dimensions (x,y,z) of (336,480,32), representing a FoV of $700 \times 1000 \times 320$ μm^3. We generate $5,000$ 2D and $15,000$ 3D training images, with an 80/20 training/testing split.

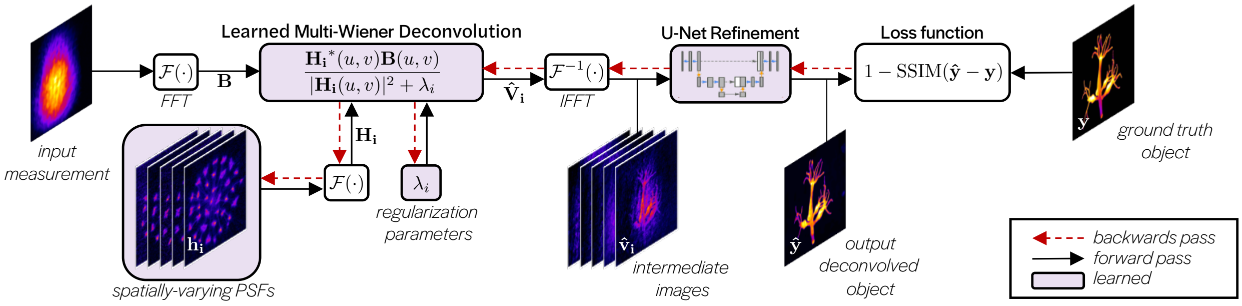
Learned Multi-Wiener Deconvolution
$\frac{\mathbf{H_i}^*(u,v)\mathbf{B}(u,v)}{|\mathbf{H_i}(u,v)|^2+\lambda_i}$
$\mathcal{F}(\cdot)$
FFT
$\mathbf{B}$
input measurement
$\hat{\mathbf{V}}_\mathbf{i}$
$\mathbf{H_i}$
$\mathcal{F}(\cdot)$
λ_i
regularization parameters
$\mathbf{h_i}$
spatially-varying PSFs
$\mathcal{F}^{-1}(\cdot)$
IFFT
U-Net Refinement
Loss function
$1 - \mathrm{SSIM}(\hat{\mathbf{y}} - \mathbf{y})$
$\mathbf{y}$
ground truth object
$\hat{\mathbf{v}}_\mathbf{i}$
intermediate images
$\hat{\mathbf{y}}$
output deconvolved object
backwards pass
forward pass
learned

Figure 4.2: **MultiWienerNet architecture**. Our pipeline consists of two parts: 1) a learnable multi-Wiener deconvolution layer which is initialized with knowledge of the system's spatially-varying PSFs and outputs a set of deconvolved intermediate images. 2) A U-Net refinement step which combines and refines the intermediate images into a single output image. Both parts are jointly-optimized during training using simulated data. After training, experimental measurements are fed into the optimized MultiWienerNet for fast spatially-varying deconvolution.

Our network consists of two components: a differentiable Wiener deconvolution layer, and a refinement CNN, Fig. 4.2. Wiener deconvolution is a fast and simple approach that is used for linear shift-invariant systems given a known PSF and noise level. It consists of a single Fourier filtering step, which can be efficiently computed using FFTs. However, when the assumption of shift-invariance does not hold, Wiener deconvolution results in degraded image quality in the areas of the image in which the PSF differs from the one assumed. Hence, instead of performing Wiener deconvolution with a single PSF [55], we approximate the behavior of our spatially-varying system using M PSFs taken from different field points. Our Wiener deconvolution layer thus performs M Wiener deconvolutions, resulting in M intermediate deconvolved images, as shown in Fig. 4.2. Each will have sharp features in a different region of the image, corresponding to the area in the FoV from which the PSF was taken. These M intermediate images are then fed into the refinement CNN which combines and refines the images to produce the final image/volume.

Mathematically, our differentiable Wiener deconvolution layer can be described as follows:

$$\hat{\mathbf{V}}_i(u,v) = \frac{\mathbf{H}_i^*(u,v)\mathbf{B}(u,v)}{|\mathbf{H}_i(u,v)|^2+\lambda_i}, \quad i = 1,2,...,M, \tag{4.4}$$

where (u,v) are frequency-space coordinates, $\mathbf{B}(u,v)$ is the Fourier transform of the measurement, $\hat{\mathbf{V}}_i(u,v)$ is the ith Fourier transform of the estimated scene intensity, $\mathbf{H}_i(u,v)$ is the Fourier transform of the ith PSF, * denotes a complex conjugate and λ_i is a regularization parameter related to the signal-to-noise ratio (SNR) of the measurement. Note that the intermediate images are obtained after taking an inverse Fourier transform, $\hat{\mathbf{v}}_i = \mathcal{F}^{-1}(\hat{\mathbf{V}}_i)$. Here, the $\mathbf{H}_i(u,v)$ are initialized using the M PSFs measured from different points in the

FoV. For 3D deconvolution, PSFs are sampled at multiple depth planes. Both $\mathbf{H}_i(u, v)$ and λ_i are learnable and optimized during training. Finally, the M intermediate images are fed into the U-Net refinement step which consists of a 2D U-Net for the 2D deconvolution problem or a 3D U-Net for the 3D problem [85].

For 2D deconvolution, $\mathbf{H}_i$ is initialized with measured PSFs from field points sampled on a 3×3 grid across the FoV, giving $M = 9$. For 3D, the PSFs are sampled on a $3 \times 3 \times 32$ grid across the FoV, giving $M = 288$. After the model is initialized with the M measured PSFs, the simulated pairs of measurements and ground truth volumes/images are used in training to update the parameters of the MultiWienerNet, including $\mathbf{H}_i$, λ_i, and all the parameters in the U-Net. We use a structural similarity index measure (SSIM) loss and L1 loss, which generally outperforms mean squared error (MSE) or L1 loss on its own. Training details are outlined in Supplement 1.

After training, we test our model on $1,000$ images in a held-out test set from the online datasets, and on experimental data from the Miniscope3D setup. We compare our results against iterative spatially-varying FISTA, a U-Net, and the U-Net with a single Wiener deconvolution [55]. Our results show that the MultiWienerNet achieves more than 625× speedup in 2D reconstruction and 1600× speedup in 3D reconstruction as compared to FISTA, while also providing better PSNR and image quality, especially towards the edges of the FoV, where off-axis aberrations dominate. Our network also outperforms current deep-learning approaches while only being slightly slower. In addition, despite being trained solely on simulated data, the MultiWienerNet generalizes well to experimental data. Though FISTA achieves slightly higher resolution near the center of the FoV (Fig. 4.3(a)), MultiWienerNet performs better overall.

4.4 Adapting network to new systems

To adapt our method to new systems, an accurate forward model is needed in order to generate simulated measurements for training. If a good forward model for the microscope already exists (e.g. Zemax model, existing simulator), such a model could be used in this step to generate simulated measurements. If such a model does not exist, or does not sufficiently model spatial-variance, we suggest a calibration procedure (detailed below) to measure the spatially-varying PSFs, paired with a low-rank forward model to simulate measurements. Given a good forward model, simulated measurements can be generated using existing online microscopy datasets and used to train the network.

System Calibration

A simple calibration procedure based on scanning a fluorescent bead (point source) across the field-of-view (FoV) can be used to characterize the spatially-varying behavior of the system. Rather than calibrating the PSF for every possible field point, we adopt the calibration method presented in [109], and sparsely sample the PSFs across the field, then use a low-

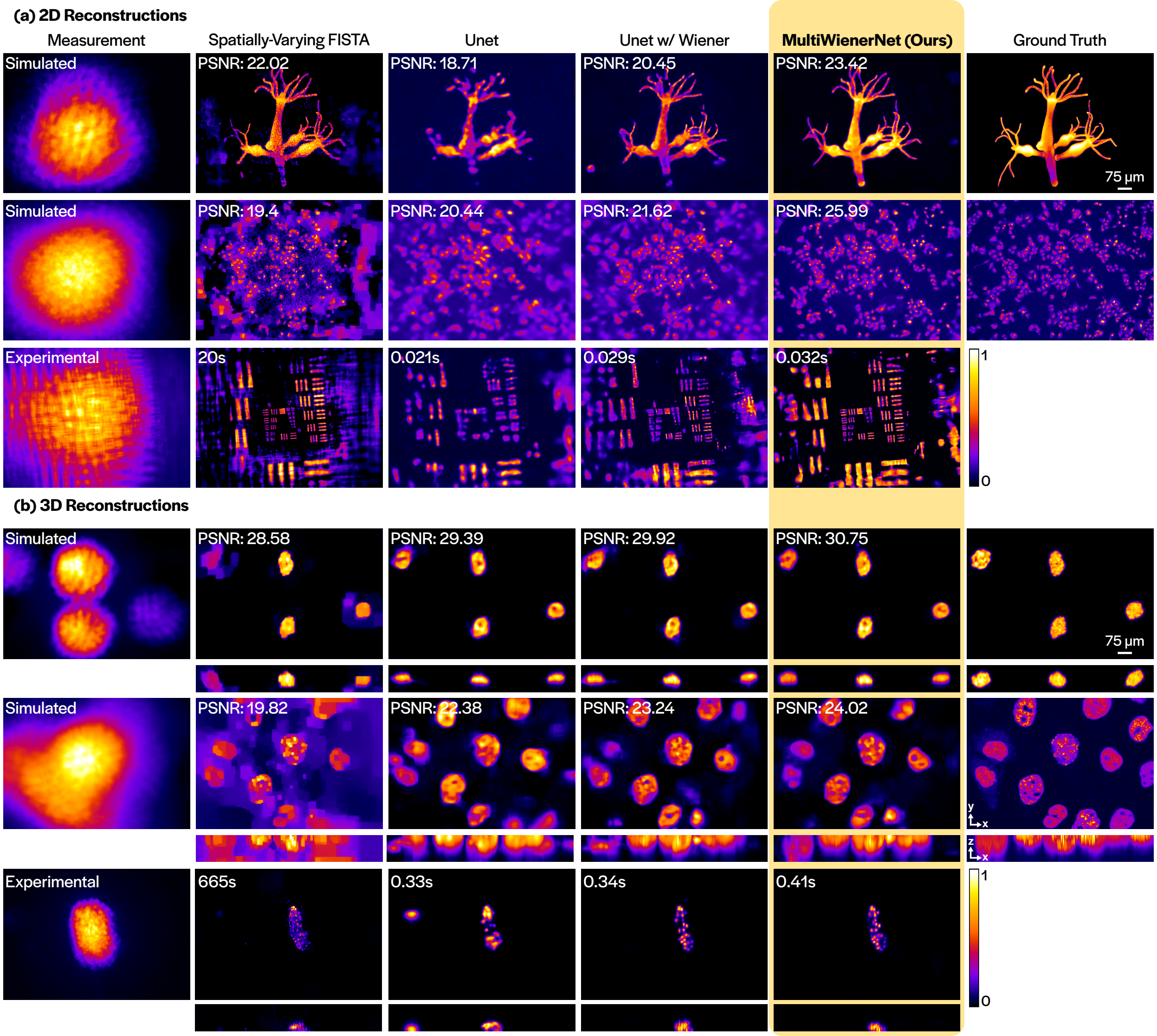

Figure 4.3: **Simulation and Experimental Results.** Deconvolution results for (a) 2D and (b) 3D, showing both simulated and experimental data. MultiWienerNet achieves better performance than other deep learning-based approaches that do not incorporate knowledge of a spatially-varying PSF (U-net and U-Net w/Wiener), and achieves better and faster ($625 - 1600\times$ speedup) results than spatially-varying FISTA, which has poor reconstruction quality at the edges of the FoV, where spatially-varying aberrations are severe. For the 3D results, xy and xz maximum projections are shown.

rank forward model to account for the shift-variance. In this work, we sample the PSF at 64 locations across the FoV for each depth. This can be done by simply scanning a bead and taking 64 calibration images across an 8×8 grid. If the bead is particularly dim, or there is noticeable noise present in the image, averaging multiple images together can help reduce noise and provide a better calibration image. If 3D deconvolution is desired, this calibration must be repeated for each depth of interest. In our case, we repeat this procedure for 32 different depths within the depth range of the microscope. To save time on calibration, it is also possible to do this calibration using a sample of unstructured beads instead of a single calibration bead [60].

Simulating measurements

To simulate measurements, we use data from existing online microscopy datasets. For the 3D dataset, the CytoPacq [21] simulator is particularly helpful in obtaining time-series 3D volumes. We follow the following pre-processing steps to generate the simulated measurement:

1. Resize the image/volume to fit the Miniscope3D image/volume size which is $336 \times 480 \times 1$ for 2D reconstruction and $336 \times 480 \times 32$ for 3D reconstruction

2. Using the sparsely sampled calibration PSFs, compute the weights $\{\mathbf{w}_r\}$, and the kernels $\{\mathbf{g}_r\}$, from a singular value decomposition (SVD) procedure outlined in [109].

3. Simulate measurements by running the resized data through the low-rank forward model in Eq. 3 of main text

4. Add appropriate levels of Gaussian and Poisson noise to the simulated measurement.

These pre-processing steps can be adapted to a new system by altering the desired image/volume size and using the calibrated PSFs for the new system.

Model initialization and network architecture

Before training, the MultiWienerNet must be initialized with the spatially-varying PSFs, $\mathbf{h}_i$, and additionally with regularization parameters, λ_i. Here, we utilize 9 PSFs from a 3×3 grid from each depth plane in our FoV to initialize $\mathbf{h}_i$. The number of initialized filters can be updated based on the level of spatial-variance of the system. Given more spatial-variance, a larger number of filters should be used; however, this comes at the price of added computational complexity. We found that 9 filters was sufficient for the Miniscope3D. For the U-net architecture, our contracting path consists of four repeated applications of two 3×3 convolutions, followed by a Scaled Exponential Linear Unit (SELU) and a convolutional down-sampling layer with stride 2. The expansive path consists of four repeated applications of two 3×3 convolutions, followed by a SELU and a transposed convolution up-sampling layer with strides 2.

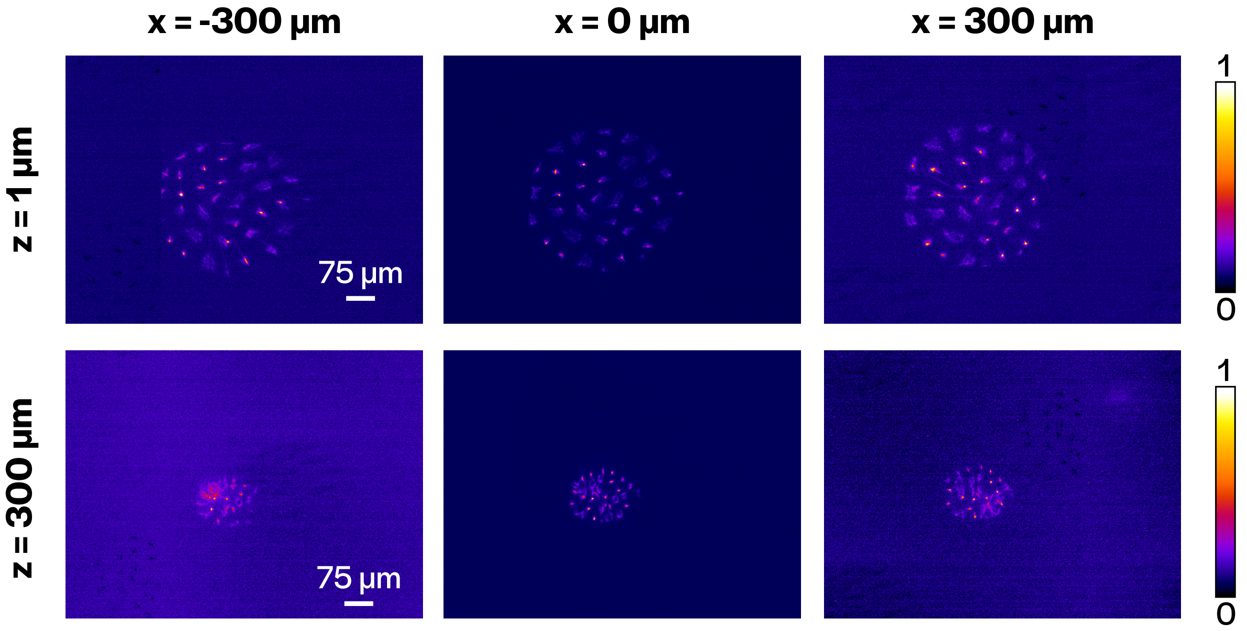

Figure 4.4: **Spatially varying PSFs**. Measured PSFs at different lateral positions for two different depth planes. Due to field-varying aberrations in the system, the PSFs change structure across the field-of-view (FoV). Note that the images are contrast stretched to show detail.

Training and Testing Details

We train for 25 epochs for the 3D reconstruction case and 100 epochs for the 2D case, using a learning rate of $1e^{-4}$. We use the ADAM optimizer [56] with default parameters throughout training. We allow the Wiener deconvolution filters and regularization parameters to update throughout training. We tested having the Wiener filters and/or the regularization parameters be fixed throughout training, but the best results were obtained when they were allowed to change. Our loss function is an average of a structural similarity index measure (SSIM) loss and L1 loss. For SSIM, we use an 11 × 11 Gaussian filter of width 1.5. We test the performance on both simulated and experimental data. The experimental sample in main-text Figure 3(a) is a 1951 USAF fluorescent resolution target, and in 3(b) is a freely moving stained tardigrade (water bear) captured at 40 frames per second.

4.5 Additional Results

Timing results

We compare our reconstruction time against FISTA with a spatially-varying model and existing deep learning based models (U-Net, U-Net w/Wiener) by averaging over 100 runs on GPU. The results are summarized in Table 4.1. We can see that our method is significantly faster than existing spatially-varying methods, and has comparable speeds to existing deep learning based methods. For 2D, we can perform reconstructions at 30fps for 336×480 images. For 3D, we can perform reconstructions at 2.4fps for 336×480×32 volumes. This is significantly faster than what is possible without using deep learning-based methods (0.05 fps for 2D, 0.0015 fps for 3D), and could enable interactive previewing. Timing results were performed on an RTX 2080 Ti GPU. The stopping criterion for FISTA was chosen to avoid unnecessary iterations. FISTA terminates if the solution is stable within a certain threshold, the loss is below a certain threshold, or the number of iterations exceeds a certain threshold.

Learned filters

We allow the network to learn the Wiener deconvolution filters. We initialize the network with a grid of measured filters, 3×3 for 2D and $3 \times 3 \times 32$ for 3D, each is centered on a different region in the FoV. Fig. 4.4 shows initial filters at different positions demonstrating the shift-variance of the system. By allowing the network to update the filters, the network can find filters that better approximate the shift-variance or better recover certain frequencies. Fig. 4.5, shows the initialized and learned filters. The learned filters maintain similar structure in the central region as the initialized ones, while adding more information towards the edges of the filter. At first glance, it is unclear if the extra information added to the learned filters is useful. However, by examining the intermediate result of the Wiener deconvolution step using both the initialized and learned filter in Fig. 4.6, we find that the deconvolved image using the learned filter has much sharper features than the one using the initialized filter. This allows the multiple learnable Wiener deconvolutions to provide the CNN with sharp intermediate images that the CNN can further refine to remove low frequency artifacts. Note that we do not require the learned filters to be positive.

Table 4.1: **Reconstruction timing comparisons (GPU)**

	FISTA	U-Net	WienerNet	**MultiWeinerNet**
2D	20s	0.021s	0.029s	0.032s
3D	665s	0.33s	0.34s	0.41s

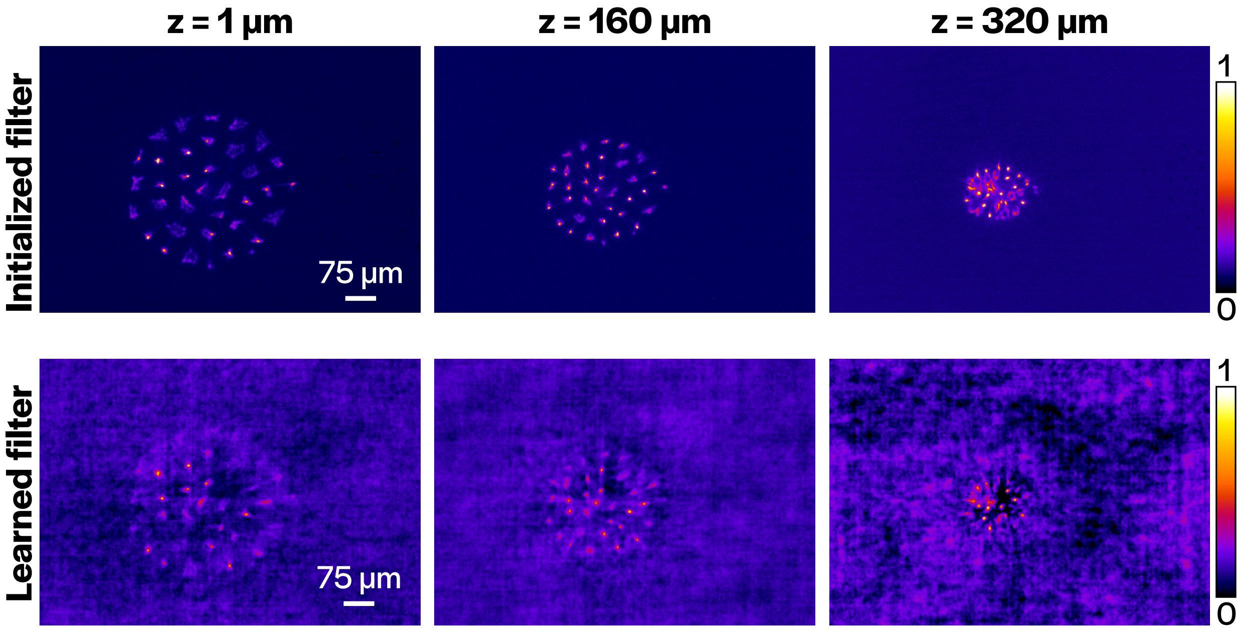

Figure 4.5: **Learned and initialized filters**. Central learned and initialized filters at different depths used for the 3D reconstruction. For visualization, we scale the learned filters from 0 to 1. In general, the filters can contain negative values. Note that the images are contrast stretched.

Deconvolution with localized PSFs

To demonstrate that our MultiWiener layer should generalize well to other imaging systems, we use Zemax to simulate PSFs from a traditional one-to-one imaging system with off axis aberrations (e.g. coma). Specifically, we simulate a dense grid of 512 × 512 PSFs using the specifications of the 2D Miniscope system [38]. From these PSFs, we simulate 2D measurements of sparse beads and a dense scene, of size 512 × 512, using the superposition principle. Fig. 4.7 shows the simulated measurement and the results of performing Wiener deconvolution on the measurement using the on-axis PSF and an off-axis PSF. As expected, deconvolution with the on-axis PSF in a system with spatially varying aberrations results in an image with good reconstruction quality only near the center portion of the image. In contrast, deconvolving with a PSF from an off-axis point results in sharper image quality in the neighborhood from which this PSF is sampled. This demonstrates how our MultiWiener stage can produce intermediate images where different portions of the image are sharp depending on where the PSF was sampled. These intermediate images can then be fed into the U-net, which will combine and refine the images to produce a sharp image across the FoV. We expect this general approach to work well in a variety of imaging systems with spatially-varying aberrations.

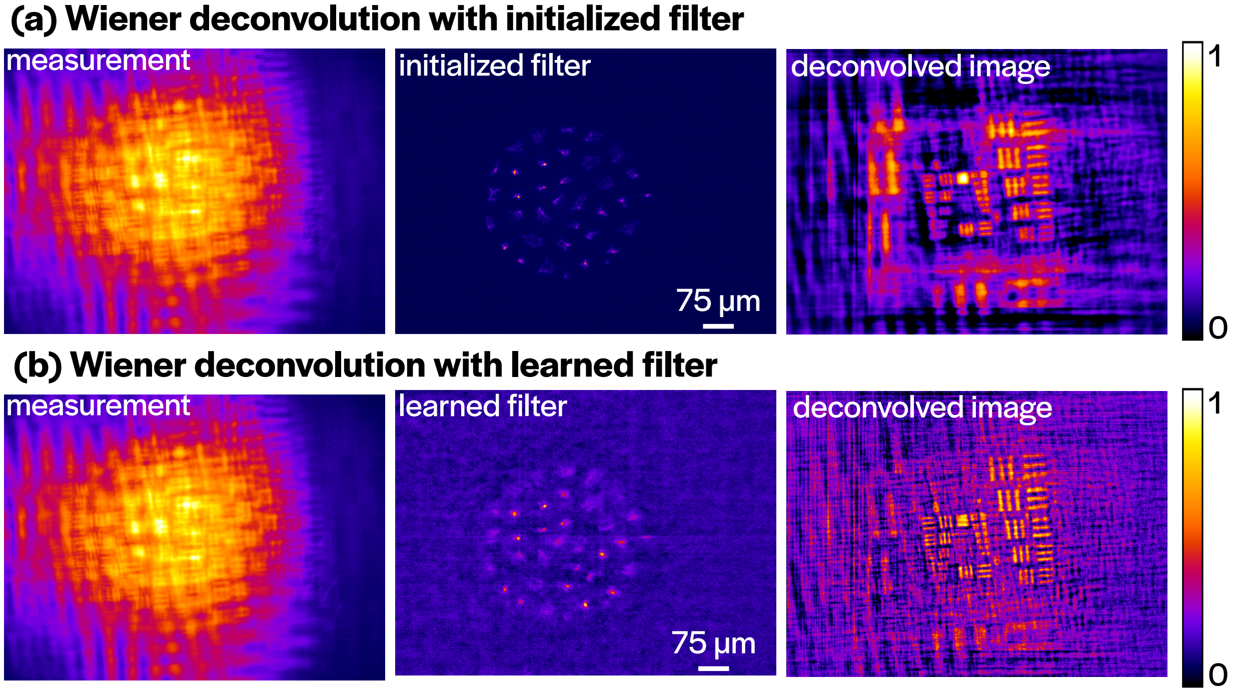

Figure 4.6: **Wiener deconvolution with learned filters**. (a) Using the central *initial* filter only, with a shift-invariance assumption. (b) Using the central *learned* filter with a shift invariance assumption results in sharper features in the deconvolved image.

In summary, we propose a new network architecture to perform fast deconvolution for microscopes with spatially-varying PSFs. Given knowledge of the system's spatially-varying PSFs, our proposed network is trained in simulation, fusing known system parameters in the form of multiple differentiable Wiener deconvolutions with a CNN refinement step. After training, our network provides a $625 - 1600\times$ speedup over existing spatially-varying deconvolution algorithms and improved reconstruction quality, especially at the edges of the FoV. The code is open-source and can be utilized for imaging system with spatially-varying aberrations.

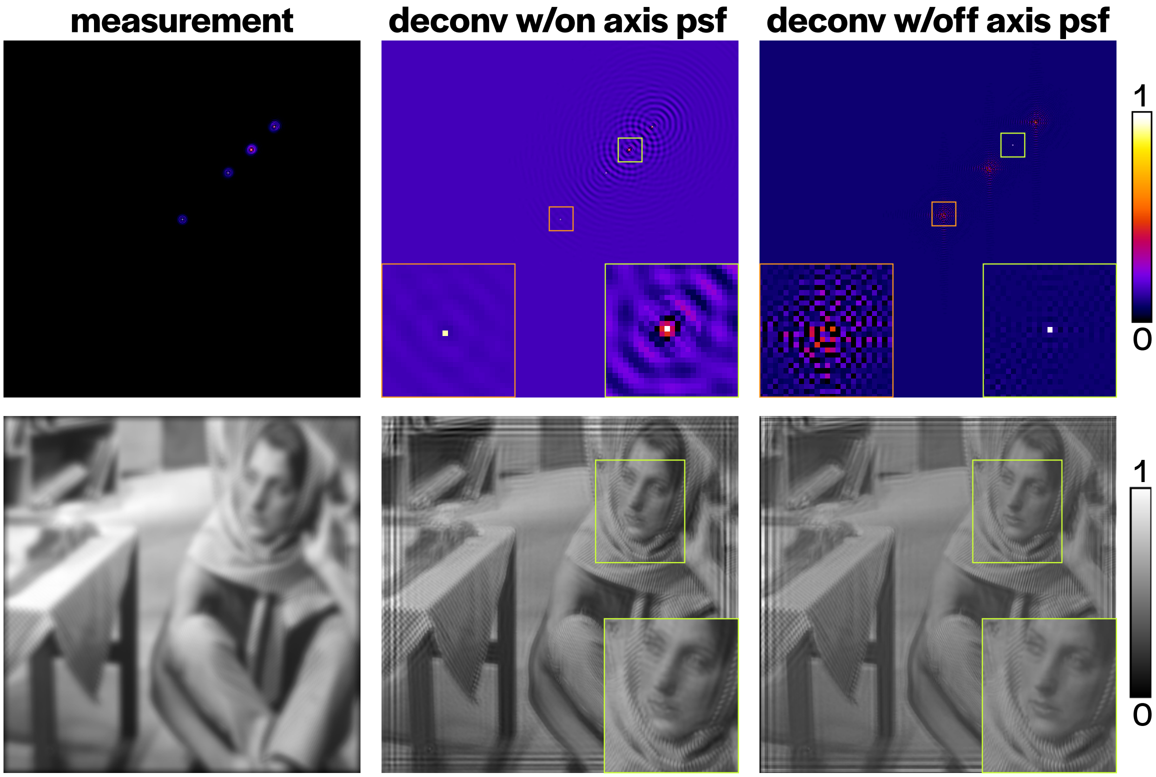

Figure 4.7: **Deconvolution with localized PSFs**. Results of performing Wiener deconvolution with on-axis and off-axis PSFs. (top) A simulated measurement showing 4 points (left), and the corresponding deconvolved image with an on-axis PSF (center) and an off-axis PSF (right). The orange inset shows an on-axis deconvolved point, that is sharper when the Wiener filter is using the on-axis PSF, and the green inset shows an off-axis deconvolved point, which is sharper when using an off-axis PSF. (bottom) A simulated measurement of a dense scene. The green inset shows fewer artifacts and sharper features when the off-axis PSF is used.

Chapter 5

Conclusion

This dissertation demonstrates how to design and model multiplexing optics to perform single-shot high-dimensional imaging (e.g. spectral and volumetric) in a compact imaging setup. *In Chapter 2*, an optimized phase mask consisting of multifocal non-uniformly spaced microlenses is used with a GRIN lens to achieve single-shot 3D imaging in a device smaller than a U.S. quarter. Since the PSF of the system is field-varying, we propose using a low-rank forward model that can incorporate field-varying aberrations achieving reconstructions with high spatial resolution. In addition, to enhance the 3D performance of the device, learned aberrations (e.g. astigmatism and tilt) are added to the phase mask to make the axial PSFs of the device as unique as possible, thus achieving high axial resolution. This chapter highlights the importance of two key aspects when designing computational imaging systems. First, having an accurate forward model is instrumental in achieving good reconstruction quality. Fig 2.4 (c) shows how the reconstruction quality changes when using a shift-invariant forward model as opposed to the low-rank field-varying forward model. A significant reduction in resolution is observed when using the shift-invariant model ($6.2\mu m$ lateral resolution with shift-invariant mode vs $2.7\mu m$ with field-varying model). Thus having an accurate forward model is a key factor for good reconstruction quality. Second, using our model for the phase mask, we are able to design the mask to target a specific lateral and axial resolution across a desired depth range. Our optimizer can change the microlenses' focal lengths, location, tilt, and add astigmatism to make the axial PFSs as unique as possible. Since we are relying on compressed sensing to recover the 3D volume from a 2D image, optimizing the phase mask using a merit function inspired from matrix coherence improves the quality of our reconstructions. This is to show that computational imaging systems perform best when they are designed with a target application in mind, an accurate forward model, and optics that are optimized for the reconstruction algorithm used.

In Chapter 3, we demonstrate a compact hyperspectral imager consisting for a diffuser combined with a tiled spectral filter array. The result is a single-shot hyperspectral system with a higher spatial resolution than that achieved by using the spectral filter alone. This chapter presents theory to quantify spatial and spectral resolution of objects with different sparsity levels. Since the reconstruction quality of compressed sensing systems rely on the object's sparsity level, we present methods to analyze the two-point spatial resolution as well as a the multi-point resolution. The two-point resolution can be thought of as a best case scenario and an upper limit performance for the system that will not be achieved for all objects. Thus, we make use of local condition number theory to predict real-world performance on complex objects. Using local condition number, we show that a diffuser achieves better performance and higher resolution for a wider range of objects than using either a low-NA or a high-NA lens with the spectral filter array.

In Chapter 4, we demonstrate a deep learning architecture that performs fast spatially-varying deconvolutions. In the previous chapters, an iterative optimization algorithm is used to recover the high-dimensional object from the 2D image. The iterative algorithm generally takes thousands of iterations to converge which prohibits real-time reconstructions. In addition, these algorithms rely on hand-tuned priors to achieve good reconstruction quality. Instead, our deep learning approach is able to provide real-time reconstructions as well as handle optical systems with field-varying aberrations. Our architecture consists of multiple differentiable Wiener filters combined with a convolutional neural network. The result of the multiple Wiener deconvolution layers is a set of intermediate images that have sharp features in different regions depending on where the PSF is sampled from. These intermediate images are then fed to a refinement convolutional neural network to blend them together and produce the final output. This chapter highlights two key points. First, combining convolutional neural networks with physics inspired layers (i.e. Wiener filters) provide better reconstruction quality than relying on convolutional neural networks alone. Second, having an accurate forward model of our optical system allows us to simulate thousands of training images that would be very difficult to capture experimentally due to the 3D nature of the objects. Despite being trained solely on simulated data, our architecture generalizes well to experimental data.

Future Directions

Below I provide a list of future directions and improvements that can be further studied and explored:

- **Volumetric optics**: In this dissertation, the multiplexing optic was either a phase mask consisting of designed microlenses or an off-the-shelf diffuser. While microlenses make a good multiplexing optic that can be easily optimized, recent fabrication developments have made metalenses, 3D diffractive optics, and 3D GRINs more available and easier to fabricate. Co-designing these optics with reconstruction algorithms can lead to better performing computational imaging systems.
- **Physics inspired deep learning**: Since iterative reconstruction algorithms require an accurate forward model for good reconstructions. That forward model can also be used to simulate training data to enable deep learning based reconstruction. This has the advantage of being real-time and handling more complex objects. We combined

differentiable Wiener filters with CNNs to achieve fast and good reconstructions for field-varying optical systems. Designing different physics inspired layers and combining it with different deep learning architectures can produce better and faster reconstructions.

- **Using time priors**: In this dissertation, we focused on using spatial sparsity priors. However, for many imaging applications we have access to data at different time points (e.g. videos of neural firings or videos of freely moving samples). In this case, time priors can be used to greatly improve reconstruction quality. While incorporating these priors with iterative optimization will result in an even slower reconstruction algorithm, I propose using recurrent neural networks as the means to incorporate time priors. If combined with physics-inspired layers, it has the potential to achieve faster and better reconstructions.

Bibliography

[1] Jesse K Adams et al. "Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope". In: *Science Advances* 3.12 (2017), e1701548.

[2] Hamed Akbari et al. "Hyperspectral imaging and quantitative analysis for prostate cancer detection". In: *Journal of Biomedical Optics* 17.7 (2012), p. 076005.

[3] Nick Antipa et al. "DiffuserCam: lensless single-exposure 3D imaging". In: *Optica* 5.1 (2018), pp. 1–9.

[4] Nick Antipa et al. "Video from stills: Lensless imaging with rolling shutter". In: *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2019, pp. 1–8.

[5] Muthuvel Arigovindan et al. "A Parallel Product-Convolution approach for representing depth varying Point Spread Functions in 3D widefield microscopy based on principal component analysis". In: *Opt. Express* 18.7 (2010), pp. 6461–6476. DOI: 10.1364/OE.18.006461. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-7-6461.

[6] Salim Arslan et al. "Attributed relational graphs for cell nucleus segmentation in fluorescence microscopy images". In: *IEEE transactions on medical imaging* 32.6 (2013), pp. 1121–1131.

[7] M Salman Asif et al. "Flatcam: Replacing lenses with masks and computation". In: *Computer Vision Workshop (ICCVW), 2015 IEEE International Conference on*. IEEE. 2015, pp. 663–666.

[8] M Salman Asif et al. "Flatcam: Thin, lensless cameras using coded aperture and computation". In: *IEEE Transactions on Computational Imaging* 3.3 (2016), pp. 384–397.

[9] Christina P Bacon, Yvette Mattley, and Ronald DeFrece. "Miniature spectroscopic instrumentation: applications to biology and chemistry". In: *Review of Scientific instruments* 75.1 (2004), pp. 1–16.

[10] Seung-Hwan Baek et al. "Compact single-shot hyperspectral imaging using a prism". In: *ACM Transactions on Graphics (TOG)* 36.6 (2017), pp. 1–12.

[11] Amir Beck and Marc Teboulle. “A fast iterative shrinkage-thresholding algorithm for linear inverse problems”. In: *SIAM Journal on Imaging Sciences* 2.1 (2009), pp. 183–202.

[12] Amir Beck and Marc Teboulle. “A fast iterative shrinkage-thresholding algorithm for linear inverse problems”. In: *SIAM Journal on Imaging Sciences* 2.1 (2009), pp. 183–202.

[13] Saima Ben Hadj, Laure Blanc-Féraud, and Gilles Aubert. “Space variant blind image restoration”. In: *SIAM Journal on Imaging Sciences* 7.4 (2014), pp. 2196–2225.

[14] David SC Biggs. “3D deconvolution microscopy”. In: *Current Protocols in Cytometry* 52.1 (2010), pp. 12–19.

[15] Ashish Bora et al. “Compressed sensing using generative models”. In: *arXiv preprint arXiv:1703.03208* (2017).

[16] Stephen Boyd et al. “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.

[17] Michael Broxton et al. “Wave optics theory and 3-D deconvolution for the light field microscope”. In: *Optics Express* 21.21 (2013), pp. 25418–25439. DOI: `10.1364/OE.21.025418`. URL: `http://www.opticsexpress.org/abstract.cfm?URI=oe-21-21-25418`.

[18] Emmanuel J Candès and Carlos Fernandez-Granda. “Towards a mathematical theory of super-resolution”. In: *Communications on pure and applied Mathematics* 67.6 (2014), pp. 906–956.

[19] Emmanuel J Candès and Michael B Wakin. “An introduction to compressive sampling”. In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 21–30.

[20] Xun Cao et al. “Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world”. In: *IEEE Signal Processing Magazine* 33.5 (2016), pp. 95–108.

[21] Centre for Biomedical Image Analysis. *CytoPacq*. (`https://cbia.fi.muni.cz/simulator`).

[22] Maumita Chakrabarti, Michael Linde Jakobsen, and Steen G Hanson. “Speckle-based spectrometer”. In: *Optics Letters* 40.14 (2015), pp. 3264–3267.

[23] Kuanglin Chao et al. “Hyperspectral-multispectral line-scan imaging system for automated poultry carcass inspection applications for food safety”. In: *Poultry Science* 86.11 (2007), pp. 2450–2460.

[24] Oliver Cossairt, Mohit Gupta, and Shree K Nayar. “When does computational imaging improve performance?” In: *IEEE Transactions on Image Processing* 22.2 (2012), pp. 447–458.

[25] Yuchao Dai et al. "Smooth Approximation of L_infinity-Norm for Multi-view Geometry". In: *2009 Digital Image Computing: Techniques and Applications*. IEEE. 2009, pp. 339–346.

[26] S. Dehaeck, B. Scheid, and P. Lambert. "Adaptive stitching for meso-scale printing with two-photon lithography". In: *Additive Manufacturing* 21 (2018), pp. 589–597. ISSN: 2214-8604. DOI: `https://doi.org/10.1016/j.addma.2018.03.026`. URL: `http://www.sciencedirect.com/science/article/pii/S2214860417305766`.

[27] Stephanie Delalieux et al. "Hyperspectral reflectance and fluorescence imaging to detect scab induced stress in apple leaves". In: *Remote Sensing* 1.4 (2009), pp. 858–874.

[28] Loic Denis et al. "Fast approximations of shift-variant blur". In: *International Journal of Computer Vision* 115.3 (2015), pp. 253–278.

[29] Steven Diamond et al. "Dirty pixels: Optimizing image classification architectures for raw sensor data". In: *arXiv preprint arXiv:1701.06487* (2017).

[30] Christoph J Engelbrecht, Fabian Voigt, and Fritjof Helmchen. "Miniaturized selective plane illumination microscopy for high-contrast in vivo fluorescence imaging". In: *Optics Letters* 35.9 (2010), pp. 1413–1415.

[31] Rob Fergus, Antonio Torralba, and William T. Freeman. "Random Lens Imaging". In: MIT CSAIL Technical Report 2006-058, 2006.

[32] Ralf C Flicker and François J Rigaut. "Anisoplanatic deconvolution of adaptive optics images". In: *JOSA A* 22.3 (2005), pp. 504–513.

[33] Ralf C Flicker and François J Rigaut. "Anisoplanatic deconvolution of adaptive optics images". In: *JOSA A* 22.3 (2005), pp. 504–513.

[34] Rebecca French, Sylvain Gigan, and Otto L Muskens. "Speckle-based hyperspectral imaging combining multiple scattering and compressive sensing in nanowire mats". In: *Optics Letters* 42.9 (2017), pp. 1820–1823.

[35] Nahum Gat. "Imaging spectroscopy using tunable filters: a review". In: *Wavelet Applications VII*. Vol. 4056. International Society for Optics and Photonics. 2000, pp. 50–64.

[36] Michael E Gehm et al. "Single-shot compressive spectral imaging with a dual-disperser architecture". In: *Optics Express* 15.21 (2007), pp. 14013–14027.

[37] Kunal K Ghosh et al. "Miniaturized integration of a fluorescence microscope". In: *Nature Methods* 8.10 (2011), p. 871.

[38] Kunal K Ghosh et al. "Miniaturized integration of a fluorescence microscope". In: *Nature methods* 8.10 (2011), pp. 871–878.

[39] Michael A Golub et al. "Compressed sensing snapshot spectral imaging by a regular digital camera with an added optical diffuser". In: *Applied Optics* 55.3 (2016), pp. 432–443.

[40] AA Gowen et al. “Hyperspectral imaging–an emerging process analytical tool for food quality and safety control”. In: *Trends in Food Science & Technology* 18.12 (2007), pp. 590–598.

[41] Robert O Green et al. “Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)”. In: *Remote Sensing of Environment* 65.3 (1998), pp. 227–248.

[42] Andres de Groot et al. “NINscope, a versatile miniscope for multi-region circuit investigations”. In: *eLife* 9 (2020), e49987.

[43] Changliang Guo et al. “Fourier light-field microscopy”. In: *Optics Express* 27.18 (2019), pp. 25573–25594.

[44] Jonathan Hauser et al. “Dual-camera snapshot spectral imaging with a pupil-domain optical diffuser and compressed sensing algorithms”. In: *Applied Optics* 59.4 (2020), pp. 1058–1070.

[45] Fritjof Helmchen et al. “A miniature head-mounted two-photon microscope: high-resolution brain imaging in freely moving animals”. In: *Neuron* 31.6 (2001), pp. 903–912.

[46] Andrew Hennessy, Kenneth Clarke, and Megan Lewis. “Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability”. In: *Remote Sensing* 12.1 (2020), p. 113.

[47] Wenqian Huang et al. “Development of a multispectral imaging system for online detection of bruises on apples”. In: *Journal of Food Engineering* 146 (2015), pp. 62–71.

[48] Alexander D Jacob et al. “A Compact Head-Mounted Endoscope for In Vivo Calcium Imaging in Freely Behaving Mice”. In: *Current Protocols in Neuroscience* 84.1 (2018), e51.

[49] Daniel S. Jeon et al. “Compact Snapshot Hyperspectral Imaging with Diffracted Rotation”. In: *ACM Trans. Graph.* 38.4 (July 2019). ISSN: 0730-0301. DOI: `10.1145/3306346.3322946`. URL: `https://doi.org/10.1145/3306346.3322946`.

[50] Ulugbek S Kamilov. “A parallel proximal algorithm for anisotropic total variation minimization”. In: *IEEE Transactions on Image Processing* 26.2 (2016), pp. 539–548.

[51] Ulugbek S Kamilov. “A parallel proximal algorithm for anisotropic total variation minimization”. In: *IEEE Transactions on Image Processing* 26.2 (2016), pp. 539–548.

[52] Ori Katz et al. “Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations”. In: *Nature Photonics* 8.10 (2014), pp. 784–790.

[53] Robert T Kester et al. “Real-time snapshot hyperspectral imaging endoscope”. In: *Journal of Biomedical Optics* 16.5 (2011), p. 056005.

[54] Salman S Khan et al. "Towards photorealistic reconstruction of highly multiplexed lensless images". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 7860–7869.

[55] Salman Siddique Khan et al. "Flatnet: Towards photorealistic scene reconstruction from lensless measurements". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[56] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[57] Can Fahrettin Koyuncu, Rengul Cetin-Atalay, and Cigdem Gunduz-Demir. "Object-Oriented Segmentation of Cell Nuclei in Fluorescence Microscopy Images". In: *Cytometry Part A* 93.10 (2018), pp. 1019–1028.

[58] Grace Kuo et al. "DiffuserCam: diffuser-based lensless cameras". In: *Computational Optical Sensing and Imaging*. Optical Society of America. 2017, CTu3B–2.

[59] Grace Kuo et al. "On-chip fluorescence microscopy with a random microlens diffuser". In: *Optics express* 28.6 (2020), pp. 8384–8399.

[60] Grace Kuo et al. "Spatially-varying microscope calibration from unstructured sparse inputs". In: *Computational Optical Sensing and Imaging*. Optical Society of America. 2020, CF4C–4.

[61] Pierre-Jean Lapray et al. "Multispectral filter arrays: Recent advances and practical implementation". In: *Sensors* 14.11 (2014), pp. 21626–21659.

[62] Richard M Levenson et al. "Multiplexing with multispectral imaging: from mice to microscopy". In: *ILAR Journal* 49.1 (2008), pp. 78–88.

[63] Marc Levoy et al. "Light Field Microscopy". In: *ACM Trans. Graph. (Proc. SIGGRAPH)* 25.3 (2006).

[64] William A Liberti III et al. "An open source, wireless capable miniature microscope system". In: *Journal of Neural Engineering* 14.4 (2017), p. 045001.

[65] Xing Lin et al. "Spatial-spectral encoded compressive hyperspectral imaging". In: *ACM Transactions on Graphics (TOG)* 33.6 (2014), pp. 1–11.

[66] Fanglin Linda Liu et al. "Fourier diffuserScope: single-shot 3D Fourier light field microscopy with a diffuser". In: *Optics Express* 28.20 (2020), pp. 28969–28986.

[67] Fanglin Linda Liu et al. "Single-shot 3D fluorescence microscopy with Fourier DiffuserCam". In: *Novel Techniques in Microscopy*. Optical Society of America. 2019, NS2B–3.

[68] Zhaoqiang Liu and Jonathan Scarlett. "Information-theoretic lower bounds for compressive sensing with generative models". In: *IEEE Journal on Selected Areas in Information Theory* (2020).

[69] Vebjorn Ljosa, Katherine L Sokolnicki, and Anne E Carpenter. "Annotated high-throughput microscopy image sets for validation." In: *Nature methods* 9.7 (2012), pp. 637–637.

[70] A Llavador et al. "Resolution improvements in integral microscopy with Fourier plane recording". In: *Optics express* 24.18 (2016), pp. 20792–20798.

[71] Guolan Lu and Baowei Fei. "Medical hyperspectral imaging: a review". In: *Journal of Biomedical Optics* 19.1 (2014), p. 010901.

[72] Guolan Lu et al. "Spectral-spatial classification for noninvasive cancer detection using hyperspectral imaging". In: *Journal of Biomedical Optics* 19.10 (2014), p. 106004.

[73] Elie Maalouf, Bruno Colicchio, and Alain Dieterlen. "Fluorescence microscopy three-dimensional depth variant point spread function interpolation using Zernike moments". In: *J. Opt. Soc. Am. A* 28.9 (2011), pp. 1864–1870. DOI: `10.1364/JOSAA.28.001864`. URL: `http://josaa.osa.org/abstract.cfm?URI=josaa-28-9-1864`.

[74] James G McNally et al. "Three-dimensional imaging by deconvolution microscopy". In: *Methods* 19.3 (1999), pp. 373–385.

[75] Sofiane Mihoubi et al. "Multispectral demosaicing using pseudo-panchromatic image". In: *IEEE Transactions on Computational Imaging* 3.4 (2017), pp. 982–995.

[76] Kristina Monakhova et al. "Learned reconstructions for practical mask-based lensless imaging". In: *Optics express* 27.20 (2019), pp. 28075–28090.

[77] Kristina Monakhova et al. "Spectral DiffuserCam: Lensless snapshot hyperspectral imaging with a spectral filter array". In: *Optica* 7.10 (2020), pp. 1298–1307.

[78] Tobias Nöbauer et al. "Video rate volumetric Ca 2+ imaging across cortex using seeded iterative demixing (SID) microscopy". In: *Nature Methods* 14.8 (2017), p. 811.

[79] Antony Orth et al. "Gigapixel multispectral microscopy". In: *Optica* 2.7 (2015), pp. 654–662.

[80] Nurmohammed Patwary and Chrysanthe Preza. "Image restoration for three-dimensional fluorescence microscopy using an orthonormal basis for efficient representation of depth-variant point-spread functions". In: *Biomed. Opt. Express* 6.10 (2015), pp. 3826–3841. DOI: `10.1364/BOE.6.003826`. URL: `http://www.osapublishing.org/boe/abstract.cfm?URI=boe-6-10-3826`.

[81] Sri Rama Prasanna Pavani and Rafael Piestun. "Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system". In: *Optics Express* 16.26 (2008), pp. 22048–22057.

[82] Yifan Peng et al. "Learned large field-of-view imaging with thin-plate optics". In: *ACM Transactions on Graphics (TOG)* 38.6 (2019), p. 219.

[83] Eftychios A Pnevmatikakis and Andrea Giovannucci. "NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data". In: *Journal of Neuroscience Methods* 291 (2017), pp. 83–94.

[84] Brandon Redding et al. “Compact spectrometer based on a disordered photonic chip”. In: *Nature Photonics* 7.9 (2013), p. 746.

[85] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].

[86] Daniel Sage et al. “DeconvolutionLab2: An open-source software for deconvolution microscopy”. In: *Methods* 115 (2017). Image Processing for Biologists, pp. 28–41. ISSN: 1046-2023. DOI: https://doi.org/10.1016/j.ymeth.2016.12.015. URL: https://www.sciencedirect.com/science/article/pii/S1046202316305096.

[87] Sujit Kumar Sahoo, Dongliang Tang, and Cuong Dang. “Single-shot multispectral imaging with a monochromatic camera”. In: *Optica* 4.10 (2017), pp. 1209–1213.

[88] Sam Dehaeck. *TipSlicer*. (https://github.com/SamDehaeck/TipSlicer).

[89] Vishwanath Saragadam and Aswin C Sankaranarayanan. “Programmable Spectrometry: Per-pixel Material Classification using Learned Spectral Filters”. In: *2020 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2020, pp. 1–10.

[90] Pinaki Sarder and Arye Nehorai. “Deconvolution methods for 3-D fluorescence microscopy images”. In: *IEEE Signal Processing Magazine* 23.3 (2006), pp. 32–45.

[91] Steve Saxe et al. “Advances in miniaturized spectral sensors”. In: *Next-Generation Spectroscopic Technologies XI*. Vol. 10657. International Society for Optics and Photonics. 2018, 106570B.

[92] G Scrofani et al. “FIMic: design for ultimate 3D-integral microscopy of in-vivo biological samples”. In: *Biomedical optics express* 9.1 (2018), pp. 335–346.

[93] Jaewook Shin et al. “A minimally invasive lens-free computational microendoscope”. In: *Science Advances* 5.12 (2019), eaaw5595.

[94] Jean-Baptiste Sibarita. “Deconvolution microscopy”. In: *Microscopy Techniques* (2005), pp. 201–243.

[95] Vincent Sitzmann et al. “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging”. In: *ACM Transactions on Graphics (TOG)* 37.4 (2018), pp. 1–13.

[96] Oliver Skocek et al. “High-speed volumetric imaging of neuronal activity in freely moving rodents”. In: *Nature Methods* (2018), p. 1.

[97] Da-Wen Sun. *Hyperspectral imaging for food quality analysis and control*. Elsevier, 2010.

[98] Florent Sureau, Alexis Lechat, and J-L Starck. “Deep learning for a space-variant deconvolution in galaxy surveys”. In: *Astronomy & Astrophysics* 641 (2020), A67.

[99] Jun Tanida et al. “Color imaging with an integrated compound imaging system”. In: *Optics Express* 11.18 (2003), pp. 2109–2117.

[100] Jun Tanida et al. “Thin observation module by bound optics: concept and experimental verification”. In: *Applied Optics* 40.11 (2001), pp. 1806–1813.

[101] S. Thiele et al. “3D-printed eagle eye: Compound microlens system for foveated imaging”. In: *Science Advances* 3.2 (Feb. 15, 2017), e1602655.

[102] UCLA. *Miniscope*. ⟨`https://miniscope.org`⟩.

[103] Ashwin Wagadarikar et al. “Single disperser design for coded aperture snapshot spectral imaging”. In: *Applied Optics* 47.10 (2008), B44–B51.

[104] Peng Wang and Rajesh Menon. “Computational snapshot angular-spectral lensless imaging”. In: *arXiv preprint arXiv:1707.08104* (2017).

[105] Zhu Wang and Zongfu Yu. “Spectral analysis based on compressive sensing in nanophotonic structures”. In: *Optics Express* 22.21 (2014), pp. 25608–25614.

[106] Zhu Wang et al. “Single-shot on-chip spectral sensors based on photonic crystal slabs”. In: *Nature communications* 10.1 (2019), pp. 1–6.

[107] Kyrollos Yanny et al. “Deep learning for fast spatially varying deconvolution”. In: *Optica* 9.1 (2022), pp. 96–99.

[108] Kyrollos Yanny et al. “Miniature 3D Fluorescence Microscope Using Random Microlenses”. In: *Optics and the Brain*. Optical Society of America. 2019, BT3A–4.

[109] Kyrollos Yanny et al. “Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy”. In: *Light: Science & Applications* 9.1 (2020), pp. 1–13.

[110] Chen Zhang et al. “A novel 3D multispectral vision system based on filter wheel cameras”. In: *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE. 2016, pp. 267–272.

[111] Yide Zhang et al. “A Poisson-Gaussian Denoising Dataset with Real Fluorescence Microscopy Images”. In: *CVPR*. 2019.

[112] Weijian Zong et al. “Fast high-resolution miniature two-photon microscopy for brain imaging in freely behaving mice”. In: *Nature Methods* 14.7 (2017), p. 713.