Title

From Photonics to AI: A Holistic Framework for Next-Generation 4D Fluorescence Microscopy

Author

Carmona, Javier

Publication Date

2025

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

From Photonics to AI: A Holistic Framework for Next-Generation 4D Fluorescence Microscopy

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical and Computer Engineering

by

Javier Carmona Jr

2025

ABSTRACT OF THE DISSERTATION

From Photonics to AI: A Holistic Framework for Next-Generation 4D Fluorescence Microscopy

by

Javier Carmona Jr
Doctor of Philosophy in Electrical and Computer Engineering
University of California, Los Angeles, 2025
Professor Katsushi Arisaka, Chair

Recording the neural activity of biological organisms is paramount to understanding how they, and consequently, we, process the world. Fluorescence microscopy has served as the standard for recording this neural activity due to its ability to capture large populations of neurons simultaneously. Recent efforts in fluorescence microscopy have been concentrated on imaging large-scale volumes; however, most of these efforts have been limited by spatiotemporal and bandwidth constraints.

I present a novel system called Transverse-Sheet Illumination Microscopy (TranSIM), which captures axially separated planes onto multiple two-dimensional sCMOS sensors at near diffraction-limited resolution with 1.0 µm, 1.4 µm, and 4.3 µm (x, y, and z, respectively). The parallel use of sensors reduces the bandwidth bottlenecks typically found in other systems. TranSIM allows for the capturing of data at large-scale volumetric fields of view up to $748 \times 278 \times 100$ µm^3 at 100 Hz. Moreover, I was able to capture smaller fields of view of $374 \times 278 \times 100$ µm^3 at a faster volumetric rate of 200 Hz. Additionally, I found that the system's versatile design allowed us to change the vertical magnification programmatically rather than necessitating a change of objectives. With this baseline system, I was able to record intricate neuronal communication in both larval and adult stage fruit flies. Moreover, I was able to reconstruct the complex physiological deformation of larval stage zebrafish hearts.

Despite its advantages, TranSIM acquires sparsely sampled volumetric data, necessitating computational reconstruction techniques to infer missing planes. To address this, I leveraged deep learning-based volumetric reconstruction methods to enhance data continuity. I first explored three-dimensional convolutional neural networks (3D-CNNs) with self-attention mechanisms, which effectively capture spatial dependencies and refine structural details across planes. These time-independent networks demonstrated high performance in reconstructing static volumes, and while satisfactory, are potentially limited when capturing temporally evolving neural dynamics. To overcome these limitations, I further investigated the implementation of four-dimensional recurrent neural networks (4D-RNNs), which integrate temporal dependencies alongside spatial information. By incorporating recurrent components, in the form of long-short term memory, these networks improved temporal coherence in the reconstructions, particularly in dynamic imaging experiments.

These results highlight the potential of artificial neural networks to significantly enhance TranSIM's imaging capabilities, enabling accurate volumetric reconstructions from sparse data while preserving both spatial and temporal fidelity. This advancement paves the way for more efficient high-speed volumetric fluorescence microscopy, facilitating the study of large-scale neural networks in living organisms with unprecedented detail. All together, I have demonstrated how the combination of TranSIM coupled with neural networks can serve as the framework for next-generation 4D microscopy.

The dissertation of Javier Carmona Jr is approved.

Dan Ruan

Daniel Aharoni

Aydogan Ozcan

Katsushi Arisaka, Committee Chair

University of California, Los Angeles

2025

To my mother and father . . .

You showed me what it means to never give up.

Thank you for always believing in me.

. . .

To my brother and sisters . . .

You supported me in my toughest moments.

. . .

To my nieces and nephews . . .

You gave me the courage to keep going.

. . .

To Tia Anita . . .

May you rest in peace.

Thank you for giving me a home away from home.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Katsushi Arisaka... For going above and beyond as a mentor, profoundly shaping my trajectory. Your guidance and support have changed my life for the better, and I am forever grateful.

Aydogan Ozcan ... For you pioneering work in virtual refocusing. It was definitely inspirational work and definitely propelled me work with TranSIM.

Dan Ruan ... I thoroughly enjoyed taking you course in the Fall of 2023. It greatly motivated me to complete the neural network aspect of this work!

Daniel Aharoni ... For the great conversations and being part of my committee. You are definitely an inspiration as a researcher!

Laurent Bentolila ... For being there in a professional and personal manner. Your great guidance convinced my not to give up on this work and see if to the end!

Blake Madruga ... For being a friend and an inspiration. Congrats on making it to MIT. Looking forward to seeing what the future holds.

Elizabeth Mills ... For being a great friend and role model. I really appreciate all that you have done for me, especially in my time of need.

Steve Mendoza ... For also being a great friend! I'm proud of you and your achievements. We are on our way!

Brian Jeong ... For the countless beads and zebrafish samples, and for helping with data taking and optics alignment.

The Elegant Mind Club ... This club, organized by Dr. Arisaka, has proved invaluable to my research. Forged by some of brightest students I have ever met. A few notable mentions include: Sandra Karon, David Reynolds, Eliana Artenyan, and Altamash Mahsud.

OpenAI - ChatGPT ... For the improved workflow in coding and editing.

LaTeX Formatting ... Thank you to Leslie Lamport, Dorab Patel, Eduardo Krell, Richard B. Wales, and John Heidemann for creating a friendly UCLA thesis template.

VITA

2021	M.S. (ECE). University of California, Los Angeles.
2015	B.S. (Physics). University of California, Los Angeles.
2020-2024	Eugene Cota-Robles Fellow, UCLA.
2018-2025	Graduate Student Researcher, Electrical and Computer Engineering, UCLA.
2018-2025	Graduate Student Researcher, Ophthalmology, UCLA.
2024-2025	Graduate Student Researcher, Interventional Radiology, UCLA.
2019-2024	Lead Teaching Assistant, Physics and Astronomy Department, UCLA.
2015-2017	Development Engineer, Physics and Astronomy Department, UCLA.

PUBLICATIONS

Carmona, J., Madruga, B. A., Bentolila, L., Arisaka, K. Transverse-Sheet Illumination Microscopy. Provisional Patent. PCT/US2021/050998, US20230314781A1. (2020).

Carmona, J., et al. Transverse Sheet Illumination Microscopy (TranSIM): A High-Speed Platform for Large-Scale Volumetric Imaging. (2025).

Roberts, A., et al. Induction of Short-Term Sensitization by an Aversive Chemical Stimulus in Zebrafish Larvae. (2020).

Scaramella. C., et al. Bisphenol A Exposure Induces Sensory Processing Deficits in Larval Zebrafish during Neurodevelopment. (2022).

CHAPTER 1

Introduction

1.1 Motivation

In contemporary neuroscience, understanding how biological organisms process the world requires tools capable of capturing neural activity with both high spatial and temporal resolution. For example, model organisms like *Danio rerio* contain approximately 10^5–10^6 neurons during their larval stages, with neural processes operating on millisecond timescales to integrate and react to sensory stimuli [6, 7]. These neurons occupy a 3D volume of 750 μm $\times$ 500 μm $\times$ 250 μm, presenting a formidable challenge to current imaging modalities.

Similarly, *Drosophila melanogaster*, a widely studied model organism, presents unique challenges due to its neuronal density and size. During the larval stage, a *Drosophila* brain spans approximately 590 μm $\times$ 340 μm $\times$ 120 μm and contains about 10^4 neurons [**?**], while the adult brain, with a volume of 900 μm $\times$ 500 μm $\times$ 200 μm, houses nearly 10^5 neurons [8, 9, 10, 11]. These neural circuits operate on millisecond timescales, necessitating imaging tools capable of capturing activity across large volumes with high temporal resolution.

Traditional fluorescence microscopy excels in spatial resolution but often sacrifices temporal performance. In contrast, methods optimized for temporal resolution, such as local field potential recordings, lack spatial comprehensiveness. The ultimate goal of neuroscience imaging remains to visualize large volumes of neural tissue with millisecond resolution, allowing researchers to decipher the dynamics of sensory integration, signal propagation, and motor output on the brain-wide scale. Figure 1.1 illustrates this expansive phase space as a *4D Space-Time* diagram.

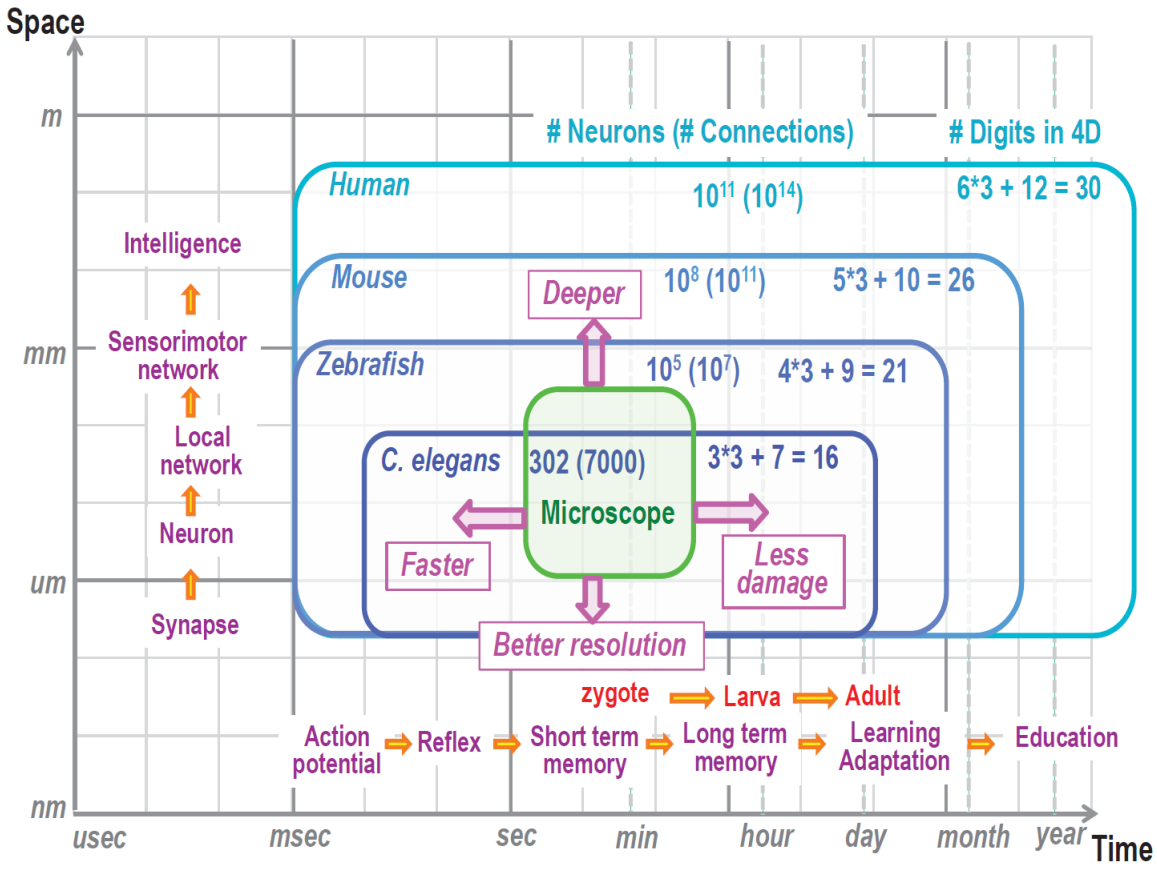

Figure 1.1: *4D Space-Time Diagram of Organisms.* This diagram depicts the brain's spatial and temporal scales in a logarithmic space. From synaptic processes operating on nanometer scales and millisecond timescales to whole-brain dynamics over decades, it highlights the orders of magnitude required to study neural activity comprehensively. Courtesy of Dr. Katsushi Arisaka.

1.2 Transverse-Sheet Illumination Microscopy

To bridge this gap, my master's work introduced *Transverse-Sheet Illumination Microscopy* (TranSIM), a system capable of high-speed volumetric imaging by spatially multiplexing excitation planes and utilizing parallelized detection [12]. TranSIM achieved field-of-view volumes of 748 μm $\times$ 278 μm $\times$ 100 μm at 100 Hz and smaller volumes at up to 200 Hz. By capturing six axially separated imaging planes simultaneously, the system addressed

bandwidth limitations that plague traditional methods while maintaining near-diffraction-limited resolution.

Figure 1.2 illustrates TranSIM's positioning within the microscopy landscape. This system's capabilities fill critical gaps in the *volume-speed* phase space, enabling brain-wide imaging at temporal resolutions sufficient to capture neural dynamics.

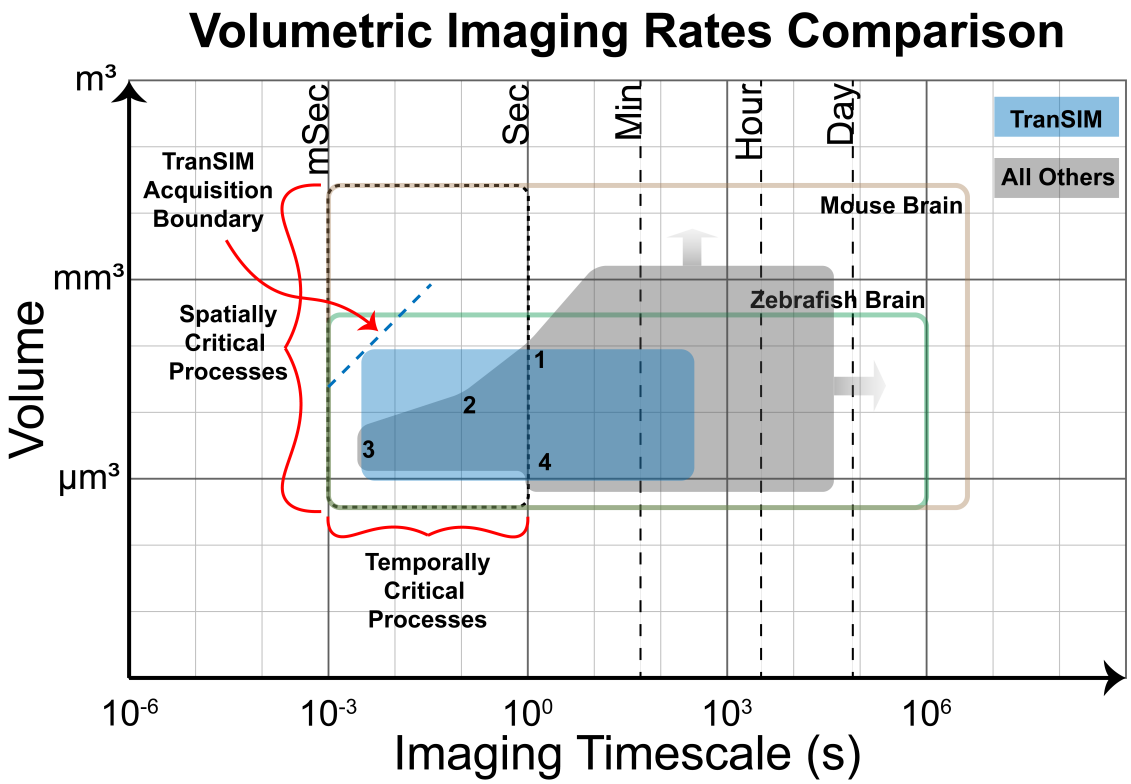

Figure 1.2: *Microscopy 4D Scanning Comparison.* Critical areas of Neuroscience research lie in the study of action potentials/brain waves and large volumetric regions in the brains of organisms. Highlighted in grey are advancements in microscopy have lead to a large exploration in the 4D occupation of brains. *Transverse-Sheet Illumination Microscopy* reaches highly sought regions of exploration in terms of *volume* and *speed*, increasing the total volumetric field of view up to near 1 mm^3, while achieving rates comparable to the fastest methods available, commercially and otherwise, with further room for improvement. For comparison, four regions indicating state-of-the-art research indicate a (1) $\sim$0.1 mm^3 region scanned at 0.8 volumes per second (VPS), Ahrens et al. (2013) [1], fastest volumetric scans; (2) $\sim$0.017 mm^3 at 10 VPS, and (3) $\sim$0.005 mm^3 at 321 VPS, Voleti et al. (2019) [2], and (4) fastest sub-micron resolution, Ahrens et al. (2013) [1] and Chen et al. (2014) [3]. TranSIM was tested up to $\sim$0.07 mm^3 at 200 VPS for a 5 minute period, denoted by the blue rectangle. The blue dashed line indicates the cut-off boundary that TranSIM can operate by changing the volumetric scanning parameters (slowing the scan rate, increasing the volume rate, or vice versa).

1.3 Incorporating Neural Networks into TranSIM

While TranSIM's hardware advancements established a solid foundation for high-speed volumetric imaging, challenges such as sparse sampling, data throughput, and real-time analysis remained significant hurdles. To address these limitations, this dissertation introduces a novel neural network architecture specifically designed to enhance TranSIM's performance. Neural networks have transformed the field of image processing, offering powerful capabilities such as denoising, super-resolution, real-time segmentation, and, critically for TranSIM, virtual refocusing [13, 14, 15].

By seamlessly integrating these advanced algorithms into TranSIM's data processing pipeline, the following advancements were achieved:

1. Sparse to true three-dimensional spatial sampling (2.5D $\rightarrow$ 3D).
2. Enhanced volumetric imaging bandwidth without compromising resolution.
3. Gentler illumination through inherent noise reduction and utilization of out-of-focus information with neural networks.

This work represents a paradigm shift in microscopy, where hardware and artificial neural networks can be combined harmoniously to advance biological science research.

CHAPTER 2

Fluorescence Microscopy Principles

Fluorescence microscopy has emerged as one of the most transformative tools driving the rapid advancements in neuroscience research over the past few decades. Its unparalleled ability to record biological systems with high contrast and spatial resolution has enabled us to explore how large populations of neurons communicate and interact. Despite its significance, I often found it challenging to locate a comprehensive resource that clearly explained the fundamental principles of fluorescence microscopy. This chapter aims to fill that gap, drawing from my own studies and experiences to provide an accessible and cohesive guide to this indispensable technique.

2.1 Fluorescence Microscopy

The phenomenon of fluorescence, first described in 1852 by Sir George Gabriel Stokes [16], refers to the ability of certain materials to absorb radiation at one wavelength and emit it at a longer wavelength. This process, illustrated in Figure 2.1, is governed by three main transitions:

1. Absorption of a photon, exciting the molecule from its ground state (E_0) to a higher energy state (E_n):

$$E_0 + h\nu_{ex} \rightarrow E_\mathrm{n}. \tag{2.1}$$

2. Non-radiative relaxation to the lowest vibrational level of the excited state (E_1):

$$E_\mathrm{n} \rightarrow E_1. \tag{2.2}$$

3. Emission of a photon as the molecule returns to the ground state:

$$E_1 \rightarrow E_0 + h\nu_{em}. \tag{2.3}$$

The energy difference between the absorbed and emitted photons, known as the "Stokes' energy shift," allows for the separation of excitation and emission spectra in fluorescence microscopy [17, 18].

This principle revolutionized microscopy by enabling high-contrast imaging of fluorescent molecules. The discovery of the Green Fluorescent Protein (GFP) by Shimomura et al. in 1962 marked a turning point [19], as it allowed for the non-invasive labeling of biological samples. GFP variants, reactive to calcium ion (Ca^{2+}) concentrations, further transformed microscopy by enabling real-time imaging of dynamic biological processes, giving rise to calcium imaging [20, 21, 22, 23, 24].

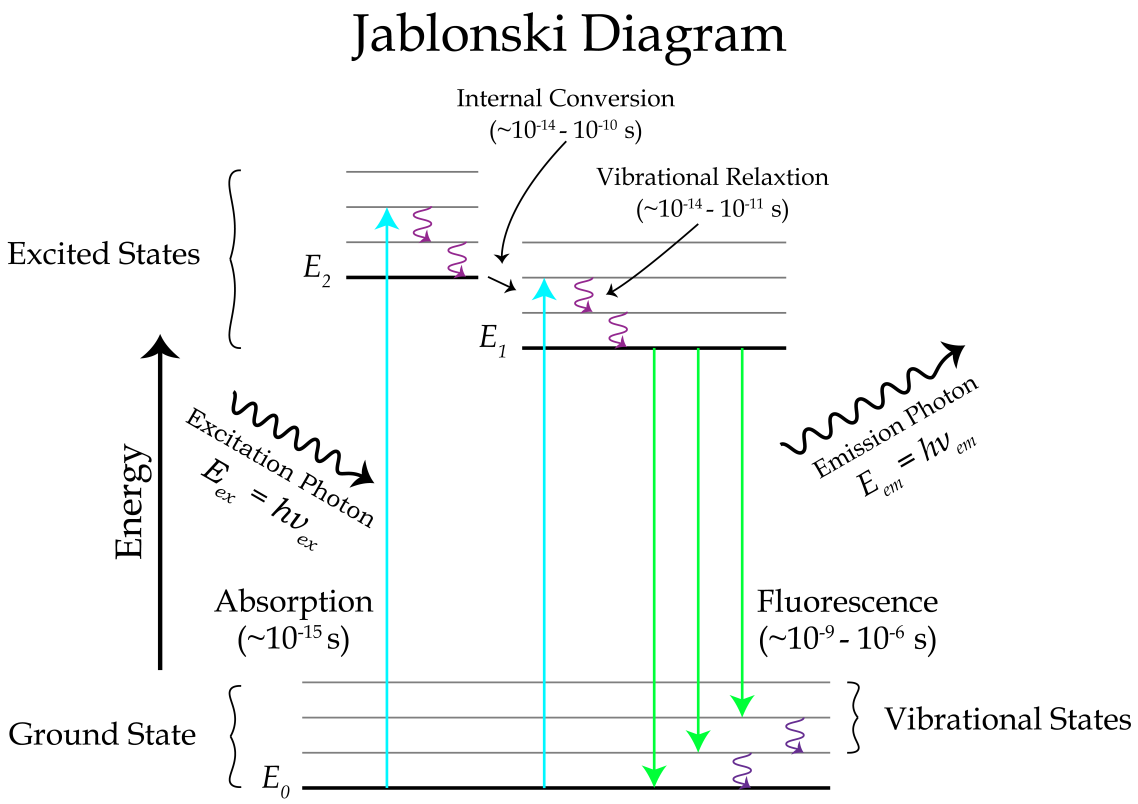

Figure 2.1: *Jablonski Transition Diagram.* The process of fluorescence involves excitation by a photon ($h\nu_{\text{ex}}$), vibrational relaxation within the excited state (E_1), and emission of a photon ($h\nu_{\text{em}}$) as the molecule returns to the ground state. The emitted photon has lower energy (longer wavelength) than the absorbed photon due to energy dissipation as heat or non-radiative processes.

2.2 Spatial Resolution and Numerical Aperture

The resolving power of a microscope is determined by its *Numerical Aperture* (NA), which defines the system's ability to capture light from a wide angular range. A higher NA corresponds to greater light collection and improved resolution. This concept is mathematically linked to the diffraction limit, as described by Born and Wolf [25].

The diffraction-limited resolution is characterized by the formation of an *Airy disk*, resulting from the interference of light waves passing through a circular aperture. The intensity

profile of the Airy disk is given by:

$$I(P) = I_0 \left[\frac{2J_1(ka\omega)}{ka\omega} \right]^2, \tag{2.4}$$

where J_1 is the first-order Bessel function, a is the aperture radius, and k is the wave number. Figure 2.2 illustrates the Airy disk pattern.

The lateral resolution of an imaging system is expressed as:

$$d_{lateral} = \frac{\lambda}{2NA}, \tag{2.5}$$

where λ is the wavelength of light. Similarly, axial resolution, describing the resolving power along the optical axis, is given by:

$$d_{axial} = \frac{2\lambda}{NA^2}. \tag{2.6}$$

These relationships highlight the importance of increasing NA to achieve finer resolution, though this also narrows the depth of field.

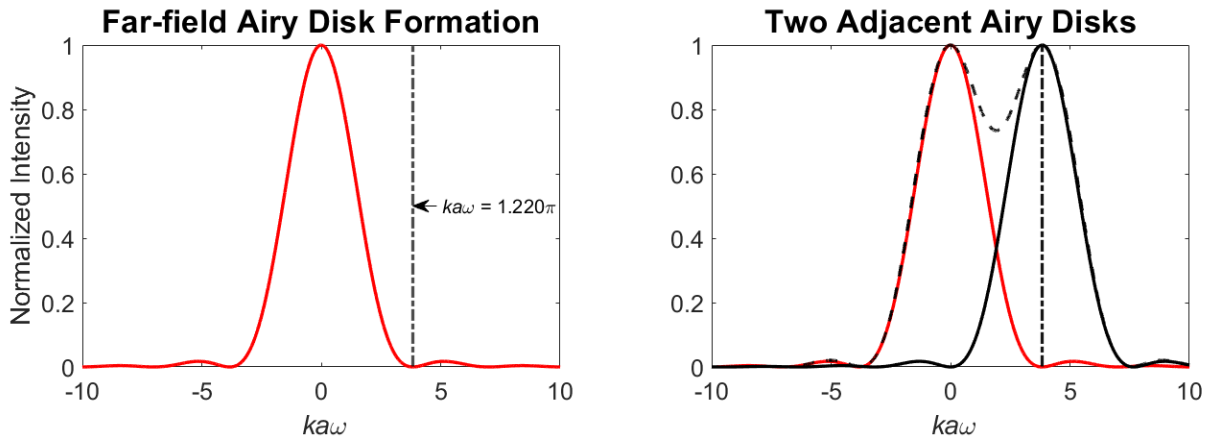

Figure 2.2: *Airy Disk Formation.* The Airy disk pattern results from the diffraction of light through a circular aperture. The resolution limit is reached when the central maxima of adjacent Airy disks are separated by the radius of the first minimum.

Understanding these principles forms the foundation for the development of advanced systems like TranSIM, which leverages high-NA objectives and parallelized imaging to overcome traditional resolution and speed constraints.

2.2.1 Bead Size Correction

The point spread function (PSF) describes the three-dimensional spatial response of an imaging system to a point source. It is a critical parameter in characterizing the resolution of fluorescence microscopy systems. The theoretical lateral and axial full-width at half-maximum (FWHM) values are traditionally determined by the system's numerical aperture (NA) and the wavelength of the illumination or detection light. However, in experimental setups using calibration beads, the bead size introduces additional convolution to the observed PSF. To accurately reflect the system's performance, this effect must be accounted for.

2.2.1.1 Lateral Resolution

The theoretical lateral FWHM ($d_{lateral}$) is defined as:

$$d_{lateral} = \frac{\lambda}{2NA}, \tag{2.7}$$

where λ is the detection wavelength and NA is the numerical aperture. In the presence of calibration beads with diameter d_{bead}, the observed lateral FWHM is modified as follows:

$$d_{lateral,observed} = \sqrt{d_{lateral}^2 + d_{bead}^2}. \tag{2.8}$$

This expression accounts for the additive convolution introduced by the finite bead size. To express the lateral PSF in terms of the Gaussian standard deviation ($\sigma_{lateral}$), the relationship between FWHM and σ is used:

$$\sigma_{lateral} = \frac{d_{lateral,observed}}{2\sqrt{2\ln(2)}}. \tag{2.9}$$

2.2.1.2 Axial Resolution

The axial FWHM (d_{axial}) is influenced by both the detection and illumination NAs, as well as their respective wavelengths ($\lambda_{detection}$ and $\lambda_{illumination}$). Without correction, the axial FWHM is given by:

$$d_{axial} = 2\sqrt{\left(\frac{\lambda_{detection}}{NA_{detection}^2}\right)^2 + \left(\frac{\lambda_{illumination}}{NA_{illumination}^2}\right)^2}. \tag{2.10}$$

To incorporate the bead size into the axial PSF, the observed FWHM is adjusted similarly to the lateral dimension:

$$d_{axial,observed} = \sqrt{d_{axial}^2 + d_{bead}^2}. \tag{2.11}$$

The corresponding Gaussian standard deviation (σ_{axial}) is:

$$\sigma_{axial} = \frac{d_{axial,observed}}{2\sqrt{2\ln(2)}}. \tag{2.12}$$

2.3 The Modulation Transfer Function (MTF)

The *Point Spread Function* (PSF) describes how an optical system responds to a point source of light. It is the fundamental descriptor of resolution in imaging systems, characterizing the spreading of light from an idealized point source due to diffraction and aberrations. Mathematically, the PSF is the system's impulse response in the spatial domain. However, in practice, it is often more useful to consider its counterpart in the frequency domain, the *Modulation Transfer Function* (MTF).

The MTF is defined as the Fourier Transform (FT) of the PSF and represents how different spatial frequencies are transmitted by the optical system. The MTF is a crucial metric for evaluating system performance, as it quantifies the contrast degradation at varying spatial frequencies. High spatial frequencies correspond to fine details, while low spatial frequencies represent broad features. Mathematically:

$$\text{MTF}(f_x, f_y, f_z) = |\mathcal{F}[\text{PSF}](x, y, z)|, \tag{2.13}$$

where f_x, f_y, and f_z are the spatial frequency components in the x, y, and z directions, respectively.

2.3.1 Nyquist Criterion and Sampling in Fluorescence Microscopy

The *Nyquist criterion* establishes the minimum sampling rate required to accurately represent a signal without aliasing. In the context of imaging, the Nyquist rate is twice the highest spatial frequency transmitted by the optical system. For a microscope, the Nyquist

sampling interval s is given by:

$$s \leq \frac{1}{2 \cdot f_{\text{cutoff}}}, \tag{2.14}$$

where $f_{\text{cutoff,lateral}} = \frac{2 \cdot NA}{\lambda}$ is the lateral cutoff frequency of the system, and $f_{\text{cutoff,axial}} = \frac{NA^2}{2 \cdot \lambda}$ is the axial cutoff frequency.

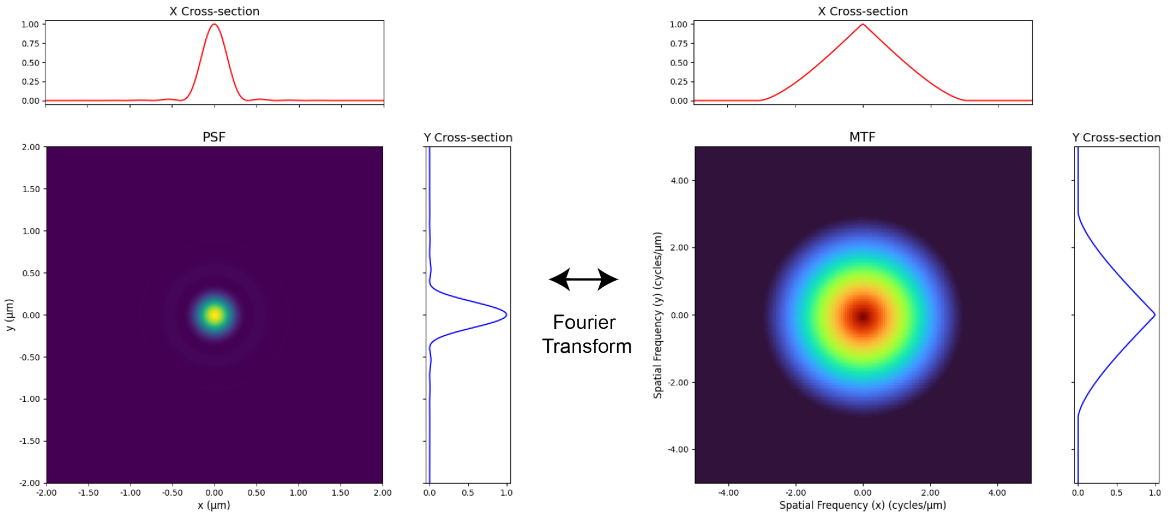

Figure 2.3: *Relationship Between PSF and MTF* (**A**) The PSF describes the system's spatial response to a point source. (**B**) The MTF, obtained as the Fourier Transform of the PSF, quantifies how contrast is transmitted at different spatial frequencies. Sampling below the Nyquist rate introduces aliasing, distorting high-frequency details. Sampling at or above the Nyquist rate ensures accurate representation of the image.

In turn, the sampling rate is determined by the pixel size of the sensor and the effective magnification of the system M. Sampling at, below, or above the Nyquist rate has distinct implications:

- **Below Nyquist (Under-sampling):** Spatial frequencies higher than the sampling frequency are aliased, introducing artifacts and distortions. This degrades image quality and can obscure fine details.

- **At Nyquist:** The minimum sampling rate ensures that all resolvable spatial frequencies are represented without aliasing. This is ideal for preserving image fidelity while minimizing data redundancy.
- **Above Nyquist (Over-sampling):** Sampling at a higher rate reduces aliasing risk but increases data storage and processing requirements. Beyond a certain point, over-sampling provides diminishing returns.

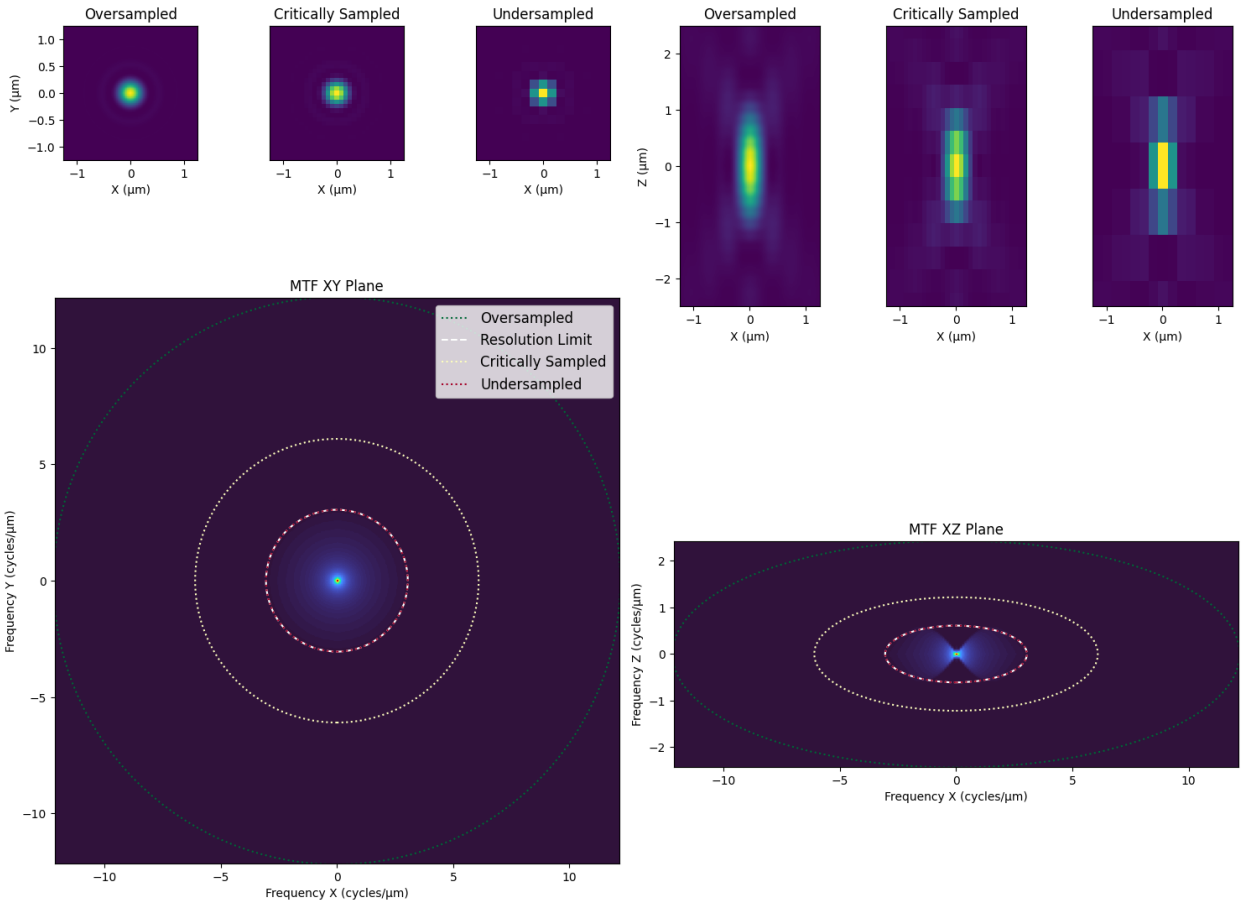

Figure 2.4: *Impact of Sampling on PSF and MTF.* (**A**) The first row shows the PSF cross–sections in the XY and XZ planes for three different sampling conditions: oversampled, critically sampled, and undersampled. The PSF represents the spatial response of the imaging system to a point source, with finer sampling improving resolution fidelity. (**B**) The second row shows the MTF cross-sections in the XY and XZ planes and optical resolution limits (dashed white lines). The MTF quantifies the ability of the system to preserve spatial frequency contrast, with undersampling introducing aliasing artifacts and oversampling reducing data efficiency.

2.4 Implications for Microscope Design

In fluorescence microscopy, the interaction between PSF and MTF determines the resolution and contrast of the system. To achieve optimal imaging performance:

- High-NA objectives improve both the lateral and axial resolution, extending the cutoff frequency.
- Proper sampling ensures an accurate representation of spatial frequencies without aliasing.
- Trade-offs can be implemented with microscopy design to ensure that parameters such as resolution, field of view, and spatial sampling are correctly balanced to achieve the desired outcome.

This foundational understanding underpins the design of systems like TranSIM, which leverage high NA, a prudent choice in spatial sampling, and parallelized imaging to address traditional resolution and speed limitations.

CHAPTER 3

Spatiotemporal Advancements in Microscopy

With the advent of Fluorescence Microscopy, a flurry of unforeseen levels of utility came into play from the tool. Finally, with the introduction of GFP, living cell dynamic imaging took life. Since then, a large portion of the development in light microscopy imaging has been in speeding up the tool to image faster and faster dynamics. For example, complex organisms like *Danio rerio*, Zebrafish, a vertebrate animal, has neural dynamic processes on the order of 1-10s of milliseconds, Eimon et al. (2018) [26]. For cardiovascular processes, like heart rate, those can beat at 120–180 bpm, Sampurna et Al. (2018)[27]. Currently, there are a lot of tools that can image at these speeds with relative ease, however, the caveat is that most methods can only do so in 2D fields of view only. Nevertheless, when imaging only in 2D, a great deal of information is lost because these are inherently 3D structures that are being studied. For that reason, people have been trying to increase not only spatial resolution of microscopes, but also temporal resolution in 3D.

3.1 Confocal Microscopy

When light illuminates the entire field without, this type of microscopy is known as *Widefield*. System wise, this is the most convenient method illuminate a sample due to its simplicity as shown in Figure 3.1. However, this type of illumination does come with its own drawbacks. Due to the decrease depth of field produced by large NA systems, most of the light that is being captured is out of focus light and therefore contaminates the light would otherwise follow resolution limit expressions.

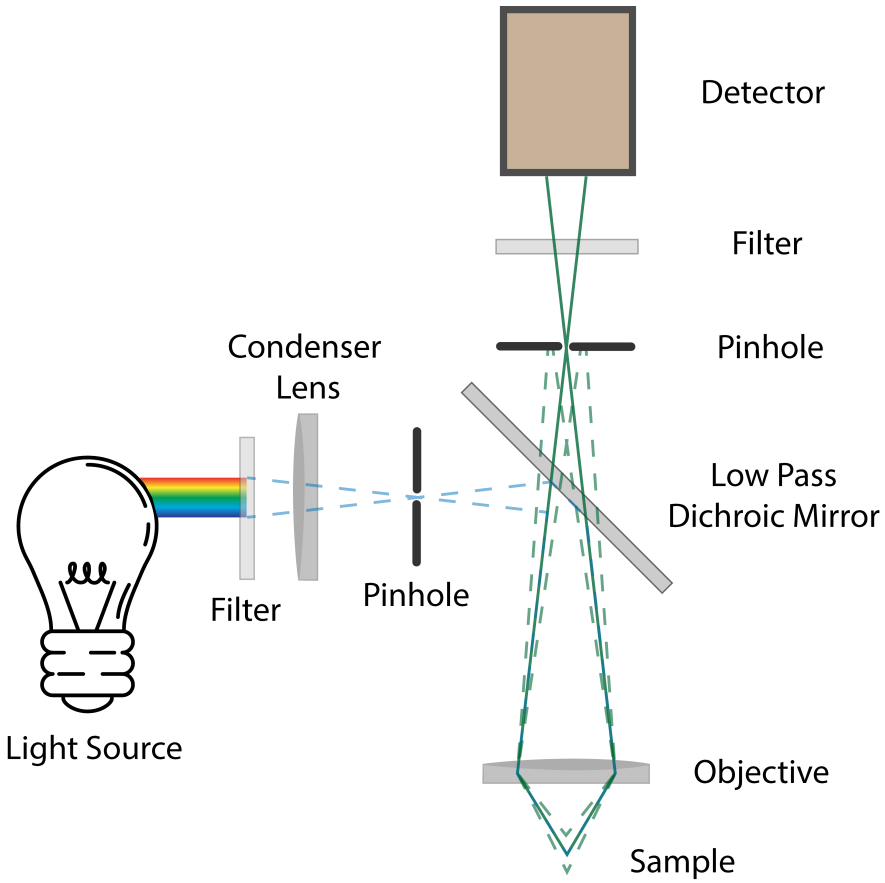

Figure 3.1: *Point Confocal Microscopy.* Invented by Marvin Minsky, in 1957, the point confocal microscope clarified a fuzzy Widefield traditional method by rejecting out of focus light.

In 1957, Marvin Minsky, came up with an ingenious method by which to remove this contaminating, out of focus light [28]. He noted, that at the imaging plane, while most of the light was was out of focus, the image was still being formed at or near its theoretical limit. One could reject the out of focus light simply by placing a pinhole in the formed image location equivalent to the object illuminated location. Due to the point confocal nature of the design, new innovative systems have since invented that allow for the image of larger areas by either translation of the sample on a motorized stage or by deviating the light source with motorized deflection mirrors which in most cases now is a coherent light source (laser). Eventually, these were able to produce up to a few images per second. However,

they remained bandwidth limited by the point scanning nature of the system.

It can been from the Airy disk formation figure that most of the intensity is contained within the radius of the first dark ring. In fact, if one were to model that first lobe as a Gaussian distribution, one can see that at,

$$(ka\omega)_{cutoff} = \frac{1.220\pi}{2}, \tag{3.1}$$

$$\omega_{confocal} = \frac{0.610}{2}\frac{\lambda}{a} = 0.310\frac{\lambda}{a}. \tag{3.2}$$

half the distance to the first dark ring, approximately 70% of the information is retained. Here is where by placing a pinhole increases the overall resolution of the system. Not only does the pinhole remove widely distributed the outer rings of the Airy disk, but also decreases the overall width of the central lobe, resulting in a increase of lateral resolution by a factor of 2. See Figure 3.2.

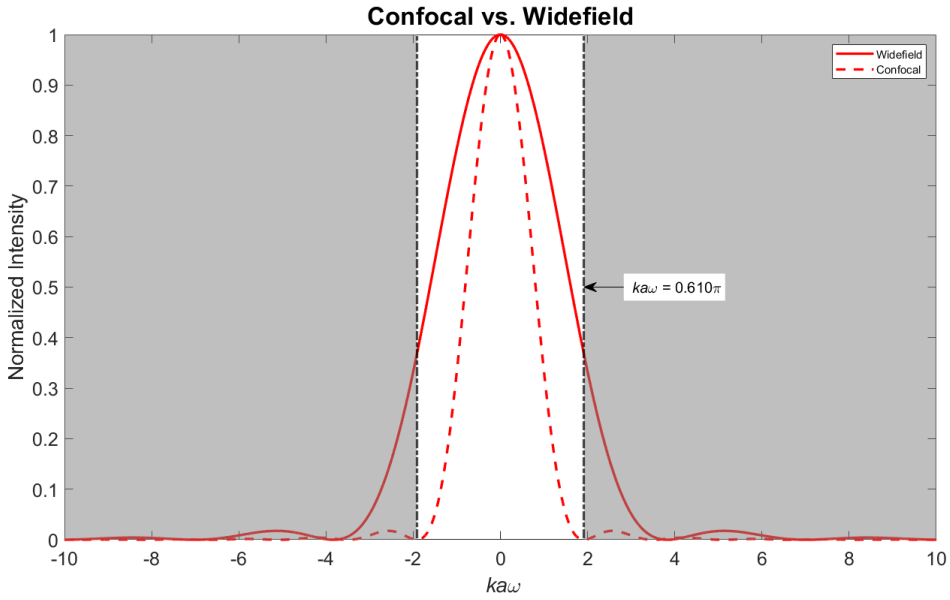

Figure 3.2: *Widefield vs. Confocal Comparison.* By placing a pinhole at approximately half the distance to the first dark lobe, the overall lateral resolution of the system can be increase by a factor of 2 while retaining nearly of the input intensity. The solid line indicates what a widefield system would see. The dashed red line indicates the compression of the first lobe down by a factor of 2 after applying a confocal pinhole.

3.1.1 Spinning Disk Confocal Microscopy

To increase the speed, David Egger and Mojmír Petráň, shortly after the invention of the point confocal, in 1967 took and old device called the Nipkow disk and reapplied to the recently invent point confocal technique [29]. This move parallelized the point-by-point process such that images capture increase several order of magnitude. Currently, state-of-the-art systems, like the Yokogawa Spinning Disk Confocal module (CSU-X1)[30] can routinely take images up to 2000 frames per second, attached to the appropriate sensor.

3.1.2 Line-Scanning Confocal Microscopy

Another method to parallel process the acquisition of data that has confocality in mind is Line-Scanning Confocal Microscopy (LSCM) [31, 32]. Instead of acquiring data on a point-by-point basis, LSCM utilizes a standard feature in modern scientific Complementary Metal–Oxide–Semiconductor (sCMOS) cameras, like the Hamamatsu Flash 4.0 v2 (Hamamatsu Photonics K.K., Hamamatsu City, Japan), which can activate only a subset of pixels and not allowing the adjacent pixel rows to collect photons. Each pixel column (times active pixels) is routed to its own Analog-to-Digital (A/D) converter. This has the same effect that a pinhole would have, but only in one dimension. Using a scanning galvanometer, a laser line can be synchronized with the rolling shutter. See Figure 3.3.

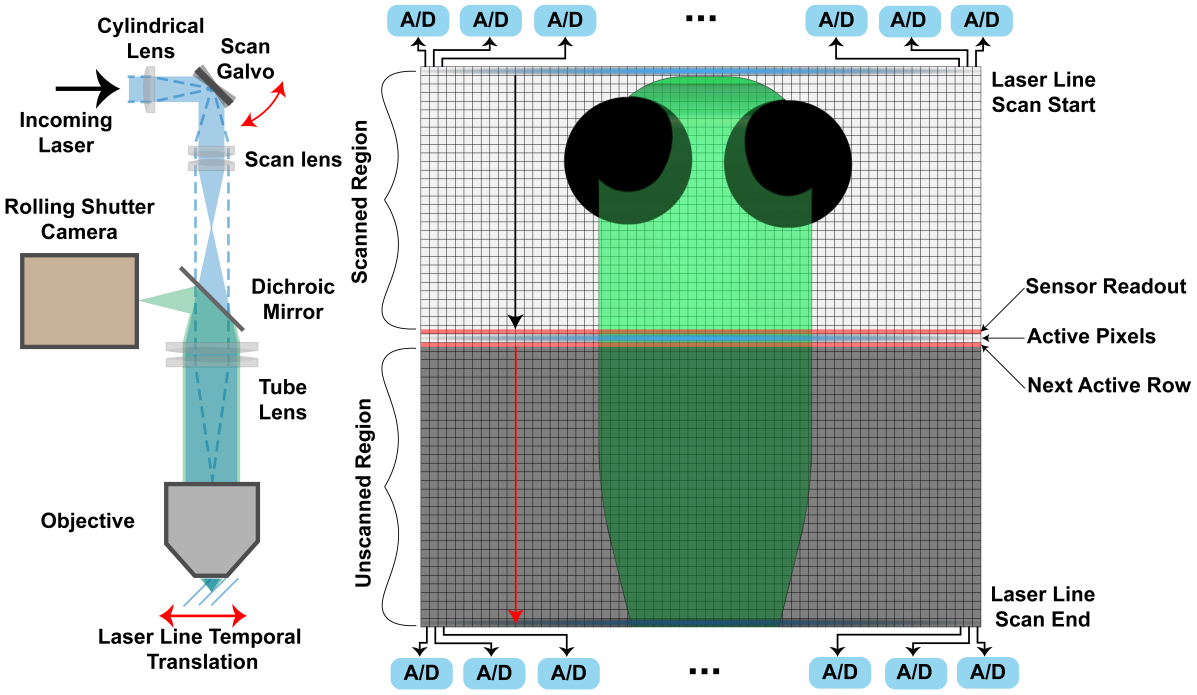

Figure 3.3: *Line-Scanning Confocal Microscopy.* By focusing light down to a line using a cylindrical lens, that scanning of that line can be synchronized with the rolling shutter of modern sCMOS cameras. The net effect is a system that is half widefield and have confocal. In the direction of the shutter, confocality exists whereby the rolling shutter rejects out of focus light, while in the orthogonal direction, widefield is present.

The resulting point spread function (PSF) is a convolution of the point confocal PSF in the X-dimension and a widefield PSF in the Y-dimension, where the PSF is elongated along the direction of the active pixel rows,

$$d_{lateral,x} = 0.310\frac{\lambda}{NA}, \tag{3.3}$$

$$d_{lateral,y} = 0.610\frac{\lambda}{NA}, \tag{3.4}$$

Although the overall PSF has increased in size, the benefits gained can arguably out weigh the negative side effects. Namely, the speed of the system can be increased dramatically. sCMOS cameras, like the Hamamatsu Flash 4.0 v2 have "light-sheet" modes that can acquire full frame images of 2048 pixels $\times$ 2048 pixels at 50 FPS. Newer sensors has the capability to capture in this mode with even larger sensor areas at 100 FPS (Hamamatsu Flash Fusion) [33] or even 121 FPS (Hamamatsu Flash Lightning) [34] under its rolling shutter mode. Previously, the bottleneck of confocal systems was the scanning speed by which you could scan and area point by point. By parallel processing the illumination and acquisition in 1 dimension, the entire process was sped up at the cost of some lateral resolution.

3.2 Light-Sheet Microscopy

By illuminating orthogonal to the detection axis, the main issues that plague single objective imaging, namely: phototoxicity, photobleaching and optical out of plane light contamination, to name a few. The first iteration of a light-sheet microscope used a cylindrical lens to create the sheet, which results in approximately a thin widefield section equivalent lateral resolution. Lately, innovative techniques have unfolded so further increase the resolution. Similarly to LSCM, a thin "pencil" like Gaussian beam can be created using a spherical lens and translated along the detection plane using a galvanometer. Ahrens et al. (2013)[1] managed to image Zebrafish in whole at 0.8 Hz using a more sophisticated version using dual detection and illumination objectives, which they named IsoView. With this illumination methodology, a confocal conjugation with the imaging sensors rolling aperture can be used,

thereby increasing the resolution in the lateral and axial dimensions. See Figure 3.4.

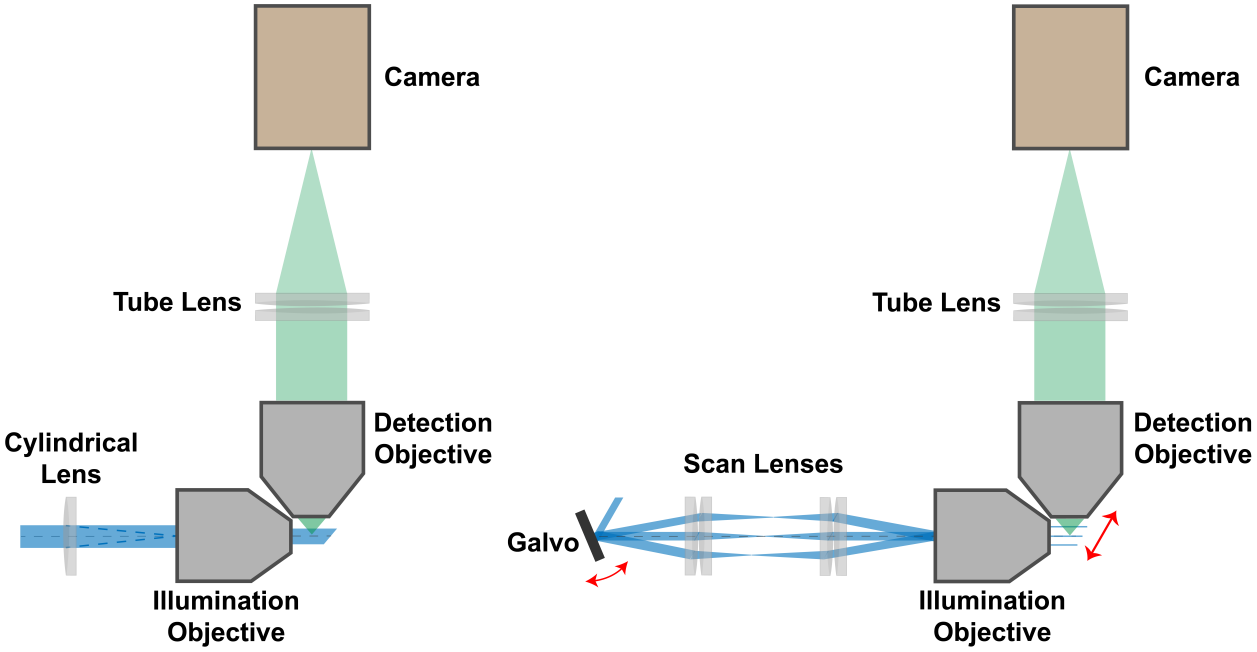

Figure 3.4: *Light Sheet Microscopy.* Conventional light-sheets (*left*) are created using a cylindrical lens to focus the light only in one dimension. To create "pencil-like" beams, *right* that are instead scanned confocally with the rolling sCMOS sensor, the methodology is to use a regular spherical lens, rather than a cylindrical lens.

Even further, sub diffraction limit beams have been created by implementing Airy beams, Vettenburg et al. (2014)[35], non-diffracting beam profiles like Bessel-Gauss beams and optical lattices of said beams which are formed by conical phase profiles, Betzig (2005) and Chen et al. (2014) [36, 3]. See Figure 3.5. The general purpose for this type of beams is not only their resolution improvements, but also because the beams propagate for longer distances than standard Gaussian beams are the same numerical aperture focusing angle.

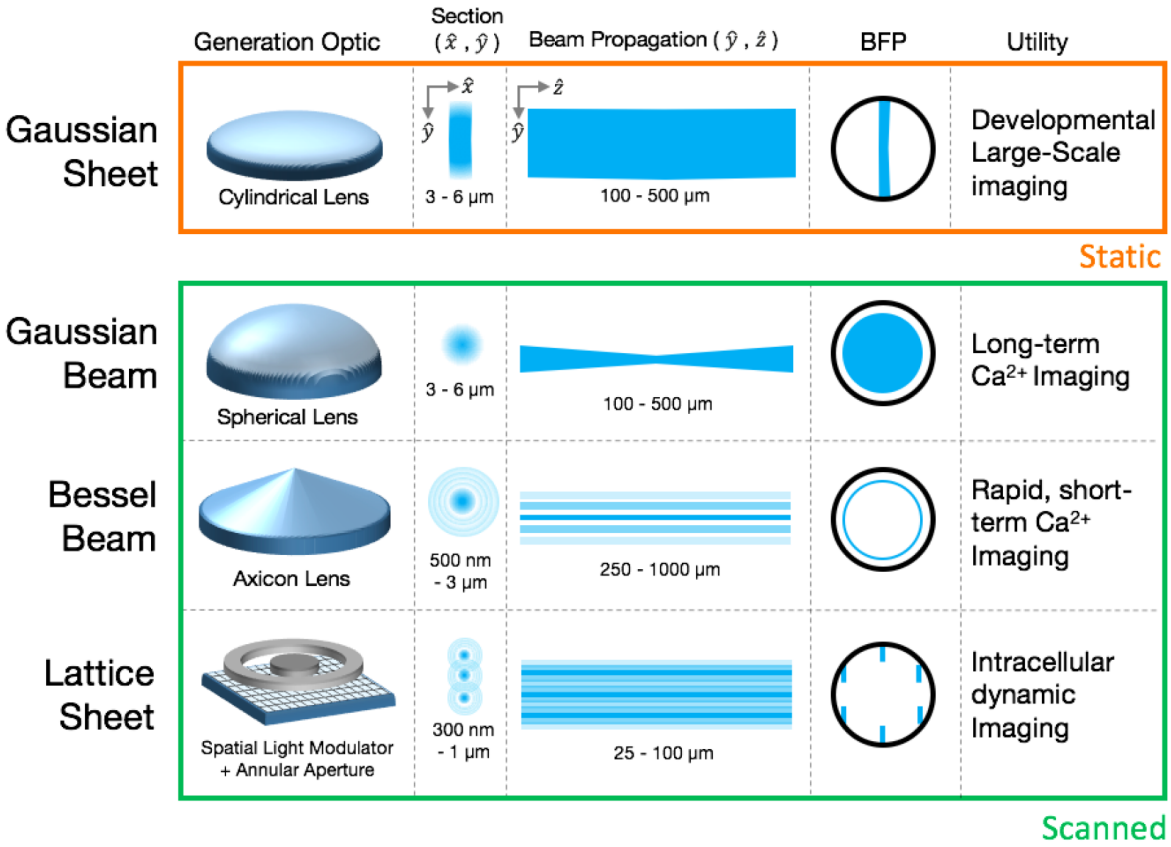

Figure 3.5: *Comparison of Light-Sheet Illumination Beams.* The first and standard beam illumination method is created using a cylindrical lens which collapses a gaussin plane wave in one dimension. The typical propagation length is 100-500 μm, however this type of beam does not afford any confocality application to a rolling shutter sCMOS sensor. For confocality, a spherical lens can be used which collapses the beam in both dimensions with the same propagation distance. Non-diffracting beams created by interference, Bessel-Gauss beams, have a much larger propagation distance, 250-1000 μm, compared to same numerical aperture collapsed Gaussian Beams. With the aid of a Spatial Light Modulator (SLM), an array of Bessel beams can be configured to interfere even further by trimming away at each adjacent beams thereby reducing the overall waist of the beams. Courtesy of Blake Madruga, adopted from his Thesis work on Configurable Bessel-Gaussian Sheet Illumination Microscopy[4].

3.3 Scanned Volumetric Imaging

In the previous sections, discussion was focused on 2D imaging by means of widefield, confocal, or light-sheet illumination. However, due to the minuscule depth of field that large numerical aperture objectives provide (on the order of microns), and the inherent need to image in 3D, methods have been developed image make this possible. There are 2 ways by which this can be achieved according the previous research. The first method is to translate the sample such that is moves in and out of the focal plane of the objective. The second method is to move the objective such that the objective's focal plane is translated relative to the sample. See Figure 3.6.

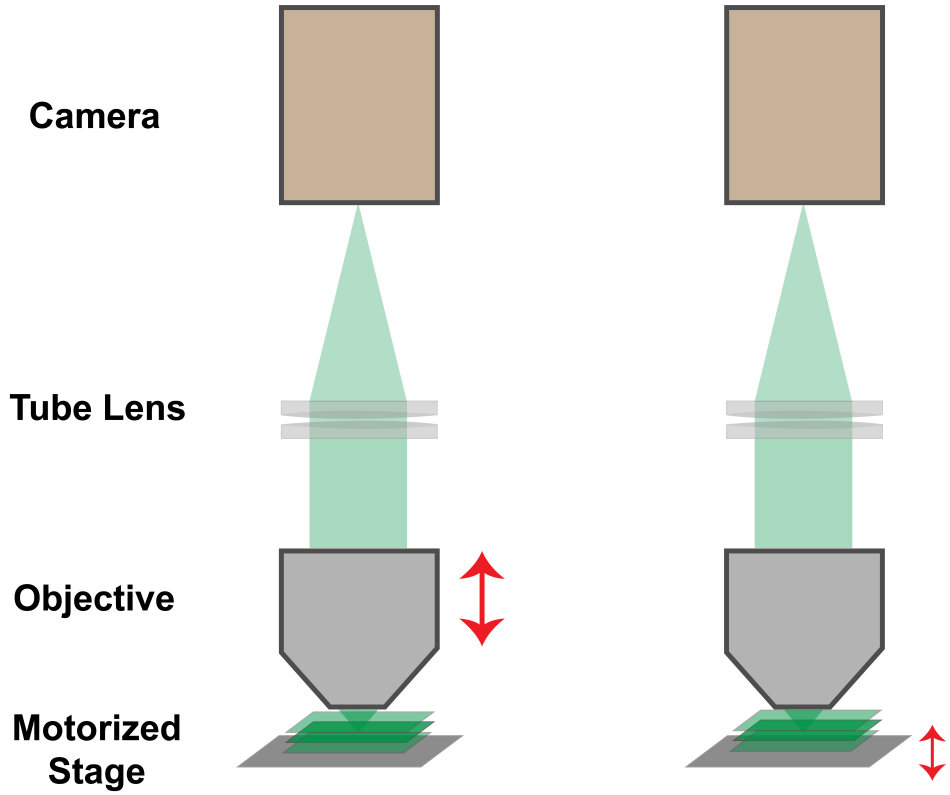

Figure 3.6: *3-Dimensional Scanning.* Two main methods for mechanic depth scanning are the translation of the objective using a piezoelectric actuator (left) and movement of the sample stage using a motorized stage (right).

3.4 Static Volumetric Imaging

Recently, a push has been made to create methods by which the slow translation of an imaging objective or sample can be avoided. The physical translation is an unavoidable requirement due to the highly compressed depth of field associated with high numerical aperture systems. Nevertheless, new methods have surfaced by which to bypass this limitation.

3.4.1 Light Field Microscopy

Standard microscopy methods rely on the capture of incident light onto a sensor. This is usually comprised on 2D information only (stemming from the fact that high NA systems have short depth-of-fields), the rest of the information has or has not already angular converged into an image. For example, an object that is further away from the objective focal plane will have already converged. Conversely with objects closer to the focal plane; they have yet to converge. With the placement of a *microlens array* at the formed image plane, the objects in the objective focal plane that are exactly correlated with the image plane while transmit through only one of the microlenses. However, if they have not yet converged or have already converged will pass through multiple microlens arrays. Now, by placing a detector at the back focal plane (BFP) of the of the microlens array, there will be product of multiple microlens BFP each containing different 2D angular information from the entire volume being imaged. This aforementioned method is known as *Light Field Microscopy* (LFM), Levoy et al. (2005) [37]. Through computational deconvolution, based on the field theory of light, the entire volume can be reconstructed [38, 39, 40]. Limitations of this system lie in the heavy computational resources required to perform this deconvolution, taking somewhere on "the order of seconds to minutes" for each volume being reconstructed [41, 42, 43, 44]. When dealing with possible dynamic processes that are on the order of milliseconds and recordings are preferred to last up to minutes or even hours, the times required to deconvolve an entire data-set become staggering.

3.4.2 Oblique Plane and SCAPE Microscopy

In 1P systems, oblique imaging techniques, also known as olbique plane microscopy (OPM), have been developed to reduce the need for axial scanning, a key bottleneck in conventional microscopy[45, 46, 47]. By tilting sheets of light at an oblique angle, fluorescence can be simultaneously illuminated and detected using the same objective lens, like traditional orthogonal geometry systems. This approach eliminates the need for mechanical movement of the objective, significantly increasing volumetric imaging speeds. With an emphasize being place on a purely optical methodology by which to obtain 3D images, Bouchard et al. (2015)[48], and most recently Voleti et al. (2019)[2] have created a novel system that does not rely on the translation of on object. Rather, it exploits an oblique light sheet illumination and detection scheme in order to capture volumes. Swept confocally-aligned planar excitation (SCAPE) microscopy, see Figure 3.7, is a single object light sheet system that has roots in both LSCM and LSM. By illuminating at an oblique angle, the once convolved depth information, now becomes visible without the need for deconvolution. This oblique light sheet produces fluorescence that is nearly orthogonal in the opposite half of the objective. The signal returns via the same pathway and is descanned by the once scanning galvanometers. Once descanned, the static image is relayed through a pair of objectives where the second objective in the pair looks at the formed image from an oblique angle. The oblique angle is orthogonal angle of tilt that the light sheet was produced. The result is a flat image plane captured by this secondary objective. Overall the system is a very clever way of realigning the image plane such that it is once again flat on the imaging sensor. However, the cost of this is technique is a reduced numerical aperture, approximately 50% reduction in effective NA. Results indicate that when partnered with a HiCAM Fluo intensified CMOS camera (Lambert Technologies, LLC Kissimmee, FL, US), volumes were generated of up to 321 VPS for voxel regions of 57 pixels $\times$ 640 pixels $\times$ 100 pixels for a maximum bandwidth of 1.2 GPixels/second.

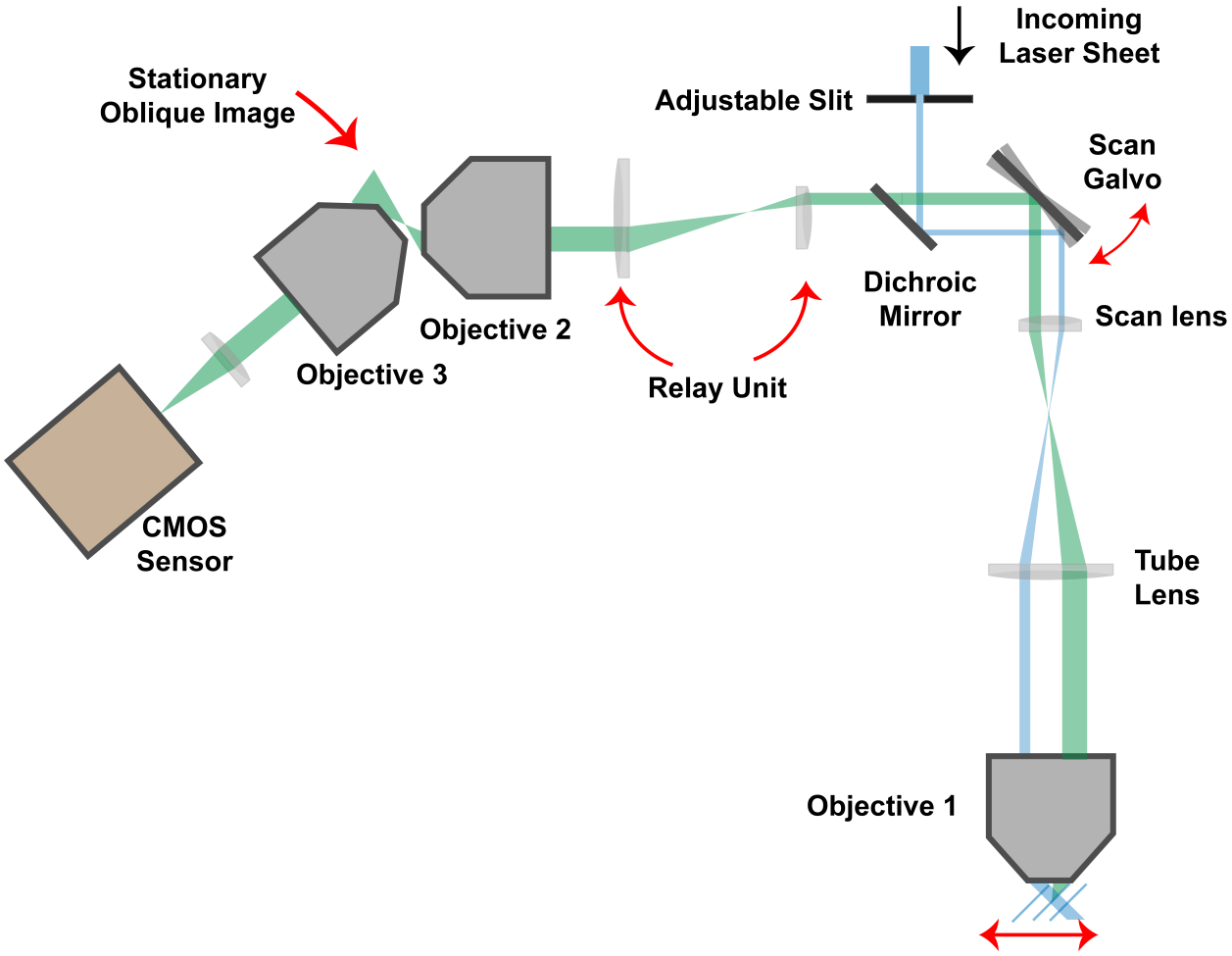

Figure 3.7: *SCAPE Microscopy.* With a single objective, SCAPE microscopy achieves depth resolved images. To do so, an oblique illumination sheet is implemented. On the detection axis, only the orthogonal directional fluorescence is utilized. After descanning, the descanned plane, is relayed to a secondary objective whereby utilizing a tertiary, the image plane is mapped onto a flat image plane and onto an imaging sensor.

CHAPTER 4

Deep Learning and Microscopy

Deep learning has fundamentally transformed fluorescence microscopy by offering powerful solutions to challenges that traditional methods struggle to address, such as high-throughput imaging, accurate segmentation, and three-dimensional (3D) reconstructions from sparsely sampled data. Recent efforts have demonstrated that convolutional neural networks (CNNs), see Figure 4.1, can significantly improve spatial resolution, segment data, denoise complex biological images, and even perform 3D refocusing without additional mechanical scanning [49, 50, 51, 52, 53] . These methods have had a substantial impact on various applications, from single-molecule localization to high-speed volumetric imaging.

Nonetheless, there remain persistent challenges in fluorescence microscopy that deep learning is well-positioned to solve. Rapid volumetric acquisition often yields under-sampled or noisy data, as increasing acquisition speeds can compromise both signal-to-noise ratio and axial resolution. Moreover, large-scale imaging of dynamic processes (e.g., neural activity or cardiac motion) requires sophisticated computational tools that can preserve fast-evolving biological features without distorting critical structural information. Such constraints are particularly relevant for emerging techniques like Transverse-Sheet Illumination Microscopy. By capturing multiple axial planes in parallel, TranSIM relieves many bandwidth bottlenecks, but the resulting volumes are sparsely sampled along the optical axis, demanding advanced reconstruction algorithms, and that is the aim of this work. This shows the importance of deep learning as an integral component of modern microscopy pipelines—especially for methods seeking to combine large fields of view, high volumetric rates, and near-diffraction-limited resolution.

In the sections that follow, we first examine how deep learning techniques for segmenta-

tion and localization (Section 4.1) have evolved to meet the unique demands of microscopy. We then move on to explore restoration and denoising (Section 4.2) and most relevant for this work, volumetric reconstruction (Section 4.3). Finally, we highlight their relevance to TranSIM's high-speed, sparse-acquisition paradigm.

4.1 Segmentation and Localization

Accurate segmentation of cells and subcellular structures forms the backbone of quantitative fluorescence microscopy. Traditional image-processing pipelines relied on thresholding or region-growing techniques that often fail when images are noisy or densely populated with overlapping objects. Deep learning, in particular the U-Net architecture, revolutionized biomedical segmentation by introducing skip connections and data augmentation strategies that excel with relatively small training sets [54]. The success of U-Net in 2D contexts quickly led to adaptations for 3D data, allowing segmentation of volumetric fluorescence datasets where tissue sections or entire organs are imaged in high detail [**?**].

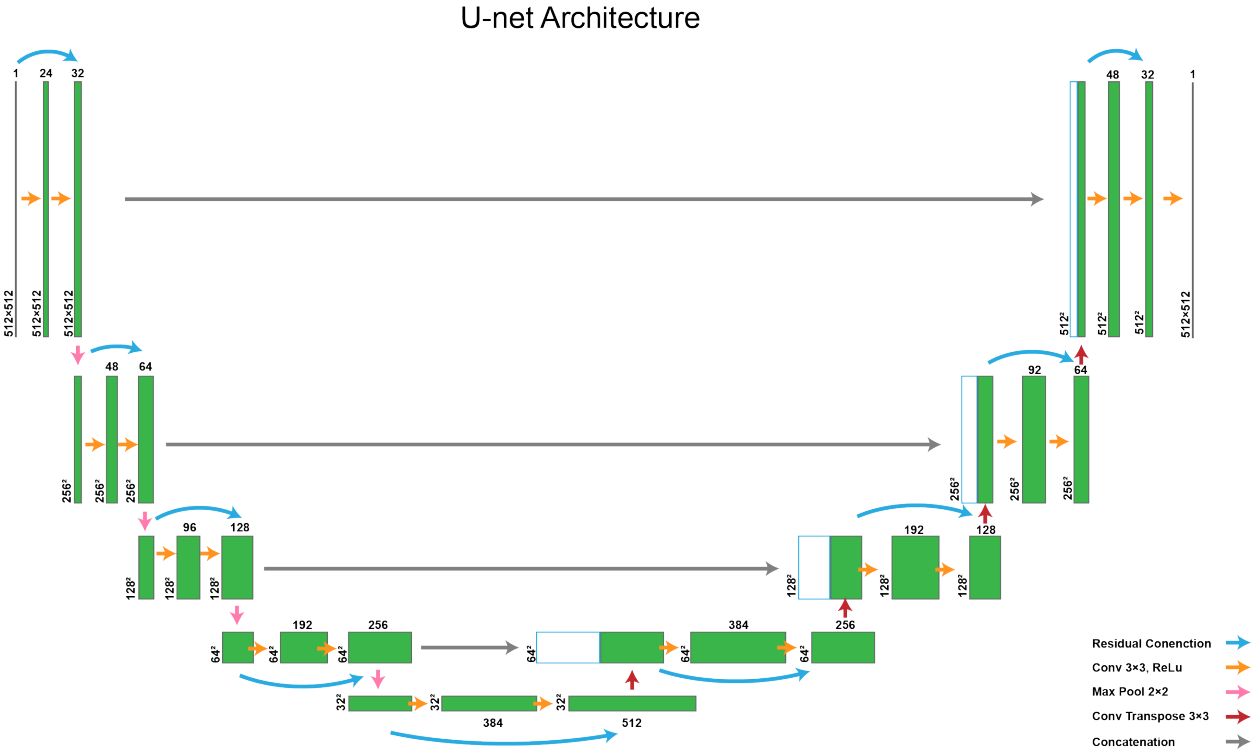

Figure 4.1: *A schematic representation of the U-Net architecture.* The network consists of a contracting path (left) for feature extraction and an expansive path (right) for reconstructing spatial details. The down-sampling and up-sampling paths are known as the encoder and decoder paths, respectively. Skip connections link corresponding layers in the contracting and expansive paths, enabling the preservation of spatial information lost during down-sampling. Each block represents a multi-channel feature map, with the number of channels increasing in the contracting path to capture complex features and decreasing in the expansive path to refine spatial resolution. Arrows denote convolutional, pooling, and up-sampling operations. This architecture enables precise pixel-level segmentation, even with limited training data.

Beyond segmentation, localization of individual fluorescent emitters has also benefited significantly from deep learning. In single-molecule localization microscopy (SMLM), CNN-based approaches can process frames at high densities of active fluorophores, overcoming fundamental limitations in traditional point-spread-function fitting methods [**?**]. For instance, network architectures inspired by U-Net or residual networks can separate closely

spaced emitters and produce more accurate position estimates at higher throughput. These networks learn to recognize the subtle diffraction patterns associated with single molecules, offering not only speed advantages over iterative fitting algorithms but also improved robustness to noise or background fluorescence.

As high-speed 3D acquisition becomes more accessible, the same principles applied to segmentation and localization in 2D extend to volumetric data. Variations of U-Net or other CNN backbones incorporate 3D convolutions and specialized layers to handle multi-plane inputs, enabling the simultaneous detection of cells, nuclei, or synapses throughout thick biological specimens [**?**].

4.2 Denoising and Image Restoration

Fluorescence microscopy images frequently suffer from noise and other degradation artifacts arising from optical imperfections, photobleaching, and limited photon budgets. Traditional denoising algorithms, such as median filtering or wavelet-based methods, often struggle to preserve fine structures in highly noisy or under-sampled datasets [**?**]. In contrast, deep neural networks (DNNs) have proven to be remarkably effective at learning complex noise distributions and restoring high-fidelity images, even under challenging imaging conditions.

One prominent example is the Content-Aware Image Restoration (CARE) framework, which employs a convolutional neural network (CNN) trained on pairs of low- and high-quality fluorescence images [55]. This supervised approach models both signal and noise characteristics by directly learning a mapping from noisy inputs to clean targets. By leveraging spatial context and multi-scale feature extraction, CARE preserves important biological structures while efficiently suppressing noise. Other supervised strategies similarly rely on paired training data, where ground-truth images are experimentally obtained via long-exposure acquisitions, physical averaging, or more advanced imaging systems.

However, obtaining high-quality reference images for supervised training can be labor-intensive or impractical, especially for dynamic biological samples. To address this challenge, self-supervised methods, such as Noise2Void and Noise2Self, have gained popularity for their

ability to denoise fluorescence images without explicit ground-truth references [51, 52]. These algorithms exploit the inherent redundancy in image data by masking or perturbing subsets of pixels during training. The network learns to predict the masked regions based on surrounding pixel intensities, effectively reducing noise while preserving fine-scale details. Such methods offer a powerful alternative when exhaustive labeled data is unavailable or difficult to acquire.

In addition to these approaches, specialized networks have been developed to tackle denoising in more specific application contexts. For example, DeepCAD and its real-time variant DeepCAD-RT have shown promise in providing high-quality denoised images for live-cell imaging scenarios, where rapid data acquisition and on-the-fly processing are critical [56]. By leveraging domain-specific training and optimized architectures, DeepCAD-based methods are able to remove noise while preserving intricate cellular structures, making them particularly suitable for fast biological experiments that cannot accommodate extensive averaging or long-exposure reference images.

Across these various strategies—supervised, self-supervised, 2D, 3D, or real-time—the core principle remains consistent: deep networks can learn sophisticated mappings between degraded input images and their high-fidelity counterparts. This learned ability is instrumental for a range of microscopy modalities, from single-molecule studies to large-volume tissue imaging. While the architectures and training paradigms differ, the end goal is unified: to recover the most accurate and biologically meaningful representation of a sample, using as few raw photons as possible. By reducing noise, these networks not only enhance visual clarity but also improve downstream quantitative analyses, such as cell tracking, morphological assessments, and the identification of subtle phenotypic changes.

4.3 Volumetric Reconstruction in Fluorescence Microscopy

Recent research has demonstrated that convolutional neural networks can be used to virtually refocus fluorescence images and reconstruct volumetric data with minimal reliance on mechanical scanning. Shin et al. presented a framework termed Recursive Light Propagation

Network (RLP-Net)[57], which learns short-distance axial refocusing and then recursively applies the same learned inference to project a two-dimensional fluorescence image into multiple planes. This approach leverages spatial invariance principles and self-supervised denoising to achieve high-speed volumetric imaging of dynamic specimens, including neurons in a live zebrafish brain, all while mitigating phototoxicity by reducing laser power and axial sampling [**?**].

The concept of computational autofocusing has also been explored from a single out-of-focus capture. Luo et al. introduced a deep learning pipeline that, once trained, processes an arbitrarily defocused image to synthesize an in-focus image offline. By doing so, they circumvent iterative axial scanning altogether. Their findings showed substantial gains in acquisition speed and reduced photobleaching, which are particularly beneficial for prolonged or large-area time-lapse studies [58]. Beyond autofocusing, Cho et al. detailed a method called deep decomposition and deconvolution microscopy (3DM), aimed at fast volumetric imaging of neuronal activity using a conventional epi-fluorescence system. Their pipeline first decomposes wide-field images into low-rank and sparse components, then employs a second network for deconvolution of the sparse activity signals. This combination effectively removes background fluorescence, preserving the crucial temporal dynamics of processes such as whole-brain zebrafish imaging.

Another line of research focuses on super-resolving or upgrading conventional fluorescence images through trained networks. Rivenson et al. illustrated that a deep neural network can learn to transform low-numerical-aperture (NA) images into high-NA equivalents, effectively enhancing lateral resolution and contrast over wide fields of view. This data-driven mapping bypasses complex physical models and instead relies on carefully aligned training pairs to bridge the gap between different optical configurations without modifying the microscope hardware [59].

Finally, Wu et al. proposed a system called Deep-Z that takes a single two-dimensional fluorescence image and attaches a user-defined digital propagation matrix to infer refocused slices in three dimensions. Their work effectively extended the depth of field by up to twentyfold for live neuronal samples, such as *C. elegans*, with minimal loss of resolution. Equally

important, the same architecture can correct for sample tilt or drift in post-acquisition steps, making it attractive for experiments requiring high-throughput volumetric monitoring of living specimens. An additional variant, Deep-Z+, was also shown to translate wide-field images to confocal-like outputs, illustrating how deep networks can cross-connect distinct optical modalities [13].

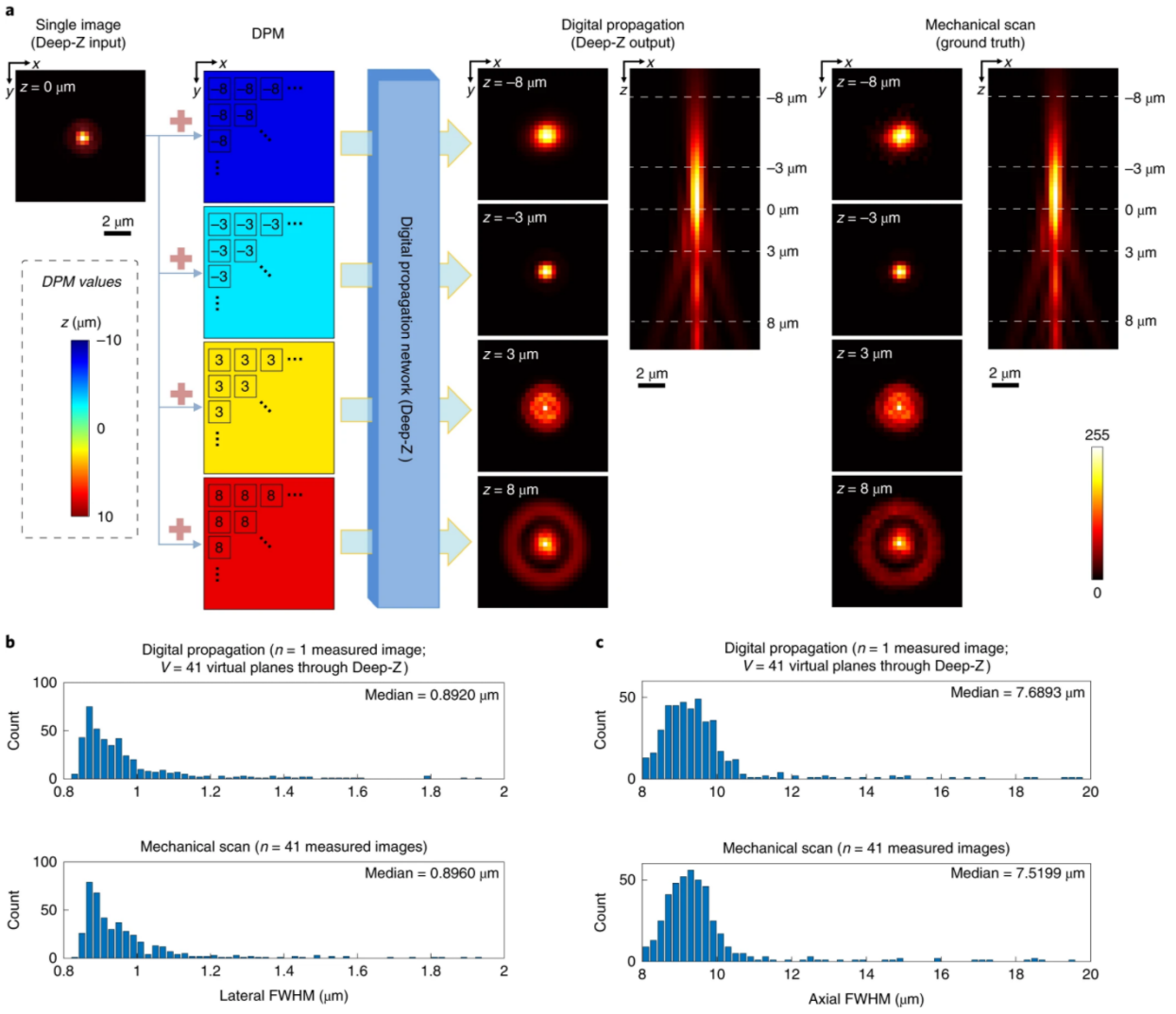

Figure 4.2: *Deep-Z Microscopy.* Single plane images are feed into a trained neural network, where the are refocused up to distance of $\pm 10\mu$m. For conformation, the Deep-Z generated refocused images are compared to the mechanical translations focused images for a fluorescent bead (a). (b) is distribution of the FWHM of 300 nm fluorescent beads in the lateral dimension for both Deep-Z propagation and mechanical objective translation. Similarly for (c), the axial dimension FWHM are measure. In both cases, it is seen that the Deep-Z propagation matches very well with the mechanical translation.

More recent efforts incorporate recurrent architectures for extended depth-of-field reconstructions. Huang et al. introduced Recurrent-MZ, which fuses multiple sparsely sampled

axial slices within a single recurrent network pass, achieving multi-plane volumetric outputs while cutting the number of physical scans by up to thirtyfold. Their framework also supports cross-modality learning, transforming wide-field images into confocal-like volumes, emphasizing the flexibility of time-sequence models for fluorescence microscopy [14].

Taken as a whole, these methods reveal how carefully trained convolutional and recurrent neural networks—often combined with domain knowledge about fluorescence imaging physics—can elevate performance, speed, and resolution in three-dimensional microscopy. They show that by minimizing brute-force axial scanning and leveraging computational reconstruction, modern fluorescence microscopy can reach unprecedented volumetric acquisition rates, even on standard hardware. Such data-driven pipelines alleviate classic trade-offs of resolution, field of view, and phototoxicity, and they highlight the continuing evolution of optical microscopy through machine learning.

CHAPTER 5

Research Objective - Interlude

The key metric in microscopy spatiotemporally relevant bandwidth. What I mean by this is the complete utilization of all microscopy parameters to maximize the bandwidth of a system. In Chapter 2, we explored how spatial sampling can have an effect on the imaging quality of data. Oversampling leads to wasted information and degradation in your bandwidth quality. If I were to image a single PSF and magnify it to occupy my entire sensor, how useful would that be? Not very, I would argue. However, on the opposite side, if I were to severely undersample my image then I might be losing information that would otherwise be be useful. When imaging we must balance these pieces of information to create the most useful tool possible. Therefore, in this work, I will justify my creation as having maximized all of the aforementioned parameters given that the available hardware at my disposal to create the highest performing system. That is, a system that does not sacrifice (much) field-of-view for speed or resolution for spatial sampling, or any combination of these. I am primarily building this tool to image whole neuron level activity and given that neurons are typically ≥ 10 μm in diameter, I can justify some spatial sacrifices that I will address later. This is achieved through a careful balance of hardware (optical) and software (neural networks).

5.1 Volumetric Rates and Bandwidth

Below in Table 5.1, a comparison of the latest microscope technology is presented to gain a broader understanding of the speeds that are currently available to conduct scientific inquiries. It is evident that the domains of time, volume, and resolution, are interrelated by a common principle, and that is *bandwidth*, grey column. Furthermore, although bandwidth

is a great indicator of the imaging speed of a system, the lack of *sensor-to-sensor* comparison renders the question more complex. Given that some systems do utilize sensors that have an intrinsically faster bandwidth, the overall bandwidth is bound to be larger. For example, when comparing a Hamamatsu Flash 4.0 v2 to a HiCam Fluo Image Sensor CMOS camera, the repetition rate of the digitizers are completely different. Succinctly, the Flash 4.0 v2 has a maximum frame rate of 50 FPS (under light-sheet mode) for a pixel area of 2048×2048 pixels^2. Conversely, the HiCam Fluo has a frame rate of 1000 FPS at 1280×1024 pixels^2. This results in a maximum bandwidth of 0.21 GPixels/s (Flash 4.0 v2) versus 1.31 GPixels/s (HiCam Fluo). Nevertheless, if paired with identical sensors, the system that utilizes the most overall sensors is the system with the theoretical maximum bandwidth. Having been said, total volume imaged per unit time is another valid form of comparison, given that much of the research is put forth in order to cover more volume, faster.

5.2 Transverse-Sheet Illumination Microscopy

TranSIM is a new microscopy method is presented that utilizes previous work, namely LSCM illumination and detection methodologies, and also novel methods to attack the problem of fast "3D" microscopy. By parallel processing the scanning method of LSCM, much like Spinning Disk Confocal parallel processed the illumination and detection of its Point Confocal predecessor, TranSIM is able to speed up acquisition by several orders of magnitude. In addition, the depth refocusing methodology, removes the slow mechanical translation seen in some previous systems, which tended to be a bottleneck when it came to volumetric imaging. To do achieve such speeds, TranSIM takes the light sheet produced by cylindrical lens and multiplexes it such that is has a lateral and axial separation, therefore, the light that is no longer convolved in its propagation. This allows for a purely optical method by which to detect the separate image planes being illumination.

Method		Camera			Scanned Rate		Volume Size			Voxel Size			Num of Voxels			Geometry	Bandwidth	Volume Rate
Name	Beam	Camera	Frame	#	Volume	Frame	x	y	z	x	y	z	x	y	z			
			FPS		VPS	FPS	µm	µm	µm	µm	µm	µm					Gpixels/s	mm^3/s
TranSIM	Gaussian	Flash 4.0	100	3	200	200	554	748	160	0.8	1.6	20	682	460	9	Epi	0.565	13.2605
	Gaussian	Fusion	100	3	400	400	624	748	160	0.8	1.6	20	768	460	9	Epi	1.272	29.8721
	Gaussian	Lightning	121	3	681	681	1229	748	160	0.8	1.6	20	1536	460	9	Epi	4.331	100.1661
TranSIM +Deep-Z	Gaussian	Flash 4.0	100	3	200	200	554	748	185	0.8	1.6	0.5	682	460	369	Epi	23.153	15.2911
	Gaussian	Fusion	100	3	400	400	624	748	185	0.8	1.6	0.5	768	460	369	Epi	52.144	34.4463
	Gaussian	Lightning	121	3	681	681	1229	748	185	0.8	1.6	0.5	1536	460	369	Epi	177.551	115.5040
IsoView	Gaussian	Flash 4.0	100	1	2	100	800	800	800	0.4	0.4	4	2000	2000	200	Ortho	1.600	1.0240
Lattice	Lattice	Flash 4.0	100	1	10	1000	30	30	20	0.15	0.15	0.2	200	200	100	Ortho	0.040	0.0002
Bessel	Bessel	Flash 4.0	100	1	5	200	60	100	40	0.15	0.15	1	400	1000	40	Ortho	0.080	0.0012
DualView	Gaussian	Flash 4.0	100	1	0.5	50	50	50	50	0.6	0.6	0.6	260	360	50	Ortho	0.002	0.0001
SCAPE	Gaussian	Zyla	100	1	10	100	260	800	265	3.3	3.3	2.6	100	240	80	Epi	0.019	0.5512
SCAPE 2.0	Guassian	HiCAM Fluo	1000	1	321	18308	197	293	78	3.9	1.07	0.86	100	640	57	Epi	1.171	1.4452
Light Field	Widefield	Zyla	100	1	20	20	700	700	200	1.4	1.4	2.6	500	500	76	Epi	0.380	1.9600
Spinning Confocal	Widefield	Flash 4.0	100	1	6	200	150	150	50	0.6	0.6	1.6	250	250	33	Epi	0.012	0.0068

Table 5.1: *Microscope Comparison.* Entire parameter space of various microscopes are listed in the area of volumetric scanning. Major research developments in microscopy have taken place in order to increase the speed are which they can image volumes. Although their suitable application domains vary wildly from system to system, the overarching connection is *bandwidth*. In the *grey* column, the total bandwidth of the system is shown in units of GigaPixels/s. Moreover, the *green* column by-passes any sensor non-equivalences, and instead focuses on the total volumetric imaging field imaging rate that is achieved in mm^3/s, or volume size times the volume rate, of each system. Also shown, are proposed versions of *Transverse-Sheet Illumination Microscopy*, in addition to its first proof-of-concept. These future additions aim to explore the possibility of increased throughput of the system, limited only by the bandwidth of the current imaging sensors. Courtesy of Dr. Arisaka, adopted microscope comparison in NIH R21 Proposal to include experimental results of TranSIM[5].

5.3 Applied Neural Networks

In Chapter 4, we saw how CNNs and in particular U-nets could be applied to fluorescence microscopy in order to increase resolution, depth-of-field, and optimize analysis, to name a few of its utility cases. Here I will continue this work with the advantage of having TranSIM, a parallel depth augmented imaging system. I will explore how I can transform a sparsely depth separated imaging device into a true three-dimensional microscope. Moreover, I am able to achieve massive field-of-views without sacrificing much in the name of spatial sampling or resolution. With that being said, in Figure 5.1, I showcase how TranSIM coupled with neural networks can transform the way we do microscopy and I hope that this opens the door to a new type of micorscopy.

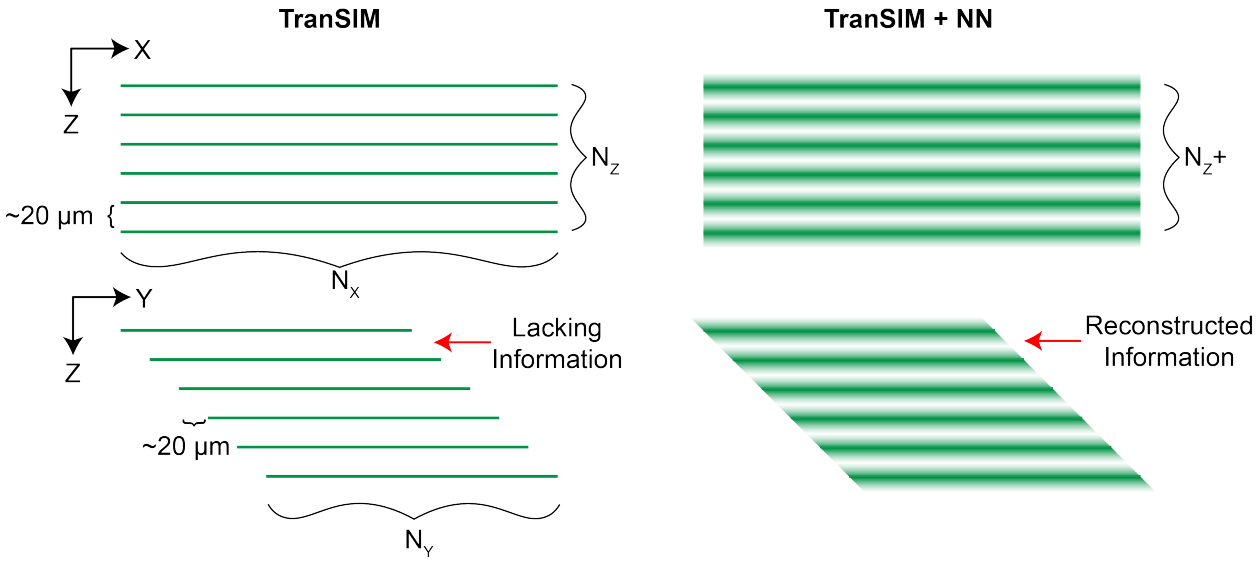

Figure 5.1: *TranSIM + Neural Networks, From Sparse to Dense.* TranSIM is only able to capture sparsely separated image planes. These planes are typically separated by a distance of 20 μm laterally and axially leading to parallelipiped volume. By training neural networks, this data can be virtually refocused and recovered to form a true three-dimensional image.

CHAPTER 6

Materials and Methods

In this chapter, I will provide an in-depth explanation of the methodologies I developed and incorporated into my work. The first part of this chapter is organized into four major sections, each detailing a distinct aspect of the methods employed throughout this research:

- The **first** section focuses on the advancements and modifications made to the TranSIM system since my Master's thesis. These updates include both hardware and optical design enhancements aimed at improving imaging performance and functionality.
- The **second** section outlines the post-acquisition data processing pipeline developed for TranSIM, specifically addressing inter-planar alignment. This pipeline resolves a critical gap in my thesis work, ensuring that the volumetric imaging data is properly aligned and spatially accurate.
- The **third** section delves into the neural network implementations, covering input data preprocessing and the architectural design of the networks (time-independent and time-dependent).
- The **fourth** amd final section of part 1 reviews the custom loss functions used for training, and the analytical approaches applied to interpret the results.

The second part of this chapter dives into to the biological data analysis that was performed to validate this system on real life data. Here I primarily worked with *Drosophila melanogaster* larvae and adult stages as the validation animal. Here I incorporated and further developed, with the aid of some undergraduate students (listed in the acknowledgments

section), and immobilization technique to minimize movement in larvae for neural network training.

This chapter not only highlights the technical innovations made during this research but also demonstrates the integration of computational techniques into the broader experimental workflow, bridging the gap between raw imaging data and biological insights.

6.1 Hardware Optical Design

Transverse Sheet Illumination Microscopy (TranSIM) employs a novel optical design that utilizes spatially multiplexed imaging schemes to achieve simultaneous imaging of multiple axially and laterally separated planes. This configuration significantly enhances volumetric imaging speed while maintaining near-diffraction-limited resolution. The system relies on a carefully engineered combination of beam multiplexing optics, precise alignment mechanisms, and a modular detection pathway to achieve parallelized data acquisition.

In Figure 6.1, the complete optical schematic of TranSIM is illustrated, highlighting the key components responsible for generating the illumination and detection profiles. The illumination system creates laterally and axially separated light sheets using a combination of beam-splitting optics and reflective cavities, while the detection system ensures that the fluorescent signals from multiple imaging planes are accurately separated and captured onto multiple sCMOS cameras.

For this proof-of-concept implementation, a 2-camera detection system was selected to maximize imaging coverage. Each camera captures three imaging planes, with each plane mapped onto a region spanning 682 pixels $\times$ 920 (or 460) pixels on the sensor. This configuration optimally utilizes the sensor area while maintaining high spatial resolution. Additionally, a "light-sheet" mode is employed, enabling the rejection of out-of-focus light and ensuring confocal-like imaging performance.

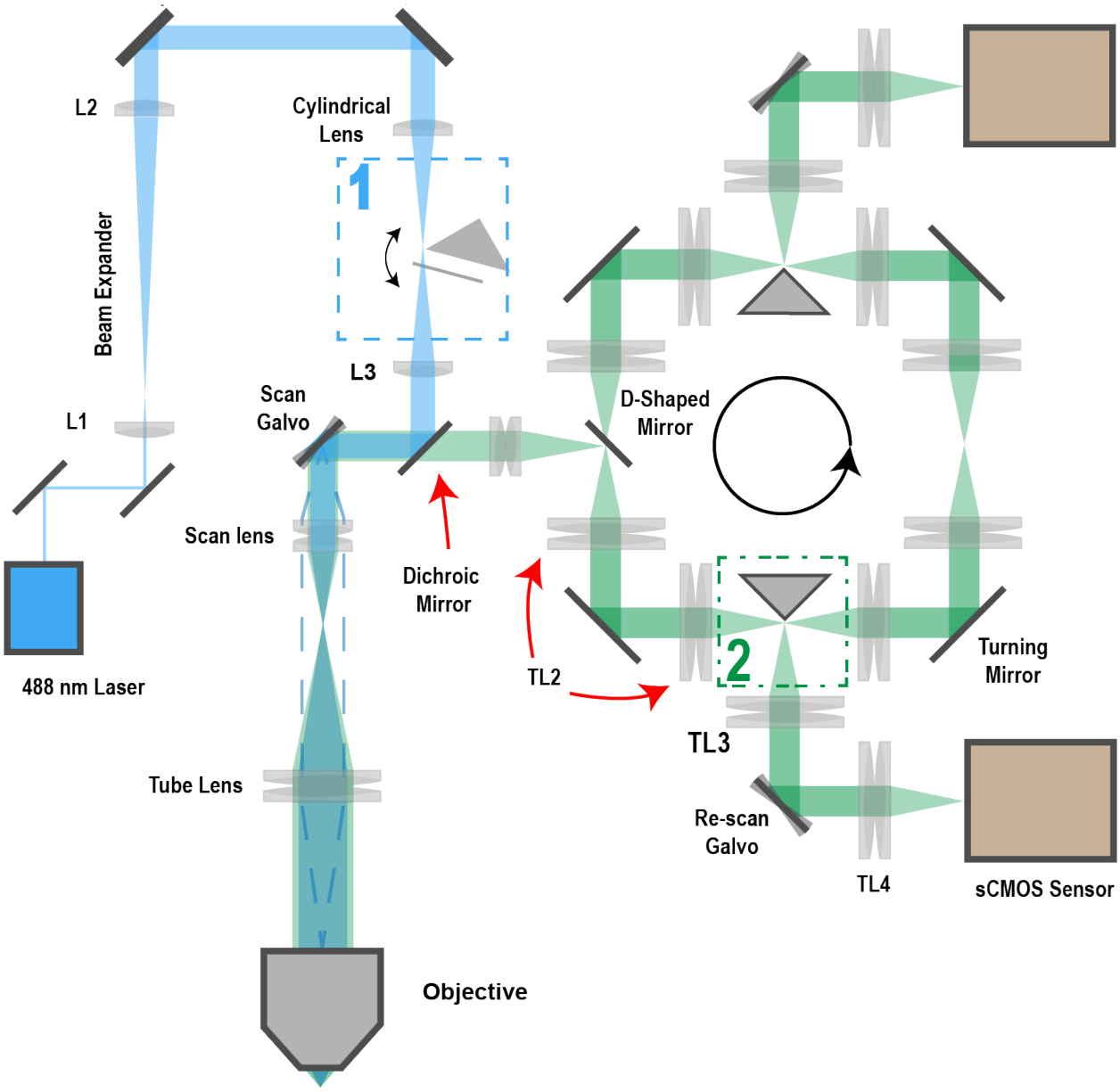

Figure 6.1: *Optical schematic of Transverse Sheet Illumination Microscopy.* The system incorporates a multiplexed illumination scheme that generates laterally and axially separated light sheets, which are then scanned across the imaging field. Fluorescent signals from multiple planes are separated using knife-edge mirrors and mapped onto sCMOS sensors, enabling simultaneous multi-plane imaging.

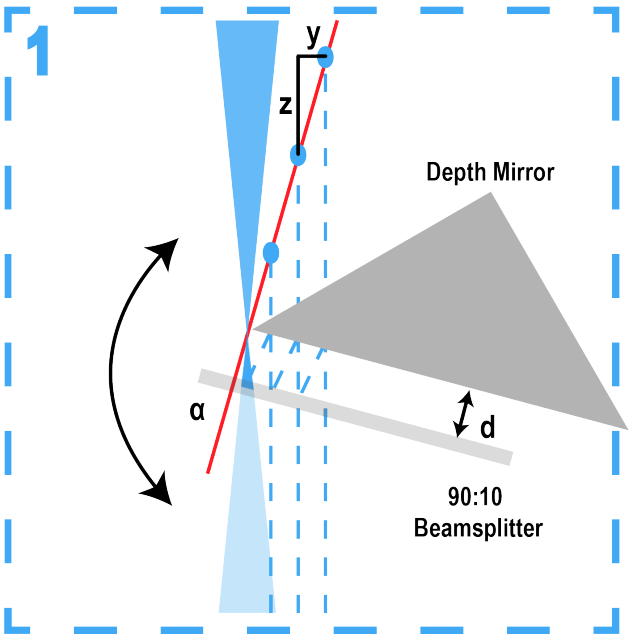

Figure 6.2: *Optical Schematic of Illumination.* By placing parallel facing partially reflecting and fully reflecting mirrors, a focused laser beam can be multiplexed *ad infinitum* with laterally and axially separated foci.

6.1.1 Illumination Design

TranSIM is a 1P system using a 488 nm, 100 mW Coherent Sapphire (Coherent Inc.) continuous wavelength laser. The laser beam diameter is expanded using a -50 mm (LC1715-A, Thorlabs) to 250 mm (LA1461-A, Thorlabs) telescope for a 5 times beam expansion from a 1 mm to 5 mm diameter. The beam is then passed through a 100 mm cylindrical lens (LJ1567RM, Thorlabs) to create the prime laser line that will be multiplexed. For multiplexing, a 90:10 R:T beamsplitter (BS) (BSX10R, Thorlabs) is placed immediately after the focal line of the cylindrical lens with angle α from the optical axis. A wedge mirror is then brought it a close as possible to the laser line without clipping it and placed parallel to the BS. The BS allows 10% of the light through and 90% is reflected to the wedge mirror which in turn fully reflects the beam back towards to the BS. Here the beam again passes through the BS with 10% of the remaining power and the process repeats. With each subsequent

bounce, the transmitted beam's power decreases exponentially

$$P_n = 0.1 \times (0.9)^{n-1} \times P_0.$$

Nevertheless, this results in an infinite set of beams that are equally laterally and axially separated. The lateral separation is introduced by the slight angle α, and the axial separation is introduced by the spatial delay introduced by the bouncing to the next lens. Additionally, the lateral and axial separations are governed by the following geometric equations,

$$y = 2d\sin(\alpha),$$

$$z = 2d\cos(\alpha).$$

The multiplexed beams are then collected with a 100 mm achromatic lens (AC254-100-A, Thorlabs) off centered such that the Nth/2 plane (of used planes on the detection side) is placed at the centered axis. A detailed schematic of this beam multiplexing system is shown in Figure 6.2. The beams are then reflected off a Dichroic mirror (T495lpxru, Chroma) towards a galvanometer for scanning. The scan lens is an effective 100 mm focal length telecentric lens created using a pair of 200 mm achromatic lenses (ACT508-200-A, Thorlabs). Similarly, the tube lens is a 200 mm effective focal length lens comprised of a pair of 400 mm achromatic lenses (ACT508-400-A, Thorlabs). Telecentric lenses were chosen to maintain constant magnification. For our imaging lens we chose a Nikon 16X 0.8 NA objective. The laser lines are then scanned laterally using the galvanometer.

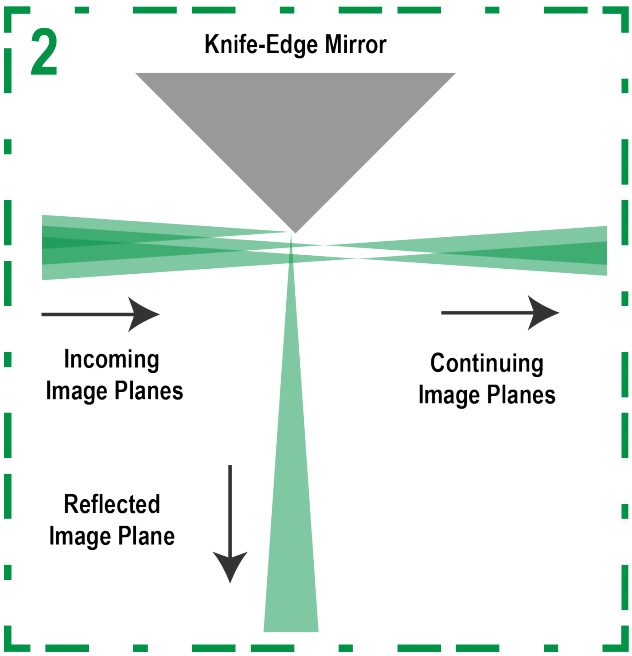

Figure 6.3: *Optical Schematic of Detection.* The once laterally and axially separated foci can be *trimmed* away and reflected onto sCMOS sensors for imaging. Here I show a detailed view of how these foci can be steered. On the round trip, the Nth + 2, image plane will be at the same position axially as the first plane, but laterally shifted to be adjacent to the previous plane leading to a set of axially refocused and laterally displaced image planes that get sent in unison to the sensor.

6.1.2 Detection Design

The detection beams are collected with the same Nikon 16X 0.8 NA objective and are initially scanned in the same manner as the illumination beams until they are de-scanned by the galvanometer. Once de-scanned, the beams are then passed through the dichroic mirror where they are injected into the cyclic module. The initial insertion is made by focusing the focal lines using a 100 mm effective focal length telecentric lens using a pair of 200 mm achromatic lenses (AC254-200-A, Thorlabs) onto a D-shaped pickoff mirror (BBD05-E02, Thorlabs). The beams are initially laterally offset in the cyclic module as this will allow space for realignment and focusing. Now inside of the cyclic module, the beams are

relayed using a 4f relays composed of a 100 mm effective focal length telecentric lenses made from pairs of 200 mm lenses (ACT508-200-A, Thorlabs) and a mirror for 90-degree reflection placed between the telecentric lenses. At the relayed focal plane, a knife-edge mirror (MRAK25-P01, Thorlabs) is brought in close enough to the beams to pick-off the first image plane. This image plane is then relayed and rescanned into an ORCA-Flash4.0 V2 sCMOS sensor (Hamamatsu) with the rolling shutter enabled. The rescanning is produced by a galvanometer. The remaining beams are relayed with a pair of identical 4f telecentric relays to the second knife-edge pick-off mirror where the second beam is sent to the sCMOS sensor. Using the last 4f relay, the remaining beams are laterally adjusted by physically translating the 4f relay to be offset from the previous three relays. The focusing is achieved by adjusting the position of the 4f relay lens distances. The beams are then allowed to pass underneath the D-shaped pick-off mirror with the beams having been realigned and refocused to be adjacent to the initial pass and the cycle repeats.

6.1.3 Inter-Planar Distance and Volumetric Field-of-View

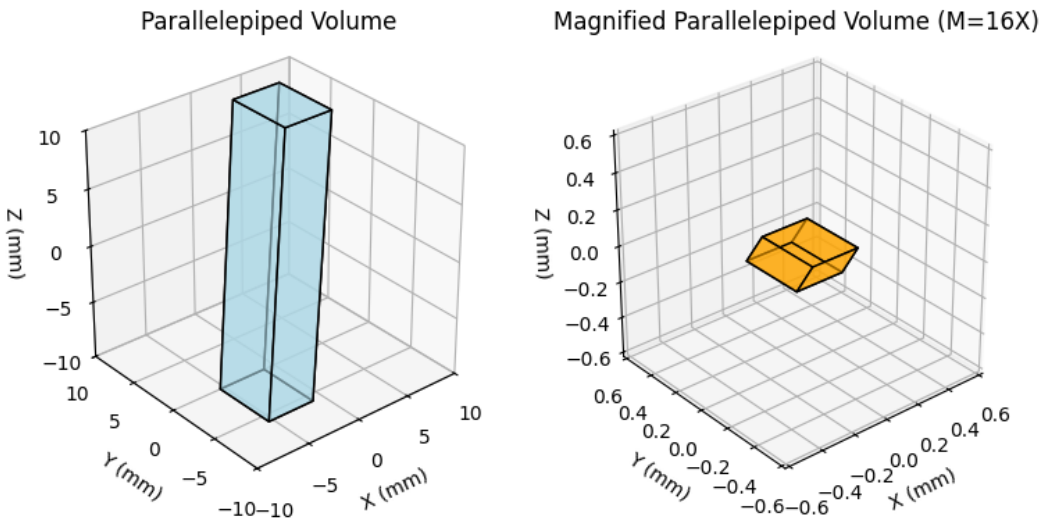

Figure 6.4: *Effect of Magnification on Volume Scaling.* The original parallelepiped volume (left) and the magnified version (right) demonstrate anisotropic scaling. The lateral dimensions (X and Y) shrink by a factor of $1/M$, while the axial dimension (Z) compresses more significantly by a factor of $1/M^2$. This mimics how magnification in microscopy increases lateral resolution while reducing axial depth, leading to a flattened appearance in the Z-direction.

TranSIM produces volumetric data with a staircase profile, resulting in a parallelepiped-like volume. The lateral and axial offsets are determined by the angle of incidence, α, on the 90:10 beamsplitter and the separation distance to the depth mirror, d. The pre-magnification lateral plane separation, y, is given by the trigonometric relation:

$$y = 2d\sin(\alpha).$$

Similarly, the pre-magnification axial displacement, z, is determined by:

$$z = 2d\cos(\alpha).$$

After applying the effective system magnification, M_{eff}, the lateral and axial distances scale inversely linearly and quadratically, respectively. The modified, magnified lateral and axial plane separations are:

Lateral (y) plane separation:

$$y' = \frac{2d\sin(\alpha)}{M_{\text{eff}}}.$$

Axial (z) plane separation:

$$z' = \frac{2d\cos(\alpha)}{M_{\text{eff}}^2}.$$

To determine the shearing of the parallelepiped volume, the magnified angle α' is given by:

$$\tan(\alpha') = \frac{y'}{z'} \quad \Rightarrow \quad \alpha' = \tan^{-1}\left(M_{\text{eff}}\tan(\alpha)\right).$$

The volumetric field-of-view (FOV) is calculated using the volume equation for a parallelepiped:

$$V_{\text{FOV}} = Bh = XYZ,$$

where:

$$B = XY, \quad h = Z = z'N_z.$$

$$\tan(\alpha') = \frac{y'}{z'} \quad \Rightarrow \quad \alpha' = \tan^{-1}\left(M_{\text{eff}} \tan(\alpha)\right).$$

Here, B represents the base area of an image plane, also defined as XY, while h represents the total volume height, also defined as Z, where the total height is determined by the number of planes N_z multiplied by the magnified axial plane separation z'.

Due to the inverse quadratic dependence on magnification for the magnified angle α', the volumetric FOV must remain within a usable range by restricting the angle of incidence to:

$$\alpha \leq 3.58°.$$

This constraint prevents excessive shearing that could compromise data integrity and ensures that the volumetric field of view is well-suited for high-resolution imaging applications.

6.1.4 Systems Control

To synchronize the illumination and detection components of TranSIM, we developed a custom Python software suite. The suite's first module generates waveforms to synchronize the electro-optical-mechanical devices at the hardware level. The system consists of two sCMOS sensors, one scanning galvanometer, two re-scanning galvanometers (one per camera), and a 10 nm precision piezo actuator (P-721, Physik Instrumente). The data acquisition (DAQ) workstation controls these devices using the analog output of an NI PCIe 6363 (Texas Instruments) DAQ card operating with a sampling rate (f_s) of 200 kSamples/second. The high sampling rate is chosen to avoid waveform aliasing. The sCMOS sensors are externally triggered using a 50% duty cycle square waveform with frequency (f_w) depending on the region of interest (ROI) used. For example, we commonly operate the system with frequencies of 10, 50, 100, and 200 Hz. With the same frequency as the camera, the scanning galvanometer is supplied with a low pass filtered triangle waveform. The waveform is synthesized based on specified parameters (frequency, amplitude, phase, offset, duty cycle) and smoothed using a Gaussian filter and a 98% duty cycle. This approach effectively attenuates high-frequency noise, yielding a controlled output signal with defined frequency characteris-

tics. For the specified waveforms, we utilized a cut-off frequency (f_c) approximately double the scanning frequency, $f_c \approx 2f_w$. The rescanning galvanometers undergo a similar waveform parametrization procedure. However, due to variability in the manufacturing of these devices, the line-scanning speed produced by the galvanometers is synchronized with the respective camera by adjusting the internal line interval speed within the camera parameters. Communication with the laser is established via a serial connection, allowing precise control over laser operations directly through the application.

6.1.5 Data Acquisition

The second module in the Python suite controls the data acquisition of the system. The ORCA-Flash4.0 V2 sCMOS sensors are connected to the DAQ via Camera Link PCIe cards (FBD-1XCLD-2PE8, Active Silicon). The data is then processed by the DAQ using custom Python software that allows for real-time visualization of the data via multi-threaded processing and OpenCV display. The software also allows simultaneous display and recording of the data and is fully synchronized with the systems control software. Data is transferred to and stored directly to NVMe drives in a raw and uncompressed h5 file format.

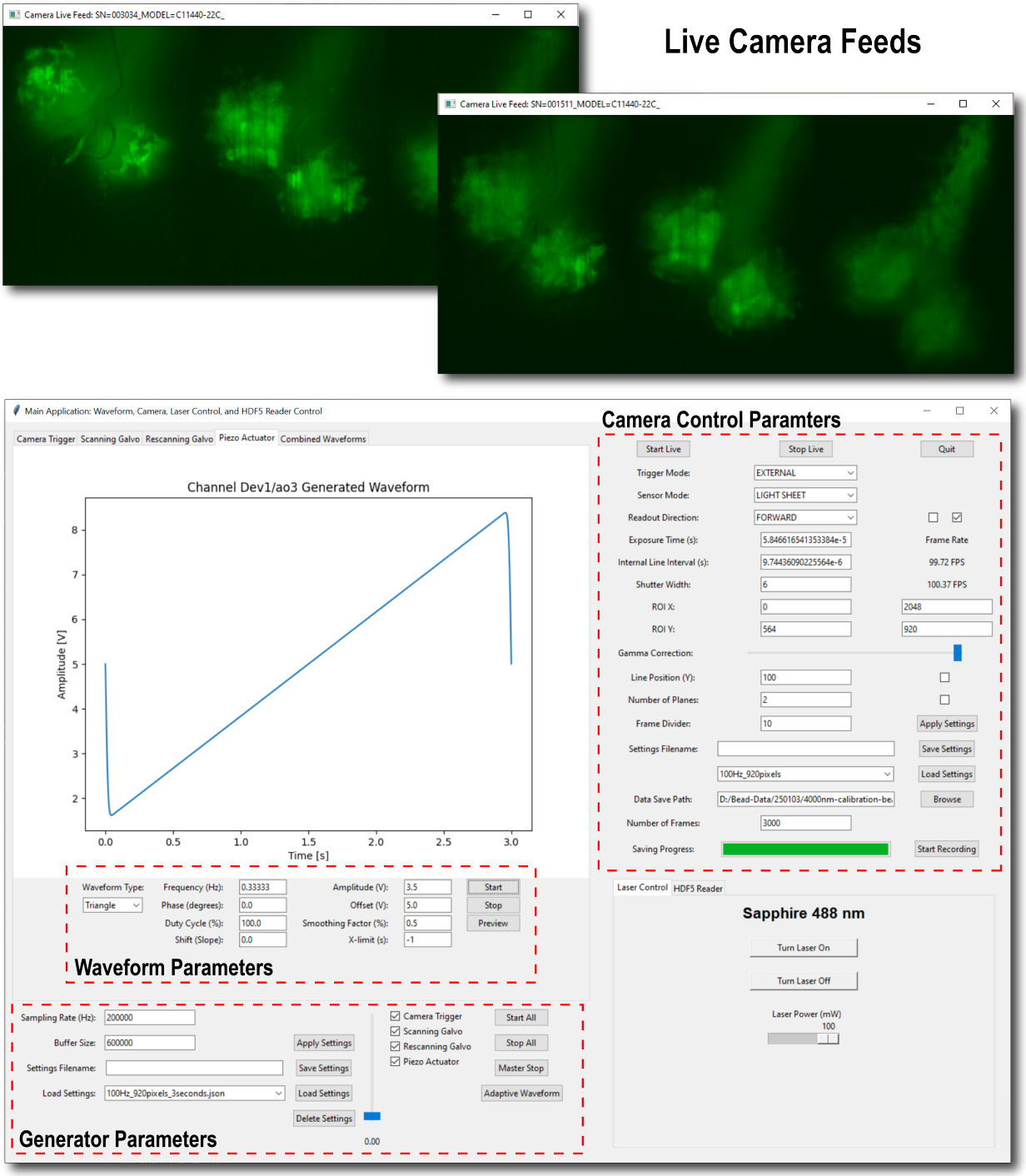

Figure 6.5: *Control Software.* Fully integrated control software for waveform generation, laser control, and camera control. The software is generalized enough to accept any number of cameras, as each camera is controlled by its own thread. Each module is also control by its own thread.

6.2 Data Processing Pipeline for Inter-Planar Alignment

One of the major challenges associated with the TranSIM system is the inherent staggered arrangement of imaging planes, which creates a staircase-like pattern in the raw data. This spatial misalignment poses significant difficulties in reconstructing coherent three-dimensional volumes from the collected data. Without proper correction, this misalignment can lead to artifacts, distortions, and inaccuracies in volumetric analyses, undermining the reliability and utility of the imaging system. To address this issue, I developed a robust and systematic data processing pipeline specifically designed for the precise alignment of imaging planes in three-dimensional space.

The core of the pipeline is the alignment algorithm, which operates in multiple stages to achieve precise inter-planar registration. Initially, a global alignment step is performed to address large-scale translational and rotational misalignments between adjacent planes. This is achieved using phase-correlation techniques, which identify the relative shifts required to coarsely align the planes. Once a baseline alignment is established, a local refinement step further adjusts the planes by optimizing their positions iteratively. This stage uses advanced similarity metrics, such as mutual information, to fine-tune the alignment and account for smaller-scale discrepancies.

Validation of the alignment process is a critical component of the pipeline. Synthetic datasets with known distortions are used to benchmark the accuracy of the algorithm, providing a controlled environment for testing. Additionally, experimental data from fluorescent bead samples are analyzed to confirm the real-world applicability of the pipeline. Metrics such as normalized cross-correlation and alignment error are used to quantify the effectiveness of the alignment, ensuring that the final dataset is both accurate and reliable.

The developed pipeline represents a significant advancement over the initial state of the TranSIM system, addressing a critical limitation in its ability to produce volumetric data. By enabling the precise alignment of imaging planes, this methodology not only enhances the quality of three-dimensional reconstructions but also extends the utility of the TranSIM system for a broader range of biological applications. The aligned datasets provide a robust

foundation for subsequent analyses, including those in the following section regarding neural networks, paving the way for deeper insights into dynamic biological processes.

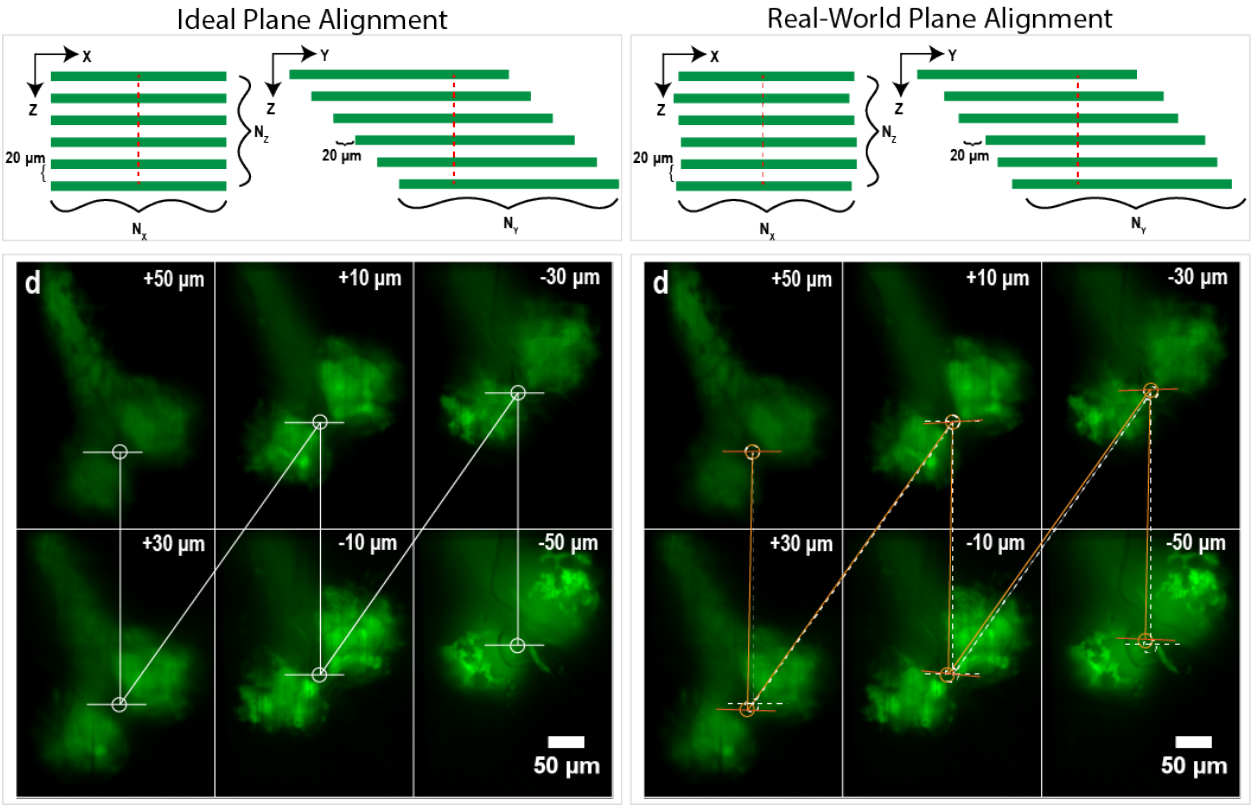

Figure 6.6: *Idealized vs. Real-World Inter-Planar Alignment.* On the left, the idealized situation assumes perfect knowledge of the inter-planar distance, allowing for straightforward alignment of imaging planes based solely on their spatial separation. This approach results in a coherent three-dimensional volume without the need for additional corrections. On the right, the real-world scenario is depicted, where each imaging plane may undergo affine transformations due to experimental factors resulting from slight optical misalignment. In white are the idealized locations and in orange are the true transformations that must be reversed for proper alignment. These transformations include translations, rotations, and scaling, which necessitate a robust computational pipeline to iteratively refine and accurately align the planes in three-dimensional space.

6.2.1 Mathematical Framework for Inter-Planar Alignment

The basic premise of this framework is to align axially and laterally separated planes by conducting a peak cross correlation search and creating a cumulative transformations matrix set that can be used at a later time with subsequent imaging data.

Fourier Domain Phase Cross-Correlation: To align two adjacent planes, $I_{\text{ref}}(x, y)$ (reference image) and $I_{\text{reg}}(x, y)$ (image to register), we first compute their translational shift using Fourier domain phase cross-correlation.

1. *Fourier Transform of the Images:*

$$\mathcal{F}_{\text{ref}}(u, v) = \mathcal{F}\{I_{\text{ref}}(x, y)\}, \quad \mathcal{F}_{\text{reg}}(u, v) = \mathcal{F}\{I_{\text{reg}}(x, y)\},$$

where (u, v) are spatial frequency coordinates.

2. *Cross-Power Spectrum:* The cross-power spectrum isolates the phase difference:

$$C(u, v) = \frac{\mathcal{F}_{\text{ref}}(u, v) \cdot \mathcal{F}^*_{\text{reg}}(u, v)}{|\mathcal{F}_{\text{ref}}(u, v) \cdot \mathcal{F}^*_{\text{reg}}(u, v)|},$$

where $(\cdot)^*$ denotes the complex conjugate.

3. *Inverse Fourier Transform for Correlation:* The inverse Fourier transform of $C(u, v)$ gives the phase correlation matrix:

$$R(x, y) = \mathcal{F}^{-1}\{C(u, v)\}.$$

The peak location in $R(x, y)$ indicates the translational shift $(\Delta x, \Delta y)$ between the two images:

$$(\Delta x, \Delta y) = \text{argmax}_{(x,y)} R(x, y).$$

Fine Affine Transformation: To refine the alignment, we apply an affine transformation to the registered image. The affine transformation matrix T is parameterized as:

$$T = \begin{bmatrix} a & b & t_x \\ c & d & t_y \\ 0 & 0 & 1 \end{bmatrix},$$

where: - a, b, c, d represent scaling, rotation, and shear, - t_x, t_y represent translation[**?**].

To optimize T, we minimize a similarity metric, such as mutual information or normalized cross-correlation:

$$\text{argmin}_T \sum_{x,y} |I_{\text{ref}}(x,y) - I'_{\text{reg}}(x,y)|,$$

where $I'_{\text{reg}}(x,y)$ is the transformed registered image:

$$I'_{\text{reg}}(x,y) = I_{\text{reg}} \left(T^{-1} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \right).$$

Cross-Correlation and Optimal Offset Search: For robustness, we calculate the alignment for a range of buffer offsets N in the z-direction. Let Δz be the expected plane separation, and $N \in [-b, b]$, where b is the buffer size. For each offset, we compute:

1. *Cross-Correlation Matrix:*

$$R_N(x,y) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}_{\text{ref}}(u,v) \cdot \mathcal{F}^*_{\text{reg},N}(u,v)}{|\mathcal{F}_{\text{ref}}(u,v) \cdot \mathcal{F}^*_{\text{reg},N}(u,v)|} \right\}.$$

2. *Peak Correlation Value:*

$$\text{peak}_N = \max_{(x,y)} R_N(x,y).$$

3. *Optimal Offset:*

$$N^* = \text{argmax}_{N \in [-b,b]} \text{peak}_N.$$

Iterative Plane Pair Alignment: To ensure proper alignment across all planes: 1. Align each plane k to the preceding plane $k-1$ using the process above, resulting in transformation matrices T_k. 2. Compute cumulative transformations to maintain global consistency:

$$T_{\text{cumulative},k} = T_{\text{cumulative},k-1} \cdot T_k.$$

Output Aligned Data: Once all planes are aligned:

1. Apply $T_{\text{cumulative},k}$ to the k-th plane:

$$I_{\text{aligned},k}(x,y) = I_{\text{raw},k} \left(T^{-1}_{\text{cumulative},k} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \right).$$

2. Save aligned planes into a single 4D (t, plane, x, y) stack for further analysis.

6.3 Neural Network Architectures

This section details the neural network architectures utilized for reconstructing volumetric data. To faithfully reconstruct the volumetric data, I first explored simply UNET architectures with attention mechanisms, however, I found them to be insufficient in the ability to generalize over time. Meaning, although 3D UNETs were able to reconstruct time-independent volumes, they failed to handle the temporal dimension of the data produced by TranSIM. Succinctly, through separation of variables, we can decompose the intensity of each pixel in the 4D data into spatially and temporally separated units.

$$I(x, y, z, t) = I_{3D}(x, y, z) \cdot T(t) \tag{6.1}$$

However, this approach fails to generalize neatly when reconstructing the data. Therefore, I opted to utilize latent lateral connections between independent volumetric UNETs and with a recurrent architecture. The key objective is to transform sparse depth input data into a densely reconstructed volumetric representation with consistent temporal transitions:

$$\mathbf{X}_{\text{input}} \in R^{T \times D_{\text{sparse}} \times H \times W} \quad \rightarrow \quad \mathbf{X}_{\text{output}} \in R^{T \times D_{\text{dense}} \times H \times W}.$$

where T is the temporal dimension, D_{sparse} and D_{dense} represent the sparse input and dense reconstructed depth dimensions, respectively, while H and W correspond to the spatial dimensions of the volumetric data.

6.3.1 Preprocessing

The network input is a volumetric dataset $\mathbf{I} \in R^{D \times H \times W \times C}$. To standardize the input data, a preprocessing step is applied to each slice of the volume.

The preprocessing involves subtracting the mean of the background data and normalizing based on the mean intensity of the top 1.0% of foreground pixel values, I_{top}. To distinguish the background and foreground pixel values, I applied a triangle threshold. This ensures that the input data has a consistent dynamic range and is robust to intensity variations.

This method was followed closely from the Deep-Z framework[13].

The mean-subtracted intensity I_{ms} is calculated as:

$$I_{\text{ms}} = I - \mu_{bg},$$

where:

$$\mu_{bg} = \frac{1}{n} \sum_{i=1}^{n} I_i,$$

and n is the total number of background pixels, calculated from the triangle threshold.

$$n = pixels[I < I_{threshold}],$$

The normalized intensity I' is computed as:

$$I' = \frac{I_{\text{ms}}}{I_{\text{top}}},$$

where:

$$I_{\text{top}} = \frac{1}{m} \sum_{i=1}^{m} I_{\text{sorted},-i},$$

and I_{sorted} is the sorted pixel intensity vector of the dataset, m is the number of pixels corresponding to the top 1.0

$$m = \lceil 0.01 \times pixels[I > I_{threshold}] \rceil.$$

This preprocessing step ensures that the input data has values primarily in the range $(0, 1)$, improving the stability and performance of the network during training. By focusing on the top 1.0% of pixel values, the normalization is robust to outliers and highlights significant intensity variations in the input data.

6.3.2 Network Architecture - Time Independent

The first neural network I explored as the viable option for TranSIM's generation data type is a 3D U-Net variant designed to process volumetric data and includes several modifications to improve reconstruction accuracy, such as residual connections and self attention. However,

it should be noted that this network architecture assumes temporal independence between the generated volumes. This can lead to issues when reconstructing data in time.

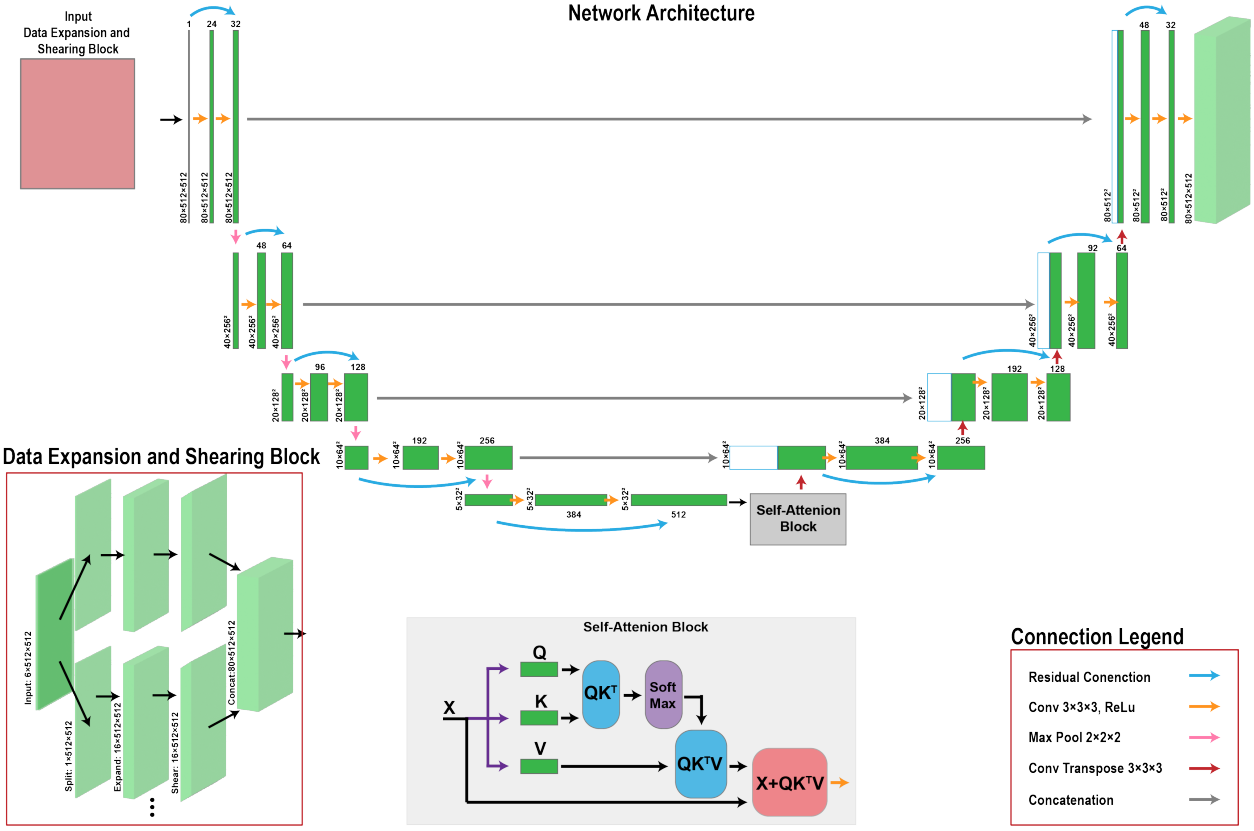

Figure 6.7: *TranSIM Time-Independent Neural Network Architecture.* The figure illustrates the neural network architecture employed for 3D image reconstruction. The network consists of a downsampling path, a bottleneck with self-attention, and an upsampling path. Each stage in the downsampling path applies convolutional layers followed by max-pooling, progressively reducing spatial resolution while capturing deeper features. The bottleneck layer incorporates a self-attention mechanism that operates across the depth dimension, enabling the network to capture global contextual information. In the upsampling path, transposed convolutions restore spatial resolution, and skip connections integrate features from corresponding layers in the downsampling path. The final output is a reconstructed 3D image, with enhanced fidelity and resolution, processed from the input image stack.

6.3.2.1 Input Layer

The input to the network is defined as:

$$\mathbf{X}_{\text{input}} \in R^{D_{sparse} \times H \times W \times C},$$

where D corresponds to the number of sparse input planes, H and W, the height and width of the image plane, and C, the color channels of the data. Here it should be noted that I only implemented a single color channel. The input is reshaped and expanded to include a depth dimension for volumetric processing.

6.3.2.2 Downsampling Path

The downsampling path reduces spatial dimensions while extracting feature representations. For each downsampling block, the operations include two convolutional layers followed by max pooling:

$$\mathbf{X}_{i+1} = \text{MaxPooling3D}\left(\text{ReLU}\left(\text{Conv3D}(\mathbf{X}_i)\right)\right),$$

where $\mathbf{X}_i$ represents the input to the i-th block. Residual connections are included to preserve low-level spatial information:

$$\mathbf{X}_i = \mathbf{X}_i + \text{Conv3D}_{\text{res}}(\mathbf{X}_i).$$

6.3.2.3 Bottleneck with Self-Attention

At the bottleneck layer, self-attention is introduced to enhance the network's capability to capture long-range dependencies. The self-attention mechanism computes:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q} \cdot \mathbf{K}^\top}{\sqrt{d_k}}\right) \cdot \mathbf{V},$$

where: - $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are the query, key, and value matrices obtained via:

$$\mathbf{Q} = \text{Conv3D}_{\text{query}}(\mathbf{X}), \quad \mathbf{K} = \text{Conv3D}_{\text{key}}(\mathbf{X}), \quad \mathbf{V} = \text{Conv3D}_{\text{value}}(\mathbf{X}),$$

- d_k is the dimensionality of the key.

The output of the self-attention [60, 61, 62] block is combined with the input tensor via a residual connection:

$$\mathbf{X}_{\text{out}} = \mathbf{X} + \text{Conv3D}(\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})).$$

6.3.2.4 Upsampling Path

The upsampling path restores spatial dimensions using transposed convolutions. For each upsampling block:

$$\mathbf{X}_{i+1} = \text{Conv3DTranspose}\left(\mathbf{X}_i\right),$$

followed by concatenation with the corresponding feature map from the downsampling path:

$$\mathbf{X}_{\text{concat}} = \text{Concat}([\mathbf{X}_{i+1}, \mathbf{X}_{\text{down}}]).$$

Residual connections are again used to refine the output:

$$\mathbf{X}_{i+1} = \mathbf{X}_{\text{concat}} + \text{Conv3D}_{\text{res}}(\mathbf{X}_{\text{concat}}).$$

6.3.2.5 Output Layer

The output layer consists of a single convolutional layer with a ReLU activation to ensure non-negative predictions:

$$\mathbf{X}_{\text{output}} = \text{ReLU}\left(\text{Conv3D}(\mathbf{X}_{\text{final}})\right).$$

6.3.3 Network Architecture - Time Dependent

To ensure high-fidelity volumetric reconstructions with temporal coherence, I implemented a modified Recurrent Convolutional Neural Network (RCNN) [14, 63, 64, 65]architecture based on a Convolutional Long Short-Term Memory (ConvLSTM3D) model[66, **?**]. Unlike traditional convolutional neural networks (CNNs) that process each frame independently, this architecture incorporates spatiotemporal dependencies across time, mitigating abrupt discontinuities in time-evolving volumetric sequences.

6.3.3.1 Depth Expansion

TranSIM captures a discrete number of imaging planes along the axial (z-direction), requiring an expansion mechanism to obtain continuous volumetric information. To address this, we define the depth expansion factor as:

$$S_z = \lceil \frac{N_z}{N_p} \rceil,$$

where N_p is the number of acquired sparse planes, N_z is the target number of dense planes, $\lceil \cdot \rceil$ denotes the ceiling function to ensure a non-fractional expansion factor. The depth expansion operation utilizes Time-Distributed (TD) 3D transposed convolutions with Rectified Linear Unit (ReLU) activations to interpolate along the z-axis:

$$\mathbf{X}_{\text{expanded}} = \text{TD}\left(\text{ReLU}(\text{Conv3DTranspose}(\mathbf{X}_{\text{input}}, \text{stride} = S_z))\right).$$

Since S_z may introduce an over-sampled volume, we trim the tensor to obtain the desired depth N_z:

$$\mathbf{X}_{\text{cropped}} = \mathbf{X}_{\text{expanded}}[:, z_{\text{start}} : z_{\text{end}}, :, :],$$

where:

$$z_{\text{start}} = \frac{D_{\text{expanded}} - N_z}{2}, \quad z_{\text{end}} = z_{\text{start}} + N_z.$$

This operation ensures that the interpolated depth maintains the appropriate axial resolution while preserving structural integrity.

6.3.3.2 Downsampling Path with Residual Learning

The encoder consists of multiple downsampling stages, each employing Time-Distributed (TD) 3D convolutional layers with residual connections to preserve low-level features and maintain temporal consistency. Each downsampling block follows the structure:

$$\mathbf{c}_k = \mathrm{TD}\left(\mathrm{ReLU}\left(\mathrm{Conv3D}(\mathrm{TD}(\mathrm{ReLU}(\mathrm{Conv3D}(\mathbf{P}_{k-1}))))\right)\right) + \mathrm{TD}(\mathrm{Conv3D}_{\mathrm{res}}(\mathbf{P}_{k-1})),$$

$$\mathbf{p}_k = \mathrm{MaxPool3D}(\mathbf{c}_k),$$

where $\mathbf{p}_{k-1}$ is the output of the previous downsampling stage, $\mathrm{Conv3D}_{\mathrm{res}}$ represents the residual connection. This design enables efficient feature extraction while mitigating information loss due to downsampling.

6.3.3.3 ConvLSTM3D Bottleneck

At the network bottleneck, a Convolutional LSTM (ConvLSTM3D) module captures temporal dependencies in the latent representation. Unlike traditional fully connected LSTMs that flatten spatial features, ConvLSTM3D preserves spatial structure by employing convolutional kernels for both input-to-state and state-to-state transformations.

The gating mechanisms for ConvLSTM3D are defined as:

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

where X_t is the input at time step t, H_t is the hidden state capturing short-term dependencies, C_t is the cell state preserving long-term dependencies, i_t, f_t, o_t are the input, forget, and output gates, W_* are convolutional weight matrices, σ is the sigmoid activation function. This design allows the network to learn dynamic features such as motion trajectories and structural deformations while maintaining temporal consistency.

6.3.3.4 Upsampling Path with Skip Connections

The decoder reconstructs high-resolution volumetric representations by progressively increasing spatial resolution using Time-Distributed (TD) 3D transposed convolutions. Skip connections are incorporated to restore fine-grained features lost during encoding:

$$\mathbf{u}_k = \mathrm{TD}\left(\mathrm{ReLU}(\mathrm{Conv3D}(\mathrm{TD}(\mathrm{ReLU}(\mathrm{Conv3DTranspose}(\mathbf{u}_{k-1}))) \oplus \mathbf{c}_{3-k}))\right),$$

where $\oplus$ represents the concatenation of skip connections. Each upsampling stage refines the reconstructed volume while preserving structural continuity.

6.3.3.5 Final Output

The final volumetric output is generated by applying a $1 \times 1 \times 1$ convolutional layer to map the high-dimensional feature representation to a single-channel volumetric intensity map:

$$\mathbf{X}_{\text{output}} = \mathrm{TD}(\mathrm{Linear}(\mathrm{Conv3D}(\mathbf{u}_3))).$$

6.3.3.6 Temporal Consistency and Motion Preservation

By leveraging ConvLSTM3D within the network bottleneck, the TranSIM neural network maintains smooth transitions across time frames, reducing flickering artifacts and ensuring that dynamic biological structures are accurately reconstructed. This architecture enables the model to capture non-rigid deformations and neural activity with high temporal precision. Meanwhile, it also maintains spatial continuity without introducing sudden discontinuities between frames. Moreover, it enhances feature representations using recurrent memory, avoiding loss of temporal coherence. The combination of a recurrent convolutional network with time-distributed convolutions and ConvLSTM3D allows TranSIM to achieve high-quality volumetric reconstructions with spatiotemporal fidelity. This design ensures that each inferred plane aligns seamlessly with its temporal neighbors, overcoming the limitations of purely spatial 3D CNN architectures. By incorporating this temporal-aware

network structure, TranSIM can reconstruct complex, time-dependent biological phenomena with significantly improved accuracy.

6.4 Loss Functions and Training

Loss functions play a fundamental role in training convolutional neural networks (CNNs) by quantifying the discrepancy between predicted outputs and ground truth data. The choice of an appropriate loss function directly influences the optimization process and the final performance of the network. Two widely used loss functions for image processing tasks are **Mean Absolute Error (MAE)** and **Mean Squared Error (MSE)**, both of which operate at the pixel level.

The Mean Absolute Error (MAE) is defined as:

$$L_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|, \tag{6.2}$$

where N represents the total number of pixels, y_i denotes the ground truth pixel value, and $\hat{y}_i$ is the predicted pixel value. MAE provides a linear penalty for errors, making it robust to outliers and ensuring stable optimization. Its constant gradients contribute to a steady convergence process, making it particularly useful in applications where noise robustness is required.

Conversely, the Mean Squared Error (MSE) is given by:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2. \tag{6.3}$$

Unlike MAE, MSE penalizes large errors more heavily due to its quadratic nature, thereby focusing the optimization process on minimizing significant deviations. This makes MSE suitable for applications requiring precise pixel-level accuracy, such as super-resolution and image reconstruction, though it may be sensitive to extreme outliers.

Despite their effectiveness, both MAE and MSE suffer from limitations in capturing

structural and perceptual differences within images. Since they treat each pixel independently, critical local features such as edges and textures may be underrepresented in the loss computation. To address these shortcomings, perceptually aware loss functions such as the Structural Similarity Index Measure (SSIM) have been introduced. SSIM accounts for luminance, contrast, and structural similarity to better align loss with human visual perception [**?**]. A hybrid loss function incorporating both pixel-wise accuracy and perceptual fidelity is formulated as:

$$\mathcal{L}_{\text{total}} = \lambda \cdot \mathcal{L}_{\text{MAE}} + (1 - \lambda) \cdot \mathcal{L}_{\text{SSIM}}, \tag{6.4}$$

where:

$$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(y, \hat{y}).$$

The SSIM function is defined as:

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}.$$

Here, y and $\hat{y}$ represent the ground truth and predicted outputs, while λ is a weighting parameter that balances the contributions of pixel-wise accuracy and perceptual quality.

6.4.1 Cosine-Weighted MAE Loss

A key challenge in training neural networks for TranSIM lies in the spatial distribution of planes. Since the imaging system produces sparsely sampled planes, the neural network must infer intermediate structures, leading to an overemphasis on planes closer to the input. To address this issue, a cosine-weighted MAE loss function is introduced to prioritize the accurate reconstruction of planes further from the original sparse inputs.

The cosine-weighted MAE loss is formulated as:

$$L_{\text{Cos,MAE}} = \frac{1}{N} \sum_{i=1}^{N} w(i) \cdot |y_i - \hat{y}_i|, \tag{6.5}$$

where the weight function $w(i)$ is defined as:

$$w(i) = 1 + (\beta - 1) \cdot \cos\left(\frac{\pi \cdot (i - \mu)}{\rho}\right).$$

Here, β is a scaling factor that modulates the impact of the weighting, μ is the midpoint of the stack, and ρ represents the periodicity of the plane distribution. This weighting function ensures that errors in regions far from the input planes contribute more significantly to the optimization process, thereby reducing bias in the learned reconstructions.

The corresponding gradient is:

$$\frac{\partial L_{\text{Cos,MAE}}}{\partial \theta} = \frac{1}{N} \sum_{i=1}^{N} w(i) \cdot \text{sign}(y_i - \hat{y}_i) \cdot \frac{\partial \hat{y}_i}{\partial \theta}.$$

Higher-weighted planes have a stronger influence on the optimization, ensuring that the model learns a more uniform distribution of accurate reconstructions across the entire volume.

6.4.2 Error Redistribution and Optimization Impact

The total error E in the system can be expressed as:

$$E = \sum_{i=1}^{N} e_i = \sum_{i=1}^{N} |y_i - \hat{y}_i|.$$

With cosine-weighted loss, the error is reformulated as:

$$E_{\text{weighted}} = \sum_{i=1}^{N} w(i) \cdot e_i.$$

This weighting redistributes the gradient impact, modifying how errors are corrected during training. The optimization update for a parameter θ_k follows:

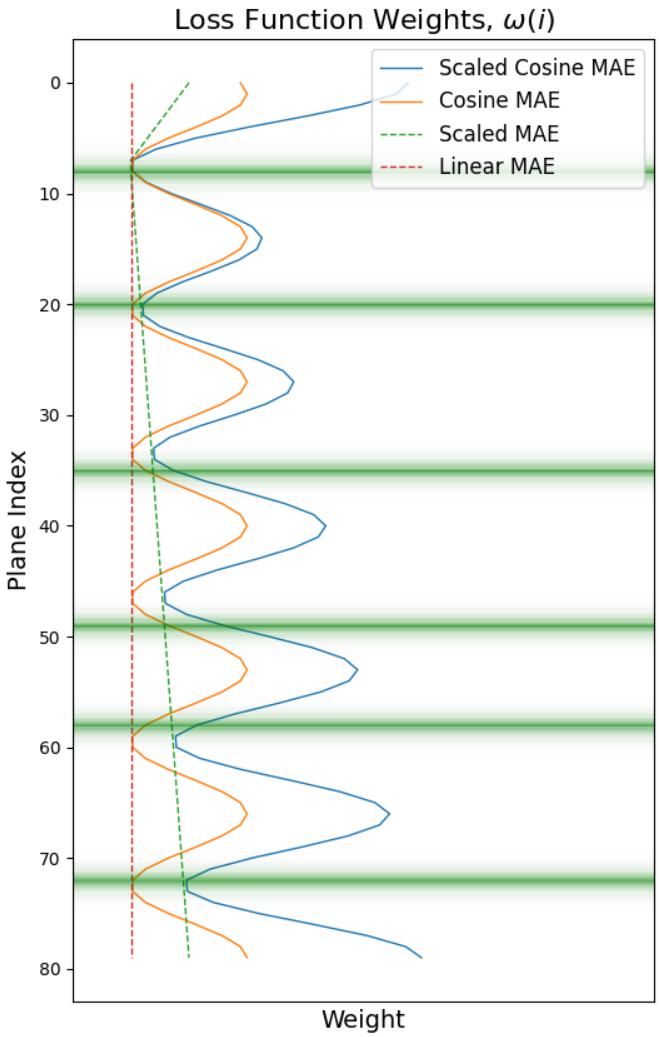

Figure 6.8: *Visualization of loss function weight distributions across imaging planes.* The plot compares different weighting strategies for the loss function: MAE, scaled MAE, Cosine–MAE, and scaled Cosine-MAE. The weights are plotted as a function of plane index, illustrating how different weighting schemes emphasize certain planes more than others. The uncertainty bands (green) indicate the approximate positional variance of the imaging planes, providing insight into the weighting's impact on reconstruction accuracy.

$$\theta_k^{t+1} = \theta_k^t - \eta \cdot \frac{\partial L}{\partial \theta_k},$$

where the weighted gradient update is:

$$\frac{\partial L_{\cos,\text{MAE}}}{\partial \theta_k} = \frac{1}{N} \sum_{i=1}^{N} w(i) \cdot \text{sign}(y_i - \hat{y}_i) \cdot \frac{\partial \hat{y}_i}{\partial \theta_k}.$$

Planes with higher weights contribute larger gradients, accelerating convergence for underrepresented regions while maintaining stability for more reliable structures.

6.4.3 Conclusion

By integrating both Cosine-Weighted MAE and SSIM loss functions, the training pipeline balances pixel-wise fidelity, perceptual quality, and structural consistency. The inclusion of cosine-weighted loss ensures improved generalization by redistributing optimization focus toward planes that require more refinement. This approach enhances both spatial and perceptual accuracy, leading to higher-quality reconstructions in TranSIM-generated volumetric data. The total loss function is defined as a follows:

$$\mathcal{L}_{\text{total}} = \lambda \cdot \mathcal{L}_{\text{Cos,MAE}} + (1 - \lambda) \cdot \mathcal{L}_{\text{SSIM}}. \tag{6.6}$$

6.5 Biological Sample Preparation

The immobilization and preparation of Drosophila melanogaster samples (larvae and adult) were critical steps in this study to ensure optimal imaging conditions and accurate calcium imaging data. A combination of mechanical, cold anesthesia, and chemical methods was employed to achieve effective immobilization while preserving the physiological integrity of the larvae. The procedure is outlined in detail below.

6.5.1 Drosophila melanogaster Larvae Stage

To facilitate the immobilization process, three key components were prepared:

Pluronic F127 Hydrogel Solution: A 25% (weight/volume) hydrogel solution was prepared by dissolving Pluronic F127 (PF127) in distilled water as per the method described by Dong et al[67]. This hydrogel served as a physical restraining medium for the larvae.

HL3.1 Saline Solution: A standard physiological saline solution, HL3.1, was prepared to act as a medium and to rinse the larvae. The preparation followed the protocol established by Feng et al [68]. Chemical Anesthesia Chamber: A custom gas chamber was designed to expose larvae to diethyl ether fumes. Approximately 10 mL of diethyl ether was placed in the chamber to serve as a chemical anesthetic, following the approach described by Kakanj et al[69]. In addition, a glass slide equipped with 100-µm spacers and an 18 mm × 18 mm coverslip (thickness: 0.13–0.17 mm) were prepared for mechanical immobilization.

Second instar larvae were selected for the experiments. The larvae were identified from 4–5-day-old media bottles, with day zero defined as the day the parent flies were introduced. Larvae were gently extracted from the media using a fine brush and placed on a prepared glass slide. The slide containing the larvae was then placed inside the chemical anesthesia chamber at a distance of approximately 7 cm from the diethyl ether source. The chamber was stored in a refrigerator maintained at a temperature of 4°C for 30 minutes. This step combined chemical anesthesia with cold exposure to effectively immobilize the larvae without causing harm.

Upon removal from the anesthesia chamber, each larva was rinsed with 0.5 µL of HL3.1 saline solution to remove residual media and prepare the surface for hydrogel application. The HL3.1 solution also provided a suitable medium for subsequent calcium imaging. Following the rinse, 20 µL of the 25% PF127 hydrogel solution was applied directly over each larva. The hydrogel served as a physical restraining agent, immobilizing the larvae for imaging. An 18 mm × 18 mm coverslip was gently placed over the larvae, applying uniform mechanical pressure. The 100-µm spacers ensured controlled flattening of the larvae, enhancing the visibility of the brain without causing excessive compression or harm.

The slides were transferred to a heating plate set to 35°C for 2–3 minutes to accelerate the solidification of the PF127 hydrogel. This step ensured that the larvae remained immobilized throughout the imaging process. After solidification, the samples were ready for calcium imaging experiments. The combination of chemical anesthesia, cold exposure, and mechanical immobilization ensured minimal movement and optimal transparency for imaging. The use of PF127 hydrogel provided a biocompatible and stable environment, while the HL3.1 solution maintained physiological conditions. The described procedure minimized potential harm to the larvae and preserved their brain activity for accurate imaging and analysis.

6.5.2 Drosophila melanogaster Adult Stage

Two *D. melanogaster* lines were used for imaging. A GFP line [**strain name**] was used to gather static images for neural network training, and a pan-neuronal GCaMP6f line was used to image dynamic signals for inference. Flies were raised on standard cornmeal medium at 50% humidity and 25°C under a 12-hour light/dark cycle. Imaging was performed on 5–10-day-old female flies[70].

Flies were cold-anesthetized at 4° C for 30 minutes and then placed on a metal shim etched with a 0.75 mm × 2 mm hole housed in a custom 3D-printed mount. The mount was kept on a cold plate at approximately 6°C for the duration of sample preparation. The thorax was pushed through the hole up to the wing nubs while the abdomen and head remained beneath the shim. The head was tilted downward until its posterior surface was parallel with the shim, exposing the posterior dorsal head cuticle. The fly was fixed in place by applying UV glue (Esslinger, #12.201) around the posterior dorsal cuticle and on either side of the thorax. The glue was cured with 3-second bursts of 365 nm light.

During dissection, the brain was bathed with saline solution (103 mM NaCl, 3 mM KCl, 10 mM trehalose, 10 mM glucose, 2 mM sucrose, 26 mM $NaHCO_3$, 1 mM NaH_2PO_4, 1.5 mM $CaCl_2$, 4 mM $MgCl_2$) to prevent desiccation. Superfine forceps (Dumont, #00) and a tungsten dissection needle (Fine Science Tools, #05) were used to open a window in the

right side of the head capsule, exposing the optic lobe. All lipids, muscles (muscles 1 and 16), and post-ocular air sacs occluding brain tissue were excised. Flies were then visually inspected for coherent bilateral movement to ensure sample viability.

CHAPTER 7

Optical Performance and Direct Acquisition

The results presented in this chapter demonstrate the improved photon efficiency, enhanced imaging performance, and expanded biological applications of the latest TranSIM system. This iteration introduces several key advancements over my previous Master's work, including a streamlined optical setup, refined image acquisition strategies, and the integration of computational techniques to achieve high-resolution volumetric imaging.

A major modification in this updated design is the transition from a three-camera configuration to a two-camera setup. Additionally, I significantly simplified the illumination scheme, reducing overall system alignment complexity while maintaining high imaging fidelity. Furthermore, I implemented a novel acquisition strategy in which the image planes are captured in reverse order—from the last to the first. This adjustment ensures a more balanced photon distribution across the imaging planes, mitigating potential signal degradation and improving the overall plane-to-plane dynamic range of the system.

With these improvements, I expanded the scope of TranSIM's biological validation to include additional animal models. Beyond previous experiments, this work extends TranSIM's application to *Drosophila melanogaster*, *Danio rerio*, and murine models, demonstrating its versatility across different biological systems. Furthermore, I developed and implemented an inter-planar axial alignment processing pipeline, addressing the inherent staggered plane misalignment in the raw imaging data. This computational correction ensures accurate volumetric reconstructions, eliminating artifacts caused by spatial misalignment and enabling more precise structural analysis.

Furthermore, TranSIM's ability to perform spatiotemporal compression enhances its adaptability for diverse imaging applications. By adjusting the rescanning waveform, spa-

tial compression enables a two-fold increase in the imaged field of view without altering scanning. Together, these advancements mark a significant evolution of the TranSIM system, integrating optical, computational, and biological innovations to push the boundaries of high-speed volumetric microscopy. The subsequent sections present detailed quantitative and qualitative analyses of these improvements, illustrating the enhanced performance and broader applicability of the system.

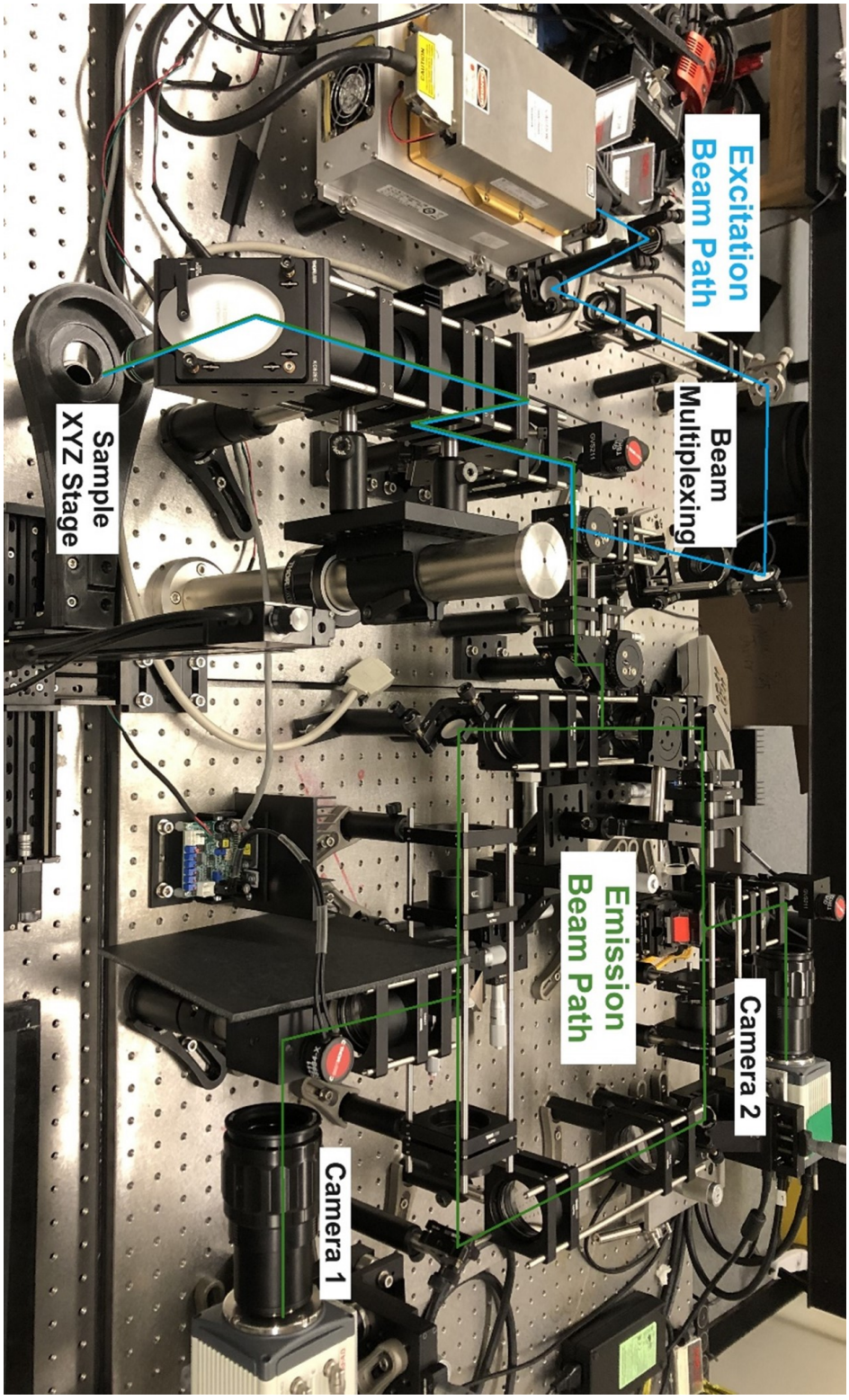

Figure 7.1: *TranSIM System Hardware.*

7.1 Hardware Implementation

In Section 6.1, I described the optical design of the system. In this section, I present the real-world hardware implementation, detailing the physical construction and integration of the components. Figure 7.1 provides a comprehensive schematic of the entire system, illustrating the full optical path from illumination to detection.

The process begins with the illumination beam path, where a 488 nm, 100 mW laser source is expanded using a beam-expanding telescope to ensure a uniform, collimated beam profile. This expanded beam is then focused into a line using a cylindrical lens, effectively shaping the excitation profile for structured illumination. The beam is subsequently multiplexed both axially and laterally to generate multiple distinct focal planes within the sample. A detailed view of the multiplexing module is provided in Figure 7.2, where the precise arrangement of optical elements responsible for generating the staggered illumination pattern is illustrated.

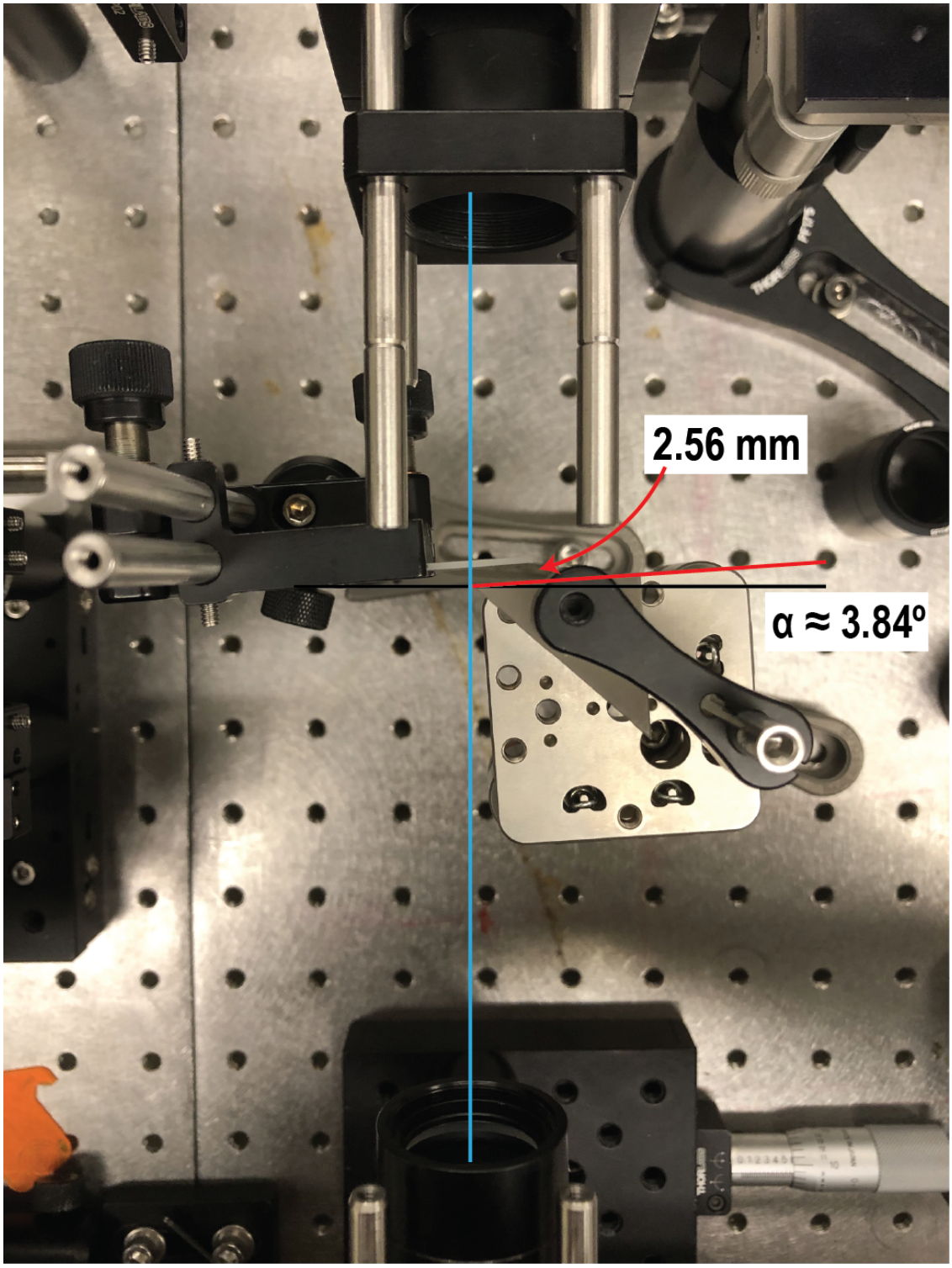

Figure 7.2: *Beam Multiplexing Module.* Measured spatial separation between 90:10 beam-splitter and depth reflection paper placed at 2.56 mm, with rotation angle $\alpha \approx 3.84^o$.

Once the beam has been multiplexed, it is collected by a relay lens and directed towards a scanning galvanometer via a dichroic mirror. The galvanometer dynamically steers the beam lines across the sample, enabling rapid scanning of the imaging planes. The fluorescent emission from the sample follows the same optical path in reverse, passing back through the scanning optics, where it is de-scanned by the galvanometer to correct for the applied motion.

After de-scanning, the fluorescence signal continues through the dichroic mirror and is directed onto a wedge mirror, which carefully aligns the beam into a resonator composed of four precisely arranged 4f relay lens pairs. Within the resonator, knife-edge mirrors are strategically placed in the beam path to "trim" image planes before relaying them toward the rescanning galvanometers. These rescanning galvanometers subsequently direct the signal onto high-sensitivity sCMOS sensors, effectively capturing the final multi-plane fluorescence images.

Enough degrees of freedom in the optical system are required to ensure proper alignment. Typically this involved placing the cameras are on tilting stages to ensure that the fluorescent signal lines can be aligned with the rolling sCMOS aperture. Moreover, on the last 4f relay, I placed the optical components on x, y, and z degrees of freedom to allow for the passing of the remainder planes underneath the D-shaped mirror.

7.2 Point Spread Function Analysis

To characterize the optical performance of TranSIM, 1 µm microspheres (TetraSpeck, ThermoFisher) were used for point spread function (PSF) analysis. Prior to imaging, the beads were ultrasonicated for 10 minutes to disperse any aggregated particles. Following sonication, 5 µL of a 2.0×10^{10} particles/mL stock solution was mixed with 50 µL of a 2% agarose (W/V) solution to ensure uniform suspension. The resulting solution, with a final concentration of 1.82×10^{9} particles/mL, was then transferred onto a glass slide with 150 µm spacers. A coverslip was placed on top to create a stable, flat imaging surface optimized for high-resolution measurements.

A detailed PSF analysis was conducted to validate TranSIM's imaging capabilities. Fig-

ure 7.8 presents cross-sectional views of the PSF in the XY and YZ planes, illustrating the system's spatial resolution in both lateral and axial dimensions. The PSF was analyzed across multiple imaging planes to assess the uniformity of resolution throughout the imaging volume. Gaussian fitting was applied to the intensity profiles to extract key optical metrics, including the Full-Width-Half-Maximum (FWHM) values. The results indicate that TranSIM achieves near-diffraction-limited performance, with FWHM values of 4.28 µm in the Z-dimension, 1.35 µm in the Y-dimension, and 0.99 µm in the X-dimension at the first imaging plane. These values remain consistent across the remaining imaging planes, highlighting the system's ability to maintain high-resolution imaging across an extended depth range.

Beyond resolution measurements, the optical efficiency of TranSIM was assessed by quantifying the relative intensity drop as a function of imaging depth. Figure 7.8 presents the intensity variation across imaging planes, providing insights into the system's capacity to maintain signal strength throughout the imaging volume. The observed uniformity in intensity distribution ensures accurate volumetric reconstruction, reducing artifacts associated with uneven illumination or signal attenuation.

These analyses confirm that TranSIM not only provides high-resolution imaging across multiple planes but also preserves signal integrity, which is crucial for accurate 3D imaging. The combination of PSF validation and intensity assessments establishes the system's robust optical performance, reinforcing its capability for high-speed, high-fidelity volumetric imaging.

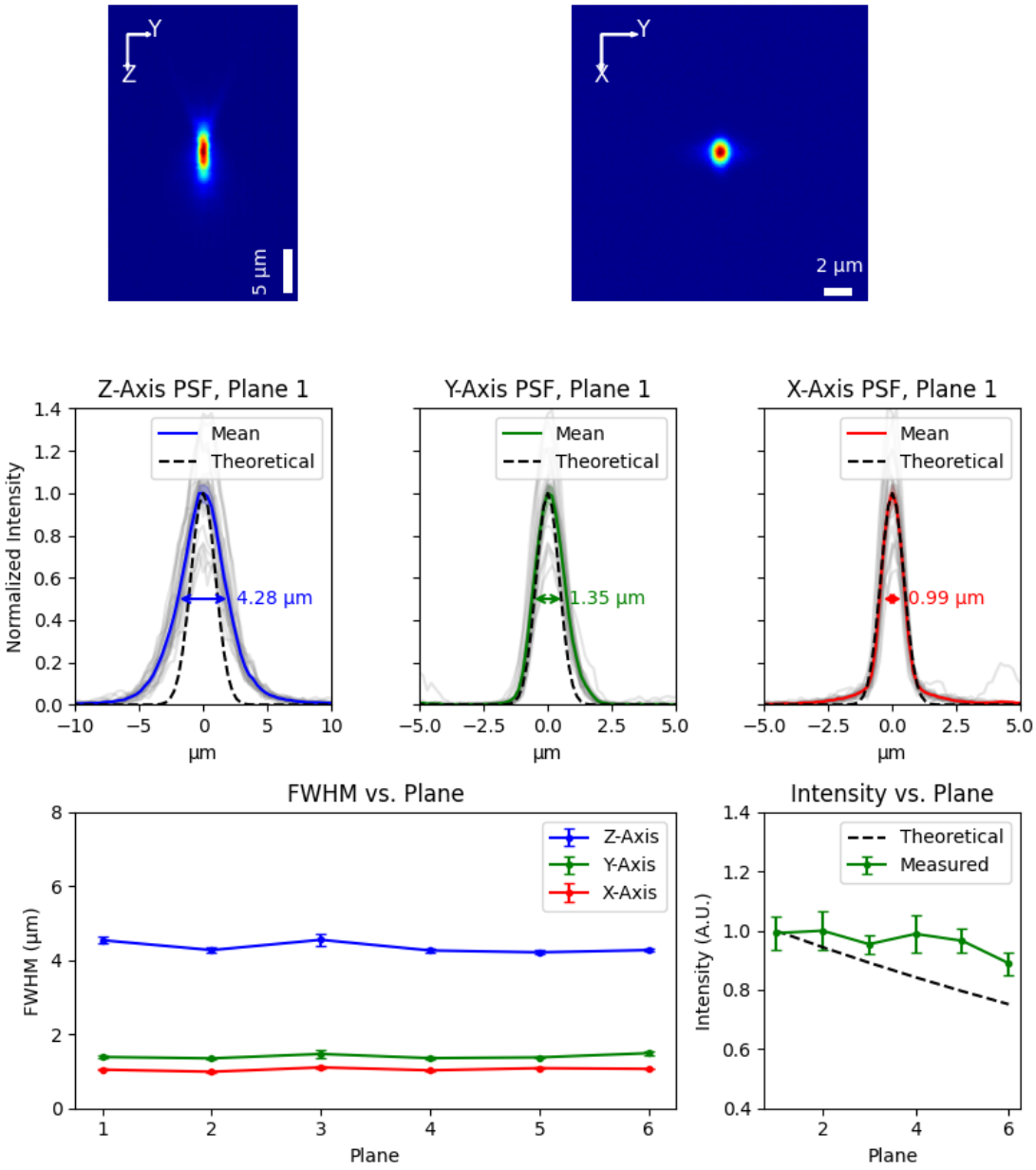

Figure 7.3: *Point Spread Function Analysis of All Planes.* Point spread function and intensity analysis of the system with cross-sectional views and independent XYZ analysis by plane. Utilized 1 µm beads and theoretical PSF was calculated using the modified FWHM taking the bead's size into account denoted in detail in the methods section. Nikon 40X 0.8NA objective was used to increase spatial sampling to isolate the possibility of aliasing by the lower magnification of the Nikon 16X 0.8NA objective.

7.2.1 Inter-Planar Alignment and Volumetric Field-of-View

To validate that the illumination planes correctly formed the expected parallelepiped volume and its magnified counterpart at the sample, a quantitative analysis of the inter-planar alignment was performed. As previously described in the Methods section, the lateral separation between the imaging planes is expected to scale inversely with magnification ($1/M$), while the axial separation follows an inverse quadratic dependence ($1/M^2$). These transformations ensure that the final imaging volume retains the appropriate geometric proportions for accurate volumetric reconstruction.

To assess the accuracy of inter-planar alignment, 4 m microspheres (TetraSpeck, ThermoFisher) were used in a controlled sample preparation designed to generate a thin, selective plane of fluorescent beads. The larger bead size was chosen for its increased brightness, enhancing visibility and facilitating precise localization across imaging planes. To achieve a confined distribution, the beads were suspended in a 2% agarose (W/V) solution and compressed into a thin layer between a glass slide and a coverslip with 50 m spacers. This configuration enabled a well-defined focal plane of beads, which was subsequently processed using the inter-planar alignment pipeline.

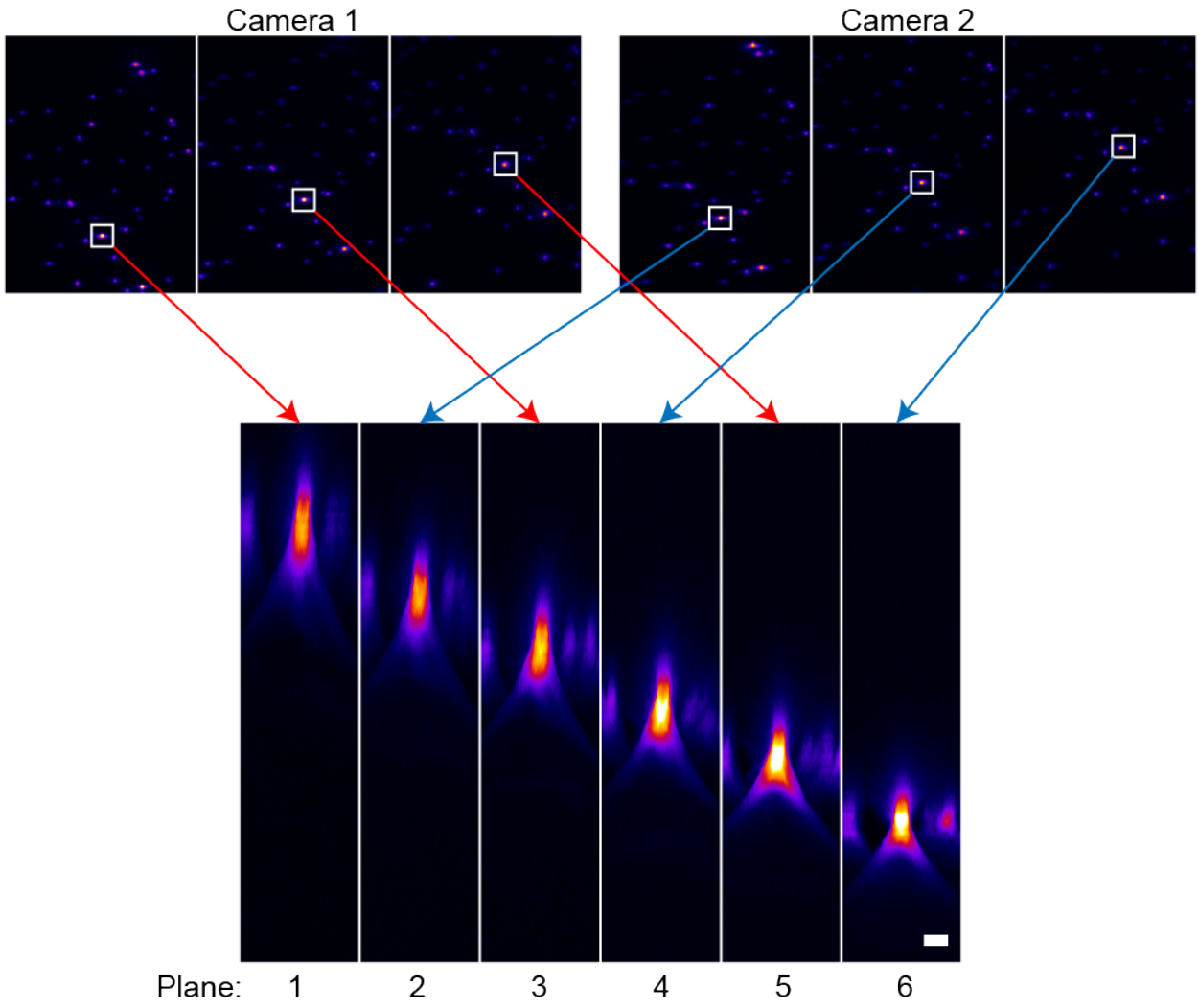

Figure 7.4: *Inter-Planar Alignment Validation.* The top row shows the maximum intensity projection (MIP) images captured from two cameras, with arrows indicating the sequential positioning of imaging planes. The bottom row traces a single 4 μm bead as it appears across different depths. Scale bar is 10 μm.

Figure 7.5 presents an analysis of the inter-planar spacing under magnified conditions. To quantify the axial and lateral shifts between planes, a cross-correlation-based alignment pipeline was employed. Gaussian fitting was applied to the cross-correlation maxima to determine the optimal transformation matrices, which were then cumulatively applied across all imaging planes. This analysis provided precise estimates of the effective inter-planar dis-

tances and confirmed that the observed transformations adhered to the expected geometric scaling laws with a measured average inter-planar distance of 20 μm with a total volume height of 100 μm. Furthermore, lateral shifts were also observed to fall within the theoretical bounds, with a vertical lateral shift per plane of 24 μm with a total lateral shift of 120 μm. Ratiometric analysis of axial plane separation and lateral shifts results in a parallelepiped volume with an angle $\alpha = 40°$. Additionally, there was some measurable horizontal lateral shifting produced by misalignment. However, this demonstrates the capabilities of the processing pipeline to correctly align the planes.

The results demonstrate that the lateral and axial spacing between imaging planes remained consistent with theoretical predictions across different magnification settings. The volumetric field-of-view was reconstructed using the calculated transformations, verifying that TranSIM effectively maintains spatial coherence across all imaging depths. The quantified inter-planar distances further validate the system's ability to perform high-fidelity volumetric imaging without significant spatial distortion.

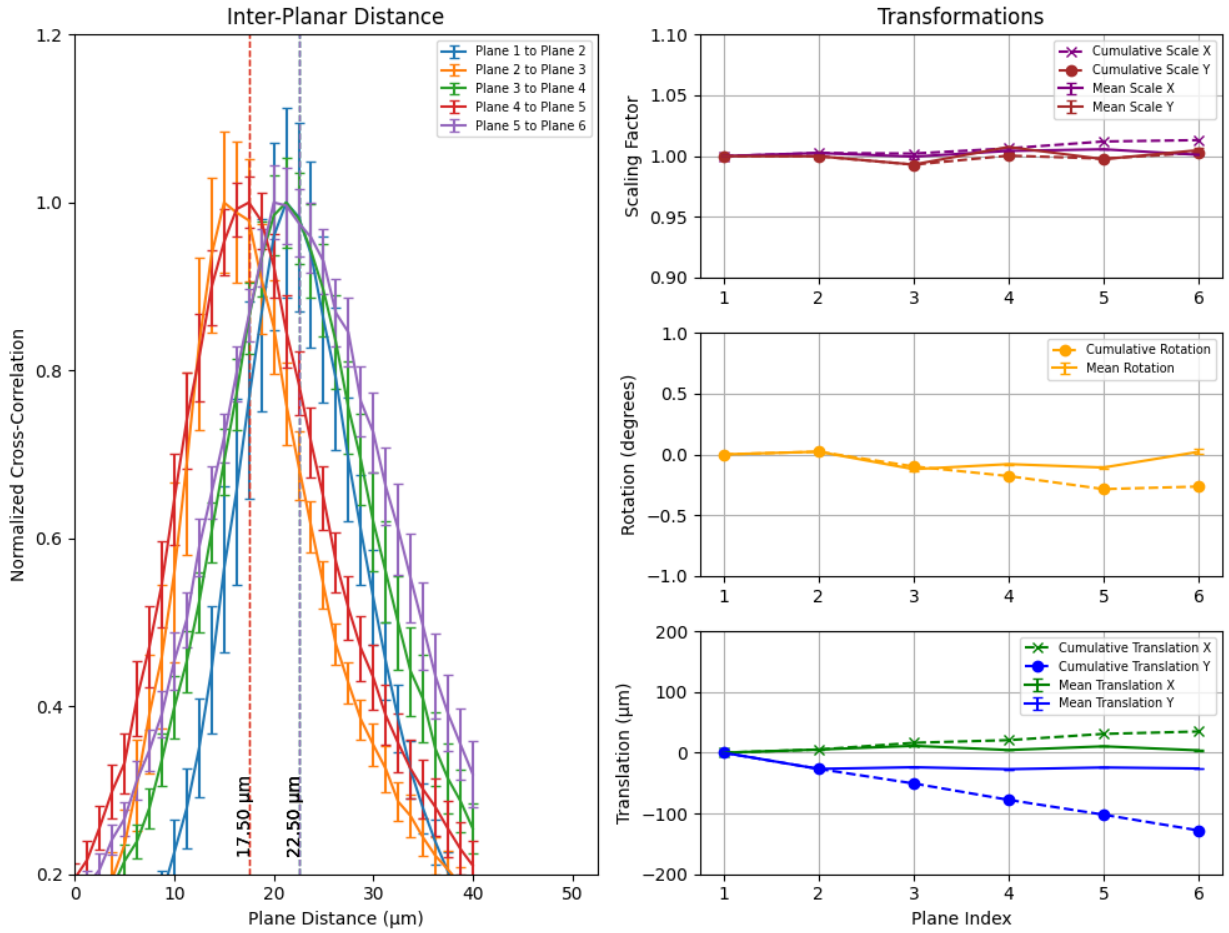

Figure 7.5: *Inter-Planar Alignment and Volumetric Scaling.* The figure illustrates the measured inter-planar distances as a function of magnification. Cross-correlation-based analysis was employed to quantify lateral and axial shifts between planes, ensuring proper formation of the imaging volume. The results confirm that lateral separation scales as $1/M$ while axial separation scales as $1/M^2$, in accordance with theoretical predictions. This validation ensures that TranSIM maintains accurate spatial relationships across imaging planes, critical for high-resolution volumetric reconstruction.

7.3 Illumination and Detection of Simultaneous Imaging Planes

The illumination and detection of simultaneous imaging planes are the cornerstone of TranSIM's design and functionality. By parallelizing the acquisition of imaging planes, TranSIM enables the rapid and efficient capture of volumetric data, significantly enhancing temporal resolution. The system achieves this by creating and detecting multiple imaging planes that are separated both laterally and axially, allowing for a comprehensive three-dimensional representation of the sample in real time.

This approach eliminates the need for traditional axial scanning mechanisms, which are commonly employed in conventional microscopy systems such as light-sheet and confocal microscopy. In these systems, volumetric imaging requires the sequential acquisition of individual planes, necessitating mechanical or optical adjustments along the z-axis. This sequential process inherently limits imaging speed and can introduce artifacts due to motion or misalignment. TranSIM's innovative beam multiplexing method circumvents these challenges by simultaneously illuminating and imaging multiple planes, drastically reducing the time required for volumetric data acquisition.

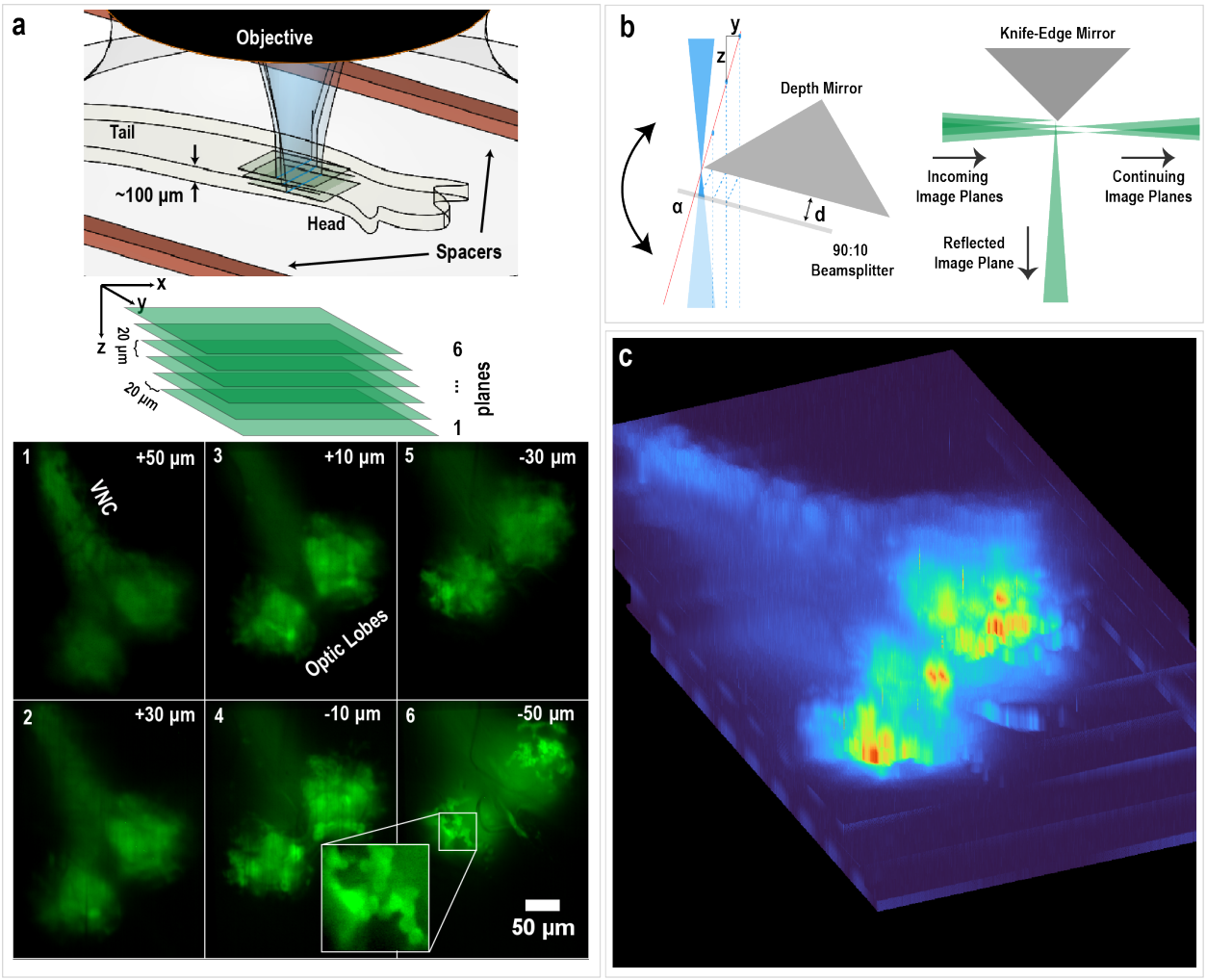

Figure 7.6: *The working principle of Transverse Sheet Illumination Microscopy.* (A) Six laterally and axially separated beams illuminate the sample, creating a parallelepiped volume. Two sensors capture these planes simultaneously, with one sensor recording odd-numbered planes (1, 3, 5) and the other capturing even-numbered planes (2, 4, 6). A zoomed-in view highlights neurons in the optic lobe. (B) The illumination beams are generated by spatially multiplexing a focused laser beam using a 90:10 beamsplitter and a reflective mirror cavity, producing separated beams based on mirror angle and spacing. The fluorescence signals retain their separation and are directed to an imaging sensor via a knife-edge mirror. (C) A 3D reconstruction of the imaging planes after alignment, showing a Drosophila melanogaster instar 2 larval brain.

7.4 Validating on Animal Models

To assess the versatility and robustness of TranSIM across different biological applications, I validated the system using multiple animal models, each selected to highlight distinct imaging capabilities. These models include zebrafish larvae for cardiovascular imaging, Drosophila melanogaster larvae and adults for neural activity studies, and mice for deep-tissue neuronal imaging. Each validation experiment was designed to test TranSIM's ability to capture high-speed dynamics, resolve fine structural details, and maintain spatial and temporal fidelity across different tissue types and depths.

For cardiovascular imaging, I utilized Danio rerio (zebrafish) larvae genetically modified to express GFP in cardiovascular tissues. This model provided an ideal testbed for assessing TranSIM's ability to capture rapid motion at high temporal resolution, particularly the beating heart. By imaging the heart in real time, I evaluated TranSIM's capacity to resolve fast-moving biological structures without motion artifacts, demonstrating its potential for functional imaging in developmental biology and cardiovascular research.

To validate TranSIM's ability to image neural activity, I conducted experiments on Drosophila melanogaster larvae and adults. Drosophila larvae were used to capture neuronal firing patterns in a developing nervous system, while adult flies provided a more complex neural architecture for testing TranSIM's performance in a densely connected network. These experiments focused on detecting fluorescently labeled neurons in the brain and visualizing spontaneous and stimulus-evoked activity at high temporal resolution. The system's capability to simultaneously image multiple planes ensured comprehensive volumetric coverage, preserving fine structural details while maintaining a high frame rate.

For deep-tissue neuronal imaging, I applied TranSIM to mouse brain tissue expressing genetically encoded calcium indicators. This model tested the system's ability to penetrate deeper into biological tissue while preserving high signal fidelity. The mouse brain presents a more challenging optical environment due to increased scattering and absorption, making it a rigorous test case for TranSIM's sensitivity and imaging depth. The ability to capture neuronal activity in real-time across multiple planes reinforced TranSIM's applicability to

neuroscience research, particularly for studying large-scale network dynamics in mammalian models.

Through these validations, I demonstrated TranSIM's capability to function across diverse biological systems, proving its efficacy in both high-speed functional imaging and structural mapping. Each animal model provided critical insights into the system's strengths and limitations, helping refine its application for future studies in neuroscience, cardiovascular research, and beyond.

7.4.1 Zebrafish Larvae Beating Heart

Studying the heart dynamics of zebrafish is crucial for understanding fundamental principles of cardiovascular development, function, and disease. Zebrafish possess a remarkable capacity for heart regeneration, a trait that is largely absent in mammals, making them a valuable model for investigating cardiac repair mechanisms. Their transparent embryos allow for direct, non-invasive optical imaging of the beating heart in real time, providing unique insights into cardiac structure, blood flow dynamics, and electro-physiological activity. Because zebrafish share many genetic and molecular pathways with humans, studying their heart function helps elucidate conserved mechanisms of cardiac development and disease progression. By using high-speed imaging techniques such as TranSIM, researchers can capture fine details of zebrafish heart contractions, valve function, and chamber interactions, offering unprecedented temporal and spatial resolution for cardiovascular research, see Figure 7.7.

Beyond fundamental biology, zebrafish serve as an essential model for studying human cardiovascular diseases and testing potential therapeutic interventions. Their rapid development and high genetic manipulability enable researchers to generate models of congenital heart defects, arrhythmias, and cardiomyopathies, providing a platform for high-throughput screening of drugs targeting cardiac dysfunction. Additionally, the ability to image heart dynamics in vivo allows for the real-time assessment of pharmacological and genetic interventions, bridging the gap between molecular studies and translational medicine. Understanding

how zebrafish hearts respond to injury, environmental stressors, or gene mutations can offer novel therapeutic insights applicable to human heart disease, particularly in the fields of heart regeneration and tissue engineering. Thus, investigating zebrafish heart dynamics not only deepens our knowledge of cardiac biology but also advances clinical research aimed at improving cardiovascular health.

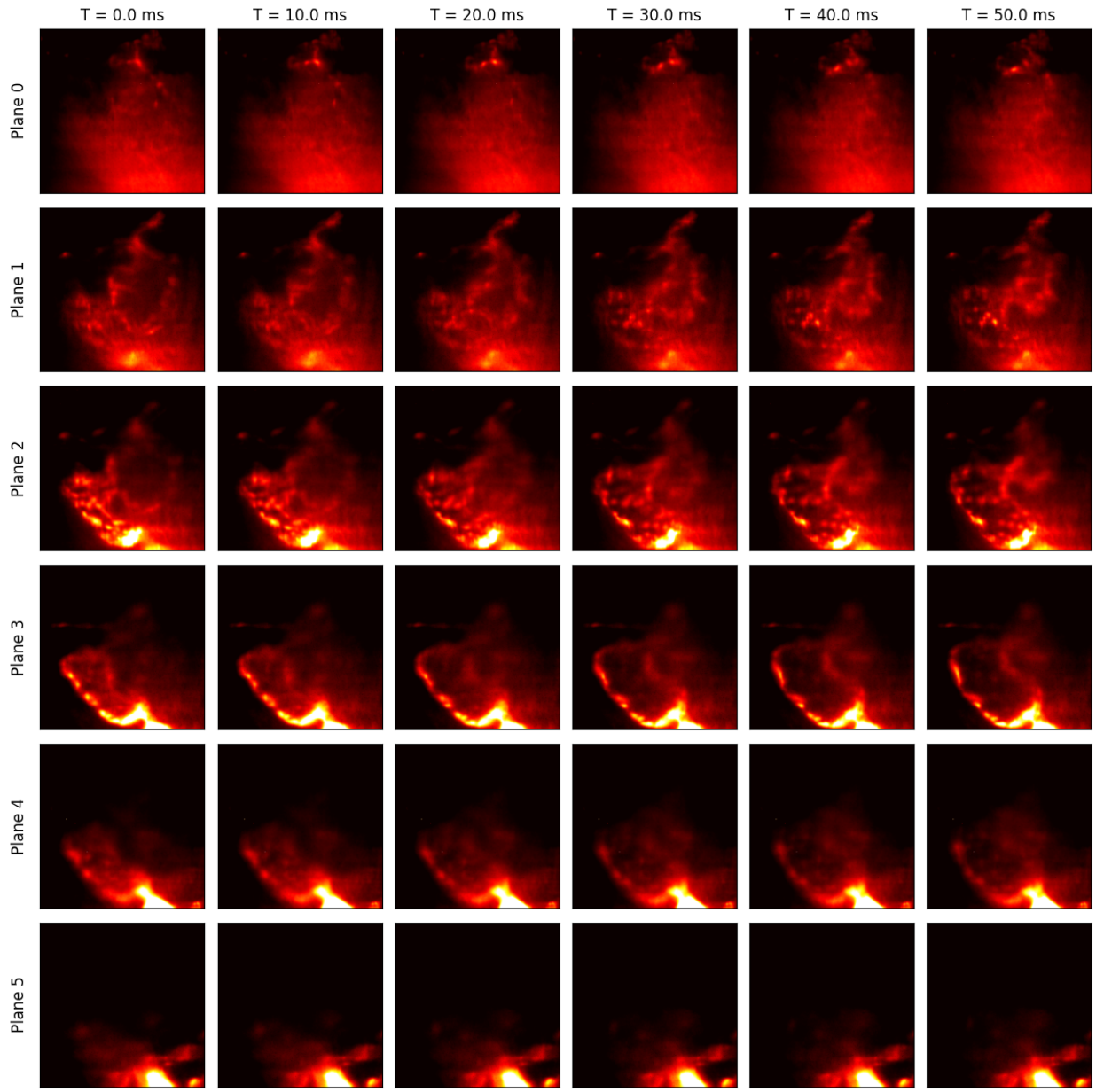

Figure 7.7: *Time-resolved Imaging of Zebrafish Heart.* Each row represents a different imaging plane (depth), spanning a total of 120 μm, while each column corresponds to a different time point (spanning 6 frames, each 10 ms apart). In planes 0–2, the atrium is observed closing, whereas in planes 3–5, the ventricle is seen opening. This simultaneous volumetric capture enables the study of cardiac dynamics across multiple depths in a single acquisition.

7.4.2 Larval Stage Fruit Fly Neural Dynamics

I imaged Drosophila instar 2 larvae at 100 Hz. I identified more than 200 neurons exhibiting significant neuronal activity, selecting them based on signal-to-noise (SNR) thresholding rather than morphology alone. The larvae were immobilized using hydrogel and mechanical pressure, as described in the methods section. This result demonstrates my system's ability to maintain both spatial and temporal fidelity during high-speed imaging at 100 Hz. The detection of such a large number of active neurons highlights TranSIM's robust sensitivity, even in a dynamic, live organism. The simultaneous imaging of multiple planes ensured that neuronal activity across varying depths could be effectively captured without compromising resolution or intensity.

To extract neuronal traces, I utilized the CaImAn Python processing pipeline. This tool initially detected over 5,000 potential neurons, but most were rejected based on SNR, allowing me to focus exclusively on well-defined signals. This approach emphasized the temporal properties of TranSIM and provided a thorough characterization of its performance. The analysis revealed clear temporal patterns of neuronal firing, further validating TranSIM's ability to resolve dynamic neural processes with high precision at significant depths. These results not only showcase TranSIM's technical advancements but also underscore its potential as a transformative tool for neuroscience research, enabling the exploration of spatiotemporal dynamics and neural connectivity in unprecedented detail.

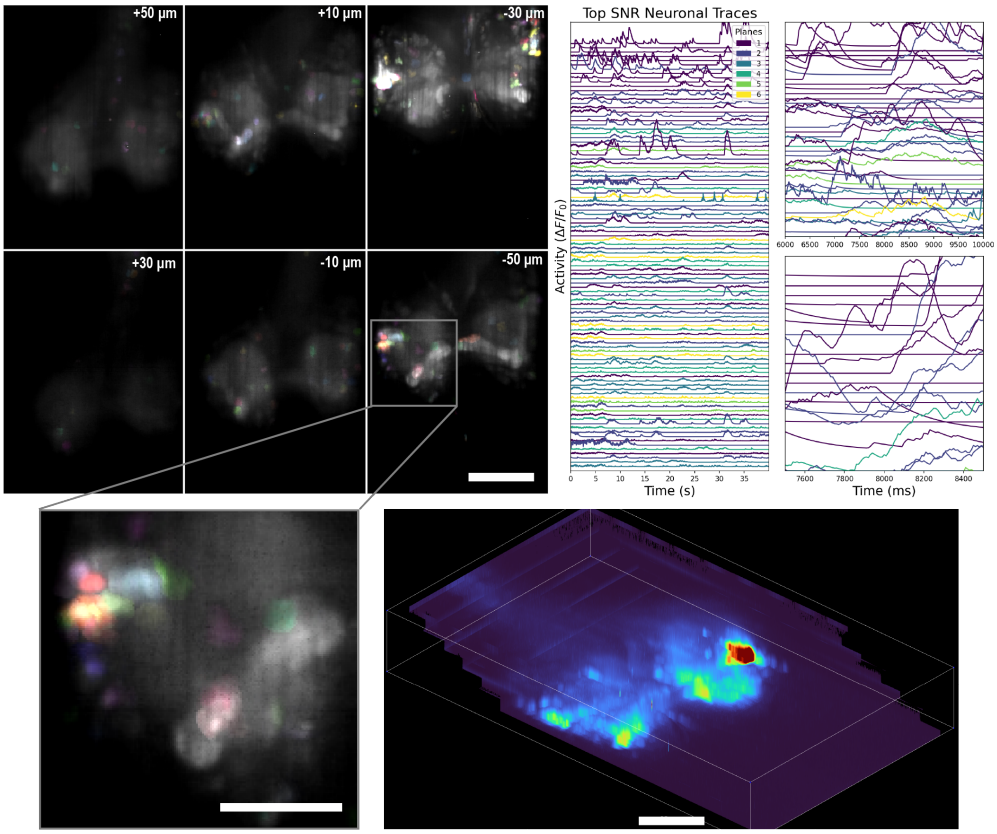

Figure 7.8: *Neural dynamics in Drosophila melanogaster larvae.* An instar 2 larva was imaged at 100 volumes per second across 6 planes with an average interplanar distance of 20 µm. The top panel (+30 µm) captures imaging approximately halfway through the optic lobe, while the last plane (-30 µm) was positioned near the ventral side at the bottom of the optic lobe. At the center of the optic lobe, we observed a high density of active neurons, which decreased further into the sample, leaving only a few neurons with a distinct commissure throughout several planes. Scale bar in the tile panel view is 100 µm. Zoomed in to panel +30 µm, we can see high neuron density clustering and not all neurons being selected. Scale bar in zoomed view is 50 µm. The top 100 neuronal traces were plotted over a 40-second time window, with zoomed-in subplots of 8 and 2 seconds to highlight fast neural dynamics. A three-dimensional rendering of the reconstructed volume, visualized via Napari, showcases the alignment of planes and highlights the staircase configuration characteristic of TranSIM. Scale bar in the three-dimensional reconstruction is 50 µm.

7.4.3 Adult Stage Fruit Fly Neural Dynamics

In the previous section, I validated the use of TranSIM on larval stage fruit flies. However, due to their simplistic nature, much of the observed neural activity was purely reflexive. To further validate the system, I extended the analysis to adult-stage *Drosophila melanogaster* using both head-fixed and head-dissected protocols, as described in the materials and methods chapter. These imaging results reveal significantly more complex neuronal activation patterns, with highly correlated neural activity across different regions of the brain. This further underscores the necessity of high-speed imaging systems like TranSIM to capture these rapid and spatially distributed neural events.

Given the large number of neurons detected, most of the active neurons identified using the CNMF processing pipeline were concentrated near the surface of the optic lobe. However, despite the scattering and optical limitations inherent to deep-tissue imaging, a subset of neurons was still identified beyond $+100\ \mu m$ into the optic lobes. In Figure 7.9, we observe the highly dynamic activation of neurons in adult-stage *Drosophila*, highlighting the increased complexity of neural interactions compared to the larval stage.

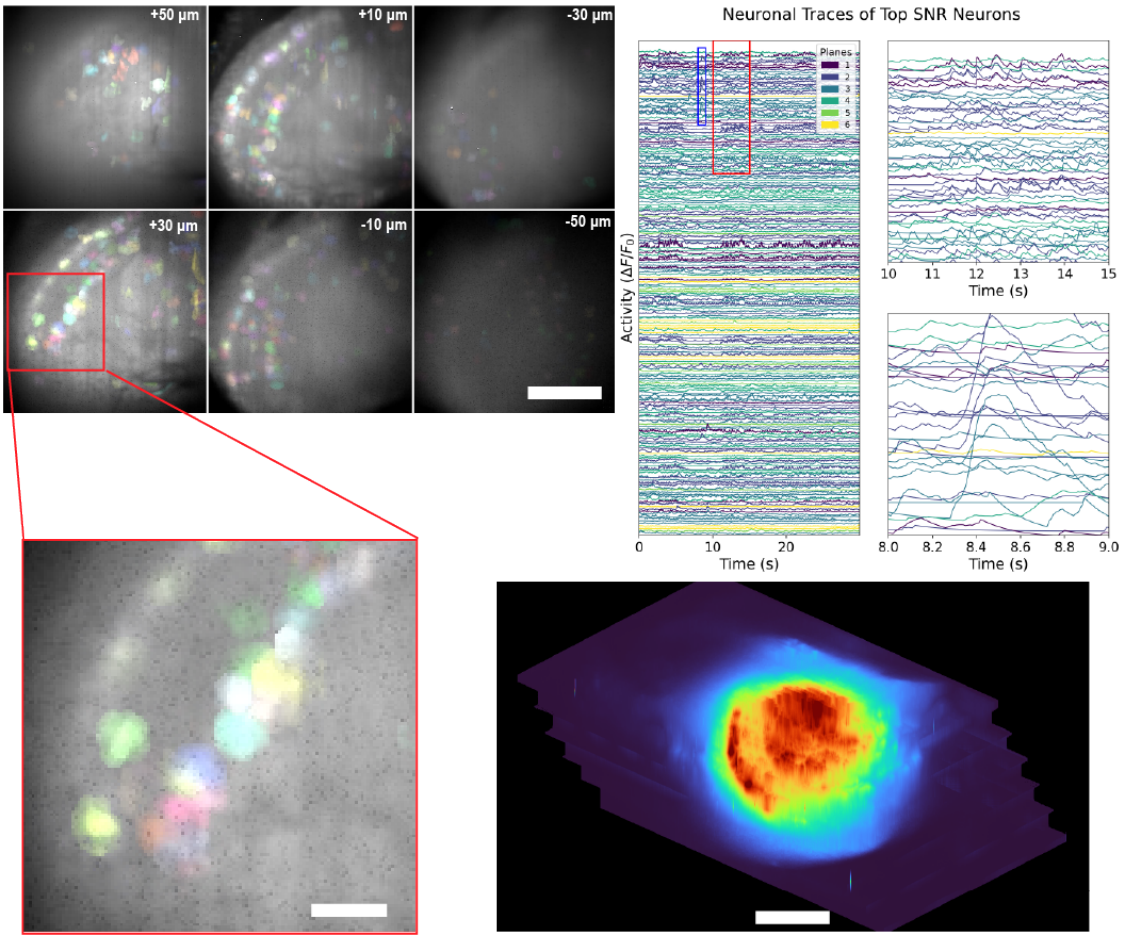

Figure 7.9: *Neural dynamics in adult Drosophila melanogaster.* An adult-stage fruit fly was imaged at 100 volumes per second across 6 planes with an average inter-planar distance of 20 µm. The top panel captures the surface of the optic lobe, while the deepest plane attempts to visualize structures inside the optic lobe. At the center of the optic lobe, we observed a high density of active neurons, which gradually decreased further into the sample, leaving only a few identifiable neurons. The tile-panel view scale bar represents 100 µm. In the zoomed-in panel at +30 µm, a high-density clustering of neurons is evident, though not all neurons were selected during processing. The scale bar in the zoomed view is 50 µm. The top 300 neuronal traces were plotted over a 30-second time window, with zoomed-in subplots of 8 and 2 seconds to highlight fast neural dynamics. A three-dimensional rendering of the reconstructed volume, visualized via Napari, showcases the alignment of planes and highlights the staircase configuration characteristic of TranSIM. The scale bar in the 3D reconstruction represents 50 µm.

To further explore the functional connectivity within the neural circuits of the adult fly, I computed the correlation matrix of the top 200 most active neurons, ranked by signal-to-noise ratio (SNR). The correlation analysis revealed multiple groups of strongly correlated neurons that exhibited synchronized activity, particularly in regions of the optic lobe associated with motion processing. A subset of these correlated neurons was extracted and visualized along with their corresponding spatial locations within the imaging plane, as shown in Figure 7.10.

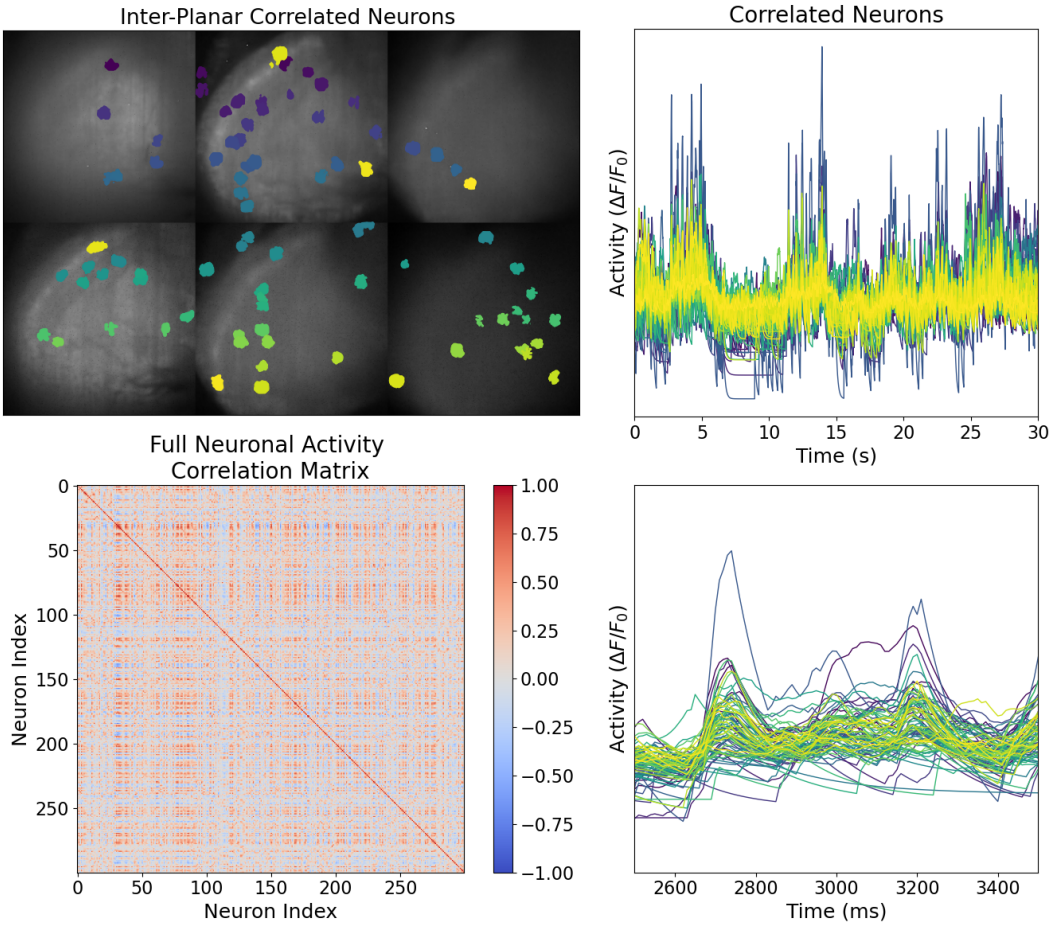

Figure 7.10: *Inter-planar correlated neurons in the adult fly optic lobe.* The figure presents an analysis of correlated neural activity based on signal-to-noise ratio (SNR) ranked neurons. (**Top-left**) A single frame from the imaging stack is displayed with detected neuron contours overlaid. The subset of neurons belonging to a highly correlated group is highlighted using distinct colors. (**Top-right**) The corresponding neuronal traces from the correlated group are plotted over a 30-second window, revealing synchronous activity patterns. (**Bottom-left**) The full correlation matrix of the top 300 neurons is shown, where the color scale represents the strength of correlation. (**Bottom-right**) A zoomed-in section of the neuronal traces illustrates fine temporal dynamics over a 2-second window, demonstrating rapid fluctuations in fluorescence intensity. The results indicate that functionally related neurons within the optic lobe exhibit strong temporal coherence, further validating the ability of TranSIM to resolve high-speed neural activity across multiple planes.

The correlation analysis reinforces the functional organization of neurons within the optic lobe, where groups of neurons exhibit strong temporal coherence, suggesting shared sensory processing roles. The extracted neural traces confirm that correlated neurons fire in synchrony over both long and short timescales. This finding further supports the ability of TranSIM to resolve high-speed neural activity across multiple imaging planes, providing valuable insights into functional interactions in the adult *Drosophila* brain.

7.5 Spatiotemporal Compression

Due to TranSIM's innovative method of de-scanning the image planes, it is possible to manipulate the scanning waveforms to achieve spatiotemporal compression, thereby enhancing the system's versatility in volumetric imaging. This is accomplished by decoupling the scanning galvanometer's input amplitude from the rescanning waveform, allowing for effective vertical magnification changes or data compression. Specifically, when a 374 µm amplitude is supplied to the scanning galvanometer, the rescanning waveform can be adjusted to spatially compress the data. This leads to a reduction in the vertical sampling at the sensor, effectively increasing the imaged field of view while maintaining resolution in the x-axis through horizontal binning. This capability, I termed *spatial compression*, provides the ability to capture larger volumetric fields without altering the physical scanning conditions at the sample.

Additionally, TranSIM supports *temporal compression*, where the rolling shutter region of interest (ROI) at the sensor is reduced, thereby decreasing the number of vertical pixels utilized per frame. By increasing the scanning frequency and reducing the ROI from 920 to 460 pixels, the frame rate can be doubled from 100 Hz to 200 Hz, enabling faster volumetric imaging while preserving the same spatial field of view. This approach not only increases temporal resolution but also reduces the data volume per unit of time.

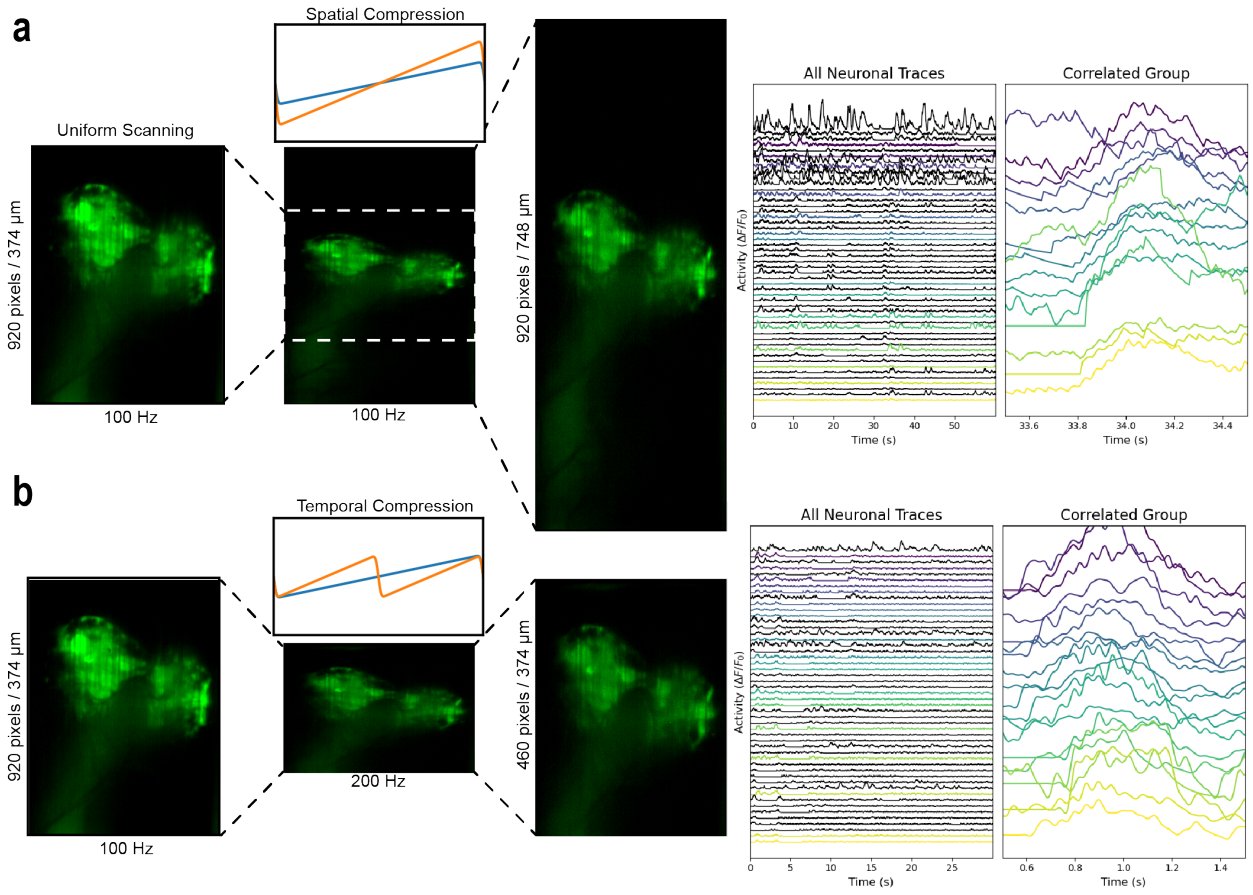

Figure 7.11: *Illustration of Spatiotemporal Compression in TranSIM.* The top row demonstrates spatial compression, where increasing the vertical scanning amplitude effectively reduces vertical sampling at the sensor, allowing for larger volumetric fields of view without compromising resolution. Using neuronal tracing with the CaImAn processing pipeline, we confirmed that spatial compression enables a two-fold increase in the total imaging volume, expanding from $374 \times 278 \times 100\,\mu m^3$ to $748 \times 278 \times 100\,\mu m^3$. The bottom row highlights temporal compression, where reducing the vertical ROI by half (from 920 to 460 pixels) leads to a doubling of the frame rate from 100 Hz to 200 Hz, enabling higher-speed volumetric imaging. These approaches allow for flexible adaptation of TranSIM to different imaging needs, whether prioritizing spatial coverage or high-speed dynamics.

Together, these spatiotemporal compression strategies significantly expand the operational flexibility of TranSIM, making it adaptable to a wide range of imaging scenarios.

Spatial compression is ideal for applications requiring large fields of view, such as whole-organism imaging or wide-area neuronal activity mapping. Conversely, temporal compression is well-suited for studying high-speed dynamic processes, such as calcium signaling or neural activity, in smaller volumes. By integrating these capabilities, TranSIM achieves a balance of spatial and temporal resolution, addressing key limitations of traditional microscopy systems.

CHAPTER 8

Filling the Gap in TranSIM with AI

Thus far, the results presented have demonstrated the direct capabilities of TranSIM. However, the system has one significant limitation: its **sparsity**. With imaging planes separated by approximately 20 μm, there is a risk of missing critical information between planes. This raises an important concern regarding the fidelity of volumetric reconstructions and the completeness of the captured data.

Addressing this challenge requires an innovative solution. In recent years, artificial neural networks have emerged as powerful tools across scientific and engineering domains, ranging from object classification to large language models, and somewhere in between and more relevant to this work, image reconstruction. Specifically, deep learning approaches offer promising techniques for reconstructing missing volumetric data, transforming sparse imaging data into dense, high-fidelity volumes.

In the methods section, I outlined the various neural network architectures explored as potential solutions for enhancing TranSIM's volumetric imaging capabilities. This chapter presents the results of these implementations, demonstrating how artificial intelligence can extend TranSIM into a true 4D microscopy platform. While multiple approaches could be considered to mitigate TranSIM's sparsity, this work focuses only a few neural network approaches due to time constraints.

This chapter is structured as follows: first, I will justify the chosen inter-planar distance and its impact on reconstruction accuracy. Then, I showcase a thorough PSF analysis of raw TranSIM PSF and generated PSFs and their comparison to theoretical simulations. Next, I will present the benefits of using a cosine weighted MAE loss function term in all subsequent neural network loss functions. Then, I will present results from the *time-*

independent neural network approach. Finally, I will discuss the results obtained using the *time-dependent* neural network model, evaluating its effectiveness in preserving temporal coherence in dynamic biological processes.

8.1 Not Too Far, Not Too Close

Choosing the optimal inter-planar spacing when volumetric imaging is a critical parameter that must be optimized to balance spatial resolution, data efficiency, and information redundancy. In TranSIM, the choice of plane separation is constrained by the axial resolution of the system, the optical transfer function (OTF), and the overlap of the point spread function (PSF) in the axial dimension.

To revisit, the axial resolution of a microscope, R_z, is given by:

$$R_z \approx \frac{2 \cdot \lambda}{\mathrm{NA}^2}, \tag{8.1}$$

where λ is the wavelength of light, and NA is the numerical aperture of the objective lens. To ensure accurate volumetric reconstruction, the Nyquist criterion states that the sampling interval must satisfy:

$$d_z \leq \frac{\lambda}{2 \cdot \mathrm{NA}^2}, \tag{8.2}$$

where d_z is the inter-planar spacing. This ensures that no spatial frequencies beyond the system's cutoff frequency are lost. However, to avoid oversampling, a more practical choice of d_z is often around $2R_z$.

The overlap of information between adjacent planes depends on the axial PSF. The PSF in the z-direction is often approximated as a Gaussian function:

$$\mathrm{PSF}(z) \propto \exp\left(-\frac{z^2}{2\sigma_z^2}\right), \tag{8.3}$$

where the axial spread parameter σ_z is related to the full-width at half-maximum (FWHM) by:

$$\sigma_z = \frac{R_z}{2\sqrt{2\ln 2}}. \tag{8.4}$$

The correlation $C(d_z)$ between two planes separated by d_z is given by the integral of the overlapping PSFs:

$$C(d_z) \propto \int_{-\infty}^{\infty} \mathrm{PSF}(z) \cdot \mathrm{PSF}(z - d_z)\, dz. \tag{8.5}$$

Substituting the Gaussian form of the PSF:

$$C(d_z) \propto \exp\left(-\frac{d_z^2}{4\sigma_z^2}\right). \tag{8.6}$$

This function decays exponentially as d_z increases, indicating that the correlation between planes diminishes as their separation grows.

8.1.1 Defining the Practical Inter-Planar Distance

From the correlation function, we define practical limits:

- If $d_z < 2R_z$, adjacent planes are highly correlated, leading to oversampling.
- If $d_z > 4R_z$, the correlation drops below 0.1, meaning that crucial spatial information is lost.

Theoretical analysis suggest that choosing d_z within the range:

$$2R_z \leq d_z \leq 4R_z \tag{8.7}$$

ensures a balance between preserving unique spatial information and avoiding redundant sampling. TranSIM leverages these principles to optimize plane separations for high-resolution 4D imaging. Therefore, with a measured axial PSF $\simeq 5\mu m$, I decided to choose a spatial separation of $d_z \simeq 20\mu m$.

8.1.2 Defining the Axial Sampling Rate

Using the aforementioned inter-planar distance considerations, the next question is determining the practical axial spatial sampling rate. The choice follows a fundamental principle: the sampling rate should be at or below the Nyquist criterion to accurately capture spatial information without aliasing artifacts.

Given that the axial resolution of the system is approximately $R_z \approx 5$ µm, the Nyquist sampling condition requires that the inter-planar spacing d_z satisfies:

$$d_z \leq \frac{R_z}{2}. \tag{8.8}$$

Substituting $R_z = 5$ µm, the optimal axial sampling rate is:

$$d_z = \frac{5}{2} = 2.5 \text{ µm}. \tag{8.9}$$

Thus, the practical axial spatial sampling rate is:

$$\frac{1}{d_z} = \frac{1}{2.5} = 0.4 \text{ planes per µm}. \tag{8.10}$$

By adhering to this sampling rate, TranSIM ensures that volumetric reconstructions contain sufficient spatial information while avoiding unnecessary oversampling, which could increase computational burden without significant gains in information. This balance optimizes both data acquisition efficiency and reconstruction fidelity.

8.2 Time-Independent Neural Network (TINN)

The first neural network I explored as a viable option for TranSIM's volumetric data generation is a 3D U-Net variant designed to process volumetric data, incorporating several modifications to enhance reconstruction accuracy, such as residual connections and self-attention. However, this network architecture assumes temporal independence between generated volumes, which can lead to inconsistencies when reconstructing dynamic time-series data.

A high-level overview of the network structure is presented in Figure 6.7. The architecture consists of a downsampling path that extracts hierarchical spatial features, a bottleneck with self-attention to capture long-range dependencies, and an upsampling path that restores spatial resolution while integrating multi-scale feature information. The network operates on sparse axial planes as input and reconstructs densely sampled volumetric stacks.

For a detailed breakdown of the network architecture, including the specific layer configurations, activation functions, and optimization strategies used, refer to the Materials and Methods section.

8.2.1 Validating Neural Network Reconstructions

In the previous section, I justified my choice of axial sampling rate based on the measured axial PSF. Here, I present the results of this choice. My initial selection was based on PSF measurements using a 40X objective, which exhibited minimal aliasing. However, the 16X objective shows some degree of aliasing or PSF broadening, likely due to optical aberrations specific to this objective.

I chose to increase the spatial sampling from the theoretical minimum of 0.4 $planes/\mu m$ to 0.8 $planes/\mu m$. This in turn further validates the ability of the network to infer spatial continuity. Later, I restrict the number of planes I reconstruct down to the theoretical minimum (albeit still more than required since the 16X PSF has been optically broadened) as this aligns with my philosophy of maximizing data efficiency.

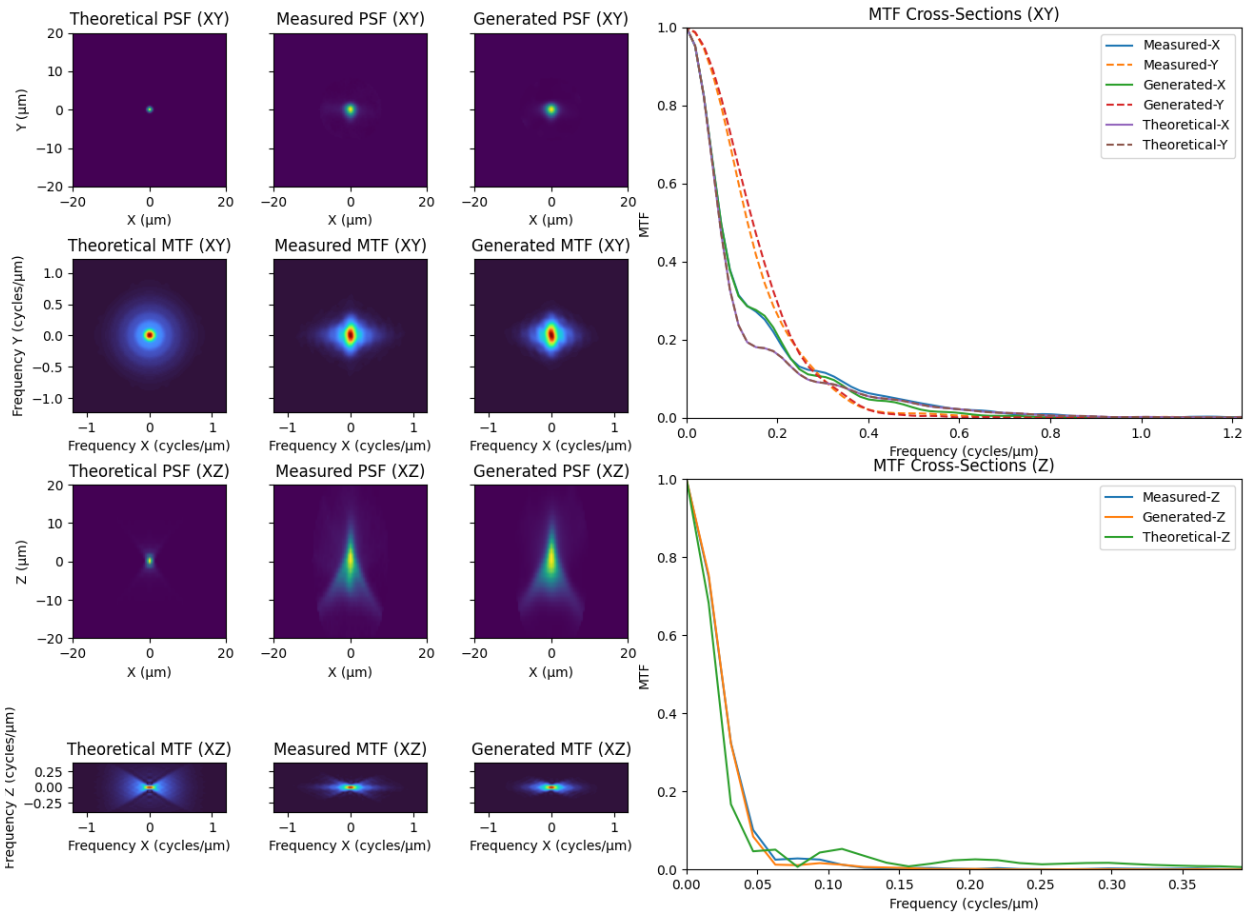

Figure 8.1: *Comparison of Real, Generated, and Theoretical PSFs and MTFs.* The first row displays the XY cross-sections of the point spread function (PSF) for the theoretical, real, and generated cases, respectively. The second row shows the corresponding modulation transfer function (MTF) representations, highlighting spatial frequency content in the XY plane. The third and fourth rows present the XZ cross-sections of the PSF and MTF, respectively, illustrating axial resolution differences. The final two plots display 1D cross-sections of the MTF in the XY and XZ planes, providing quantitative comparisons of frequency response along lateral and axial dimensions. These results validate the neural network's ability to approximate the optical system's response while revealing potential deviations from the theoretical and measured PSFs.

8.2.2 Choosing the Right Loss Function

I explored and evaluated the performance of several loss functions, including traditional ones such as MAE and MSE, as well as my own custom modifications, such as the Cosine weight adjustment as a function of depth. The rationale behind this weighted adjustment is based on the periodic sparsity of the imaging planes. Given this periodicity, the neural network may be better at inferring planes where more information is available—specifically, the sparse planes—rather than the intermediary ones. To assess this hypothesis, I evaluated the Cosine-MAE, MAE, and MSE loss functions for inter-plane reconstruction in TranSIM using the Structural Similarity Index Measure (SSIM) and Root Mean Squared Error (RMSE). The results of this evaluation are presented in Table 8.1.

Loss Function	**Avg SSIM (↑)**	**Avg RMSE (↓)**
MAE-COSINE	**0.6953**	**0.0503**
MAE	0.6898	0.0513
MSE	0.6602	0.0596

Table 8.1: Average SSIM and RMSE values for different loss functions used in TranSIM inter-plane reconstruction. Higher SSIM values indicate better structural preservation, while lower RMSE values indicate lower reconstruction error.

The results indicate that the MAE-COSINE loss function achieved the highest SSIM value of 0.6953, demonstrating superior structural preservation compared to standard MAE, which obtained an SSIM of 0.6898, and MSE, which achieved an SSIM of 0.6602. In terms of RMSE, MAE-COSINE also outperformed the other loss functions with the lowest value of 0.0503, compared to 0.0513 for MAE and 0.0596 for MSE. These results suggest that MAE-COSINE provides the most accurate intensity reconstruction while maintaining the integrity of fine structural details. A detailed plane-by-plane analysis of these metrics is presented in Figure 8.2.

The observed advantage of MAE-COSINE can be attributed to its ability to balance intensity accuracy, as captured by MAE, while prioritizing intermediary planes by increasing

their weight. This weighting strategy ensures that central planes contribute more significantly to the reconstruction, reducing inconsistencies and improving spatial coherence. In contrast, MSE underperformed in both SSIM and RMSE, likely due to its sensitivity to intensity outliers. The squared error term in MSE amplifies large deviations, which can lead to excessive blurring or misalignment in reconstructed planes. The relatively small difference in RMSE between MAE-COSINE and MAE suggests that absolute intensity differences alone do not fully capture the quality of the reconstruction. Instead, structure-aware loss functions like MAE-COSINE yield more meaningful results by preserving fine spatial details and mitigating artifacts.

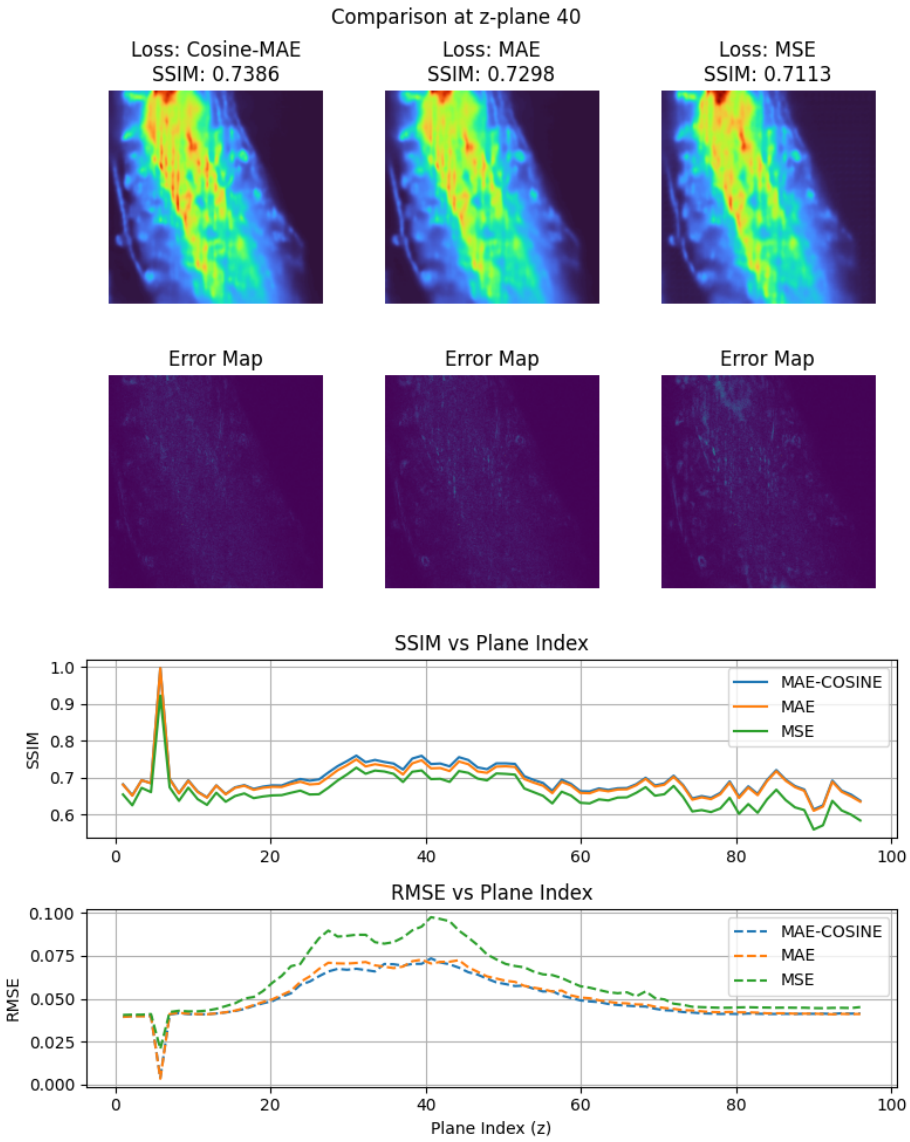

Figure 8.2: *Comparing MAE-Cosine to Standard Loss Functions* The top row presents the reconstructed planes using different loss functions: **MAE-Cosine, MAE, and MSE**. The second row displays the mean absolute difference maps between the ground truth and the reconstructed images, highlighting errors in the predictions. The two plots show the SSIM and RMSE values as a function of the plane index, demonstrating the structural similarity retention across depth and the RMSE values across the planes, respectively, illustrating the overall intensity error trends. The results indicate that Cosine-MAE achieves the best balance between preserving structural integrity and minimizing intensity errors.

8.2.3 Final Network Training

The TranSIM reconstruction network was trained using the Cosine-MAE loss function, which demonstrated optimal performance in preserving structural coherence and intensity accuracy. To further enhance perceptual quality, the Structural Similarity Index Measure (SSIM) was incorporated as a complementary term in the final loss function:

$$\mathcal{L}_{\text{total}} = \lambda \cdot \mathcal{L}_{\text{Cos,MAE}} + (1 - \lambda) \cdot \mathcal{L}_{\text{SSIM}}. \tag{8.11}$$

where the regularization term was set to $\lambda = 0.3$, prioritizing pixel-wise differences while still accounting for structural similarity.

The network was implemented using TensorFlow 2.10 and trained on Python 3.10. The Adam optimizer was utilized with a learning rate of 5×10^{-6} and all other default parameters. Training was conducted over 40 epochs with 10,000 iterations per epoch, using a batch size of 4. The input data consisted of frames with dimensions $(\text{batch}, 6, 256, 256)$, corresponding to 6 sparse axial planes per sample. This training configuration ensured stable optimization and robust convergence, allowing the network to effectively reconstruct volumetric fluorescence microscopy data with high fidelity.

8.2.4 Volumetric Reconstruction of Drosophila Larvae

After reconstructing fluorescent beads, I then trained the neural network on *Drosophila melanogaster* larvae data. Larvae were prepared using the methodology described in the biological sample preparation section. Training data volumes were acquired from fully immobilized larvae by axially scanning the imaging objective with 1.25 μm image plane spacing. From here, 10 total volumes were acquired to mimic temporal volumes. The training data was prepared by axially registering the planes with the pre-calculated affine transformations calculated from the 4 μm bead sheet, as described in the mathematical framework for inter-planar alignment section of the materials and methods. All training parameters remained the same as those mentioned in the previous subsection.

The results indicate that the network effectively reconstructs volumetric information while preserving structural integrity, despite the inherent biological complexity of the sample. Notably, the network demonstrates strong performance in areas of high fluorescence contrast, while minor deviations in SSIM maps suggest potential improvements in regions with low signal-to-noise ratios. Across selected planes, the reconstructed volumes achieved SSIM values of 0.682, 0.696, 0.723, 0.705, and 0.637, demonstrating strong structural preservation, particularly in central axial planes. The RMSE values for these corresponding planes were 0.039, 0.058, 0.062, 0.043, and 0.041, indicating low intensity reconstruction errors.

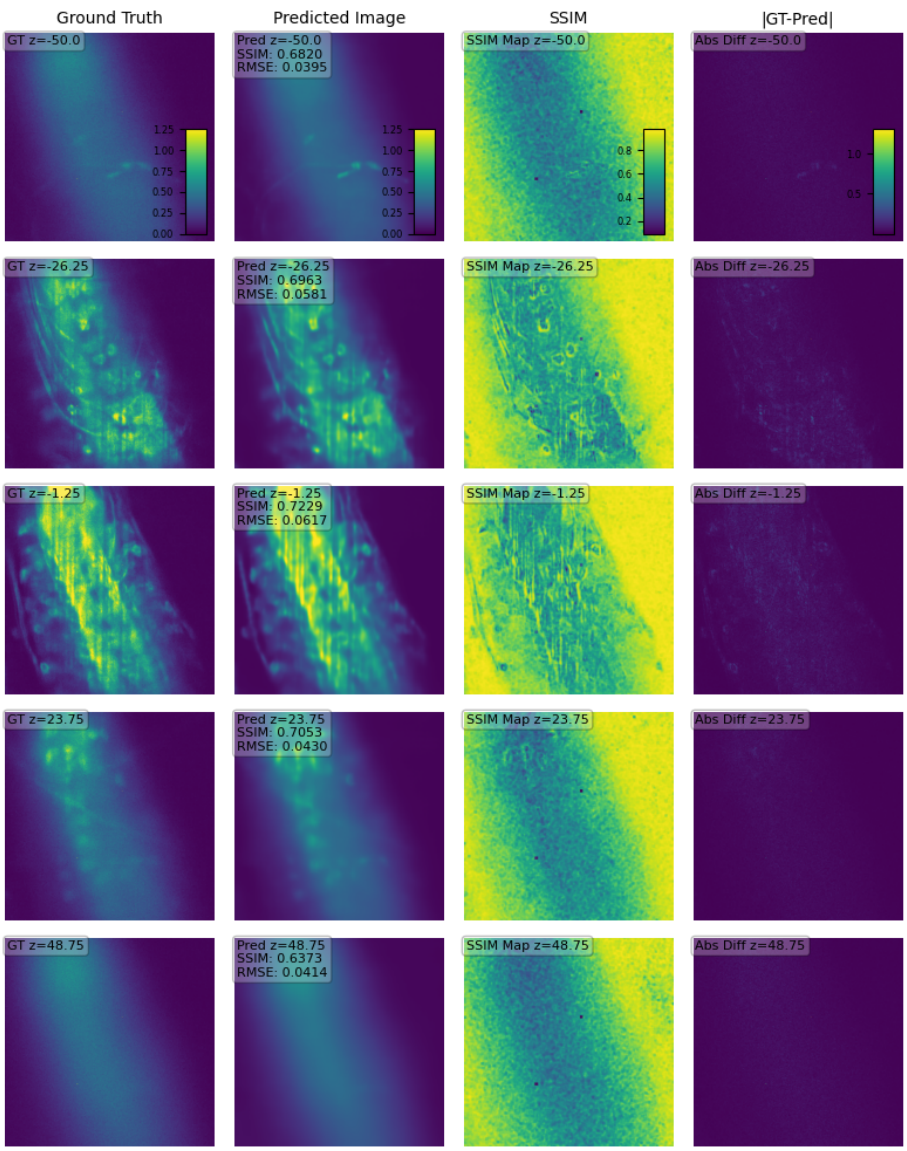

Figure 8.3: *Comparison of Ground Truth and Reconstructed Planes.* Each row corresponds to a different selected axial plane, while the columns present: (1) the original ground truth image, (2) the reconstructed plane using the neural network, (3) the Structural Similarity Index (SSIM) map highlighting structural deviations, and (4) the absolute difference between the ground truth and reconstructed image. The SSIM and RMSE values are displayed for each reconstructed plane, indicating the fidelity of the neural network's predictions. These comparisons illustrate how well the neural network preserves structural information and intensity variations across different depths.

8.2.5 Voxel Throughput Analysis

To assess the computational efficiency of the neural network when applied to volumetric reconstructions of *Drosophila* larvae, I evaluated the voxel reconstruction rate as a function of image area. The voxel throughput metric, defined as the number of reconstructed voxels per second, provides insight into the network's scalability and processing speed for varying image resolutions.

The results, summarized in Figure 8.4, demonstrate a positive correlation between image area and voxel reconstruction speed. As the image area increases, the voxel throughput also increases. Specifically, for an volume size of 48×128^2 pixels, the network reconstructs 6.61×10^6 voxels per second, whereas for the largest tested volume size of 48×640^2 pixels, the reconstruction rate reaches 1.34×10^7 voxels per second. This trend suggests that the network efficiently utilizes larger image sizes to process more voxels per second, likely benefiting from optimized memory access patterns and parallel computation.

This scalability is particularly advantageous for throughput fluorescence microscopy, as it ensures that increasing imaging resolution does not significantly compromise computational performance. The observed increase in voxel throughput reinforces the suitability of the proposed neural network for large-scale volumetric imaging applications.

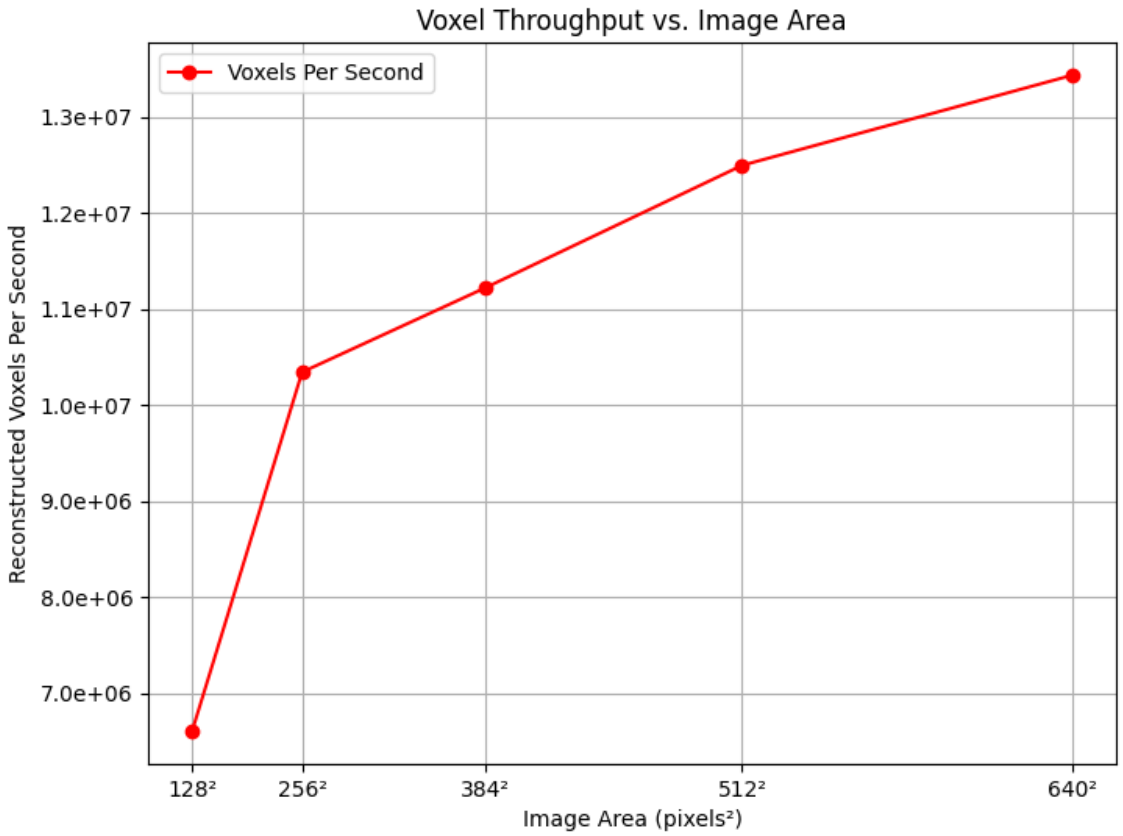

Figure 8.4: *Voxel Throughput vs. Image Area.* The plot illustrates the reconstructed voxels per second as a function of image resolution, with input sizes ranging from 128^2 to 640^2 pixels. The results indicate a positive correlation between image area and voxel throughput, suggesting that larger images allow for more efficient volumetric reconstruction. This trend highlights the scalability of the neural network and its suitability for handling high-resolution fluorescence microscopy data.

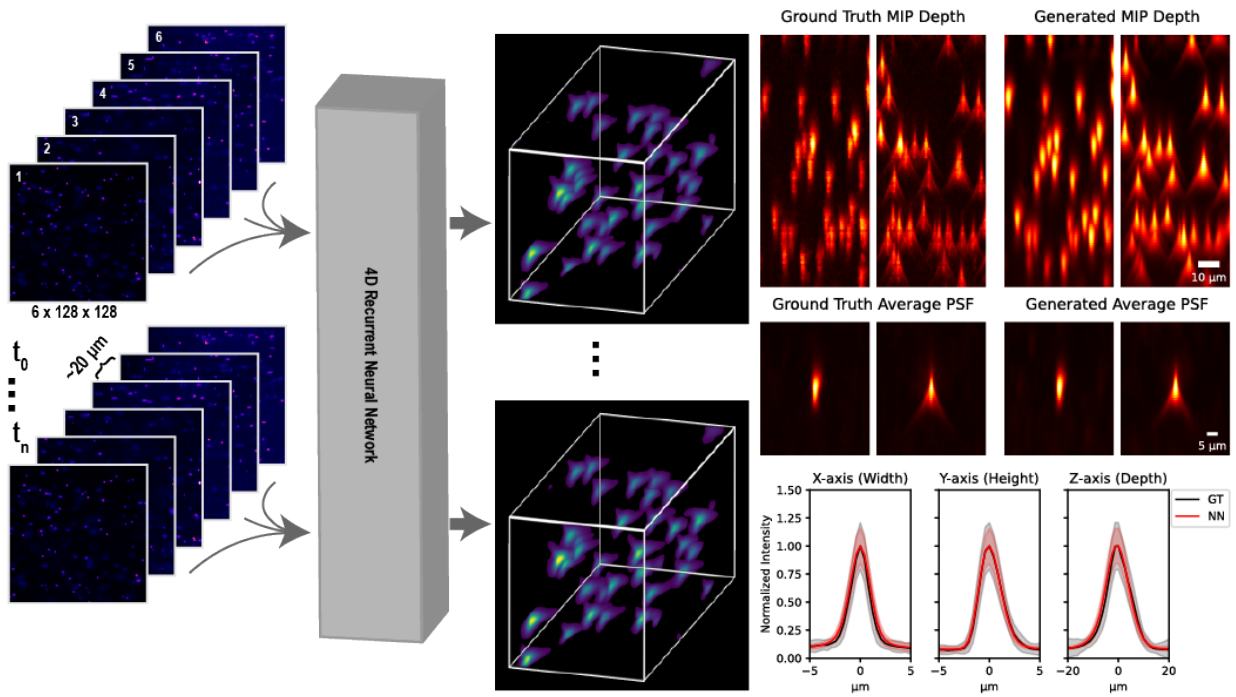

Figure 8.5: *Time-Dependent Neural Network.* To investigate the temporal discontinuity in volumetric reconstruction, a LSTM based RNN was implemented. 4D chunks of data (time sequence + 3D volume) were fed into the neural network. The neural network, returned time correlated reconstructions of the same shape. PSF analysis shows similar results to the time-independent neural network.

8.3 Time-Dependent Neural Network (TDNN)

Given that the previous neural network does not have any time dependence between the reconstructed volumes, I decided to add temporal connections between the reconstructed volumes, in particular, by making of use Recurrent Neural Networks. Specifically, I decided to implement Long-Short Term Memory latent space connections between the reconstructed volumes before bringing them up the decoder pathway. Building upon, the time-independent neural network where I justified the use of Cosine-MAE coupled with SSIM, I utilized this same loss function. For specific neural network architecture, refer to the materials and methods section, subsection time-dependent neural network.

8.3.1 Final Network Training

The network was implemented using TensorFlow 2.10 and trained on Python 3.10. The Adam optimizer was utilized with a learning rate of 5×10^{-6} and all other default parameters and with the time-independent neural network for further comparison analysis. Training was conducted over 40 epochs with 10,000 iterations per epoch, using a batch size of 1, however, unlike the time-independent network, here I had the network reconstruct 5 time steps. The input data consisted of frames with dimensions $(\text{batch}, 5, 6, 128, 128)$, corresponding to 6 sparse axial planes per sample and 5 time steps. This training configuration ensured stable optimization and robust convergence, allowing the network to effectively reconstruct volumetric fluorescence microscopy data with high fidelity. Here the only, real noticeable difference was in the height and width dimensions of the input batch training datasets. Here, because of the larger neural network architecture, the size of the images have to be reduced by a factor of two per dimension.

8.3.2 Volumetric Reconstruction

The results in Figure 8.6 present a detailed comparison between the generated and ground truth data across 96 depth planes, quantified using the SSIM, RMSE, and MAE, in accordance to the analysis used for the previous network.

The SSIM values range from approximately 0.59 to 0.76, indicating moderate-to-high structural fidelity in the reconstructions. Across all time points, mid-range depth planes consistently achieve higher SSIM values, peaking at approximately 0.75–0.76. This suggests that the neural network is most effective at reconstructing planes with more contextual information from surrounding slices. Shallower and deeper planes tend to have lower SSIM values, typically ranging from 0.59 to 0.64, indicating greater deviations in structural accuracy at the boundaries. This behavior is likely due to edge effects, where fewer adjacent planes contribute to the reconstruction. Moreover, this deviation is further enhanced by the fact that the sample is rather thin and does not occupy the entirety of volumetric depth. The highest SSIM values appear between depth planes 40–60 across most time points, supporting

the idea that reconstruction accuracy is best where the input and interpolated planes have the most balanced contributions.

The RMSE values range from approximately 0.024 to 0.08, with lower values indicating better pixel-wise accuracy. RMSE shows a near-inverse relationship with SSIM, where lower RMSE values, between 0.024 and 0.027, correspond to higher SSIM values, reinforcing the observation that better structural preservation aligns with improved intensity fidelity. The highest RMSE values, reaching approximately 0.08, are observed in mid-depth planes (40–50), despite the high SSIM, suggesting that while these planes maintain structural coherence, they still exhibit some pixel-wise intensity mismatches. Similar to SSIM, edge planes, particularly those below depth 10 or above depth 85, tend to have lower RMSE values. However, this could be attributed to lower structural complexity in these regions rather than an actual improvement in reconstruction accuracy.

Overall, the results indicate that middle planes reconstruct most effectively, with both SSIM and RMSE trends showing optimal reconstruction in the depth range of approximately 40–60. This suggests that interpolated planes in this region are the most reliable. Edge planes exhibit greater deviation, with lower SSIM and slightly reduced RMSE values, suggesting that these regions lack sufficient contextual information for accurate reconstruction. Despite minor fluctuations, the trends in SSIM and RMSE remain consistent across all five time points, indicating that the neural network's performance is stable over time.

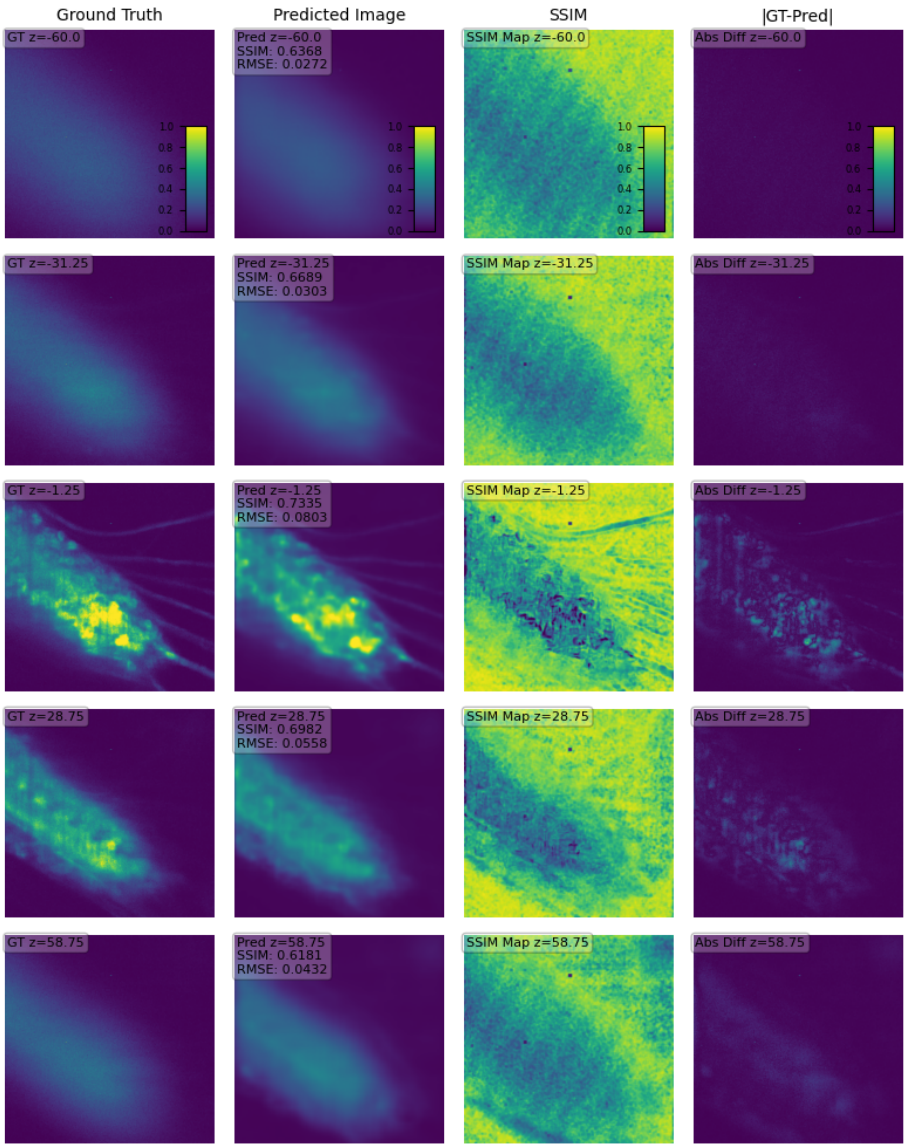

Figure 8.6: *Depth Reconstruction Analysis.* Each row corresponds to a different selected axial plane, while the columns present: (1) the original ground truth image, (2) the reconstructed plane using the neural network, (3) the Structural Similarity Index (SSIM) map highlighting structural deviations, and (4) the absolute difference between the ground truth and reconstructed image. The SSIM and RMSE values are displayed for each reconstructed plane, indicating the fidelity of the neural network's predictions. These comparisons illustrate how well the neural network preserves structural information and intensity variations across different depths.

8.4 The Noise Discrepancy

The analysis of TDNN and TINN models against both denoised and noisy ground truth data provides insights into how noise influences reconstruction performance. The SSIM and RMSE were computed for each case, and the depth-averaged results are presented in Figure 8.7.

The results indicate that TDNN performs better when compared to the denoised ground truth, achieving an average SSIM of 0.8503 and an average RMSE of 0.0386. When compared to the noisy ground truth, the SSIM drops to 0.6456, and RMSE increases to 0.0487. Similarly, TINN achieves the highest performance when compared to the denoised ground truth, with an average SSIM of 0.9248 and an RMSE of 0.0222. However, its performance slightly declines when compared to the noisy ground truth, with SSIM at 0.6992 and RMSE at 0.0346.

Since the acquired data for training was collected under real-world acquisition conditions, it inherently contained noise. The cameras were operated at their standard frame rate of 100 FPS, ensuring consistency with non-neural network data acquisition. To evaluate the impact of noise on reconstruction, an additional study was performed using DeepCAD-RT, a neural network-based denoising algorithm, to preprocess the ground truth before comparison.

As shown in Figure 8.7, the results confirm that both TDNN and TINN improve their performance when evaluated against denoised ground truth data. This suggests that much of the error in reconstruction arises from the networks' implicit denoising capabilities. Notably, TINN consistently outperforms TDNN in all cases, achieving higher SSIM and lower RMSE values, indicating its superior ability to preserve fine structural details while mitigating noise-related artifacts.

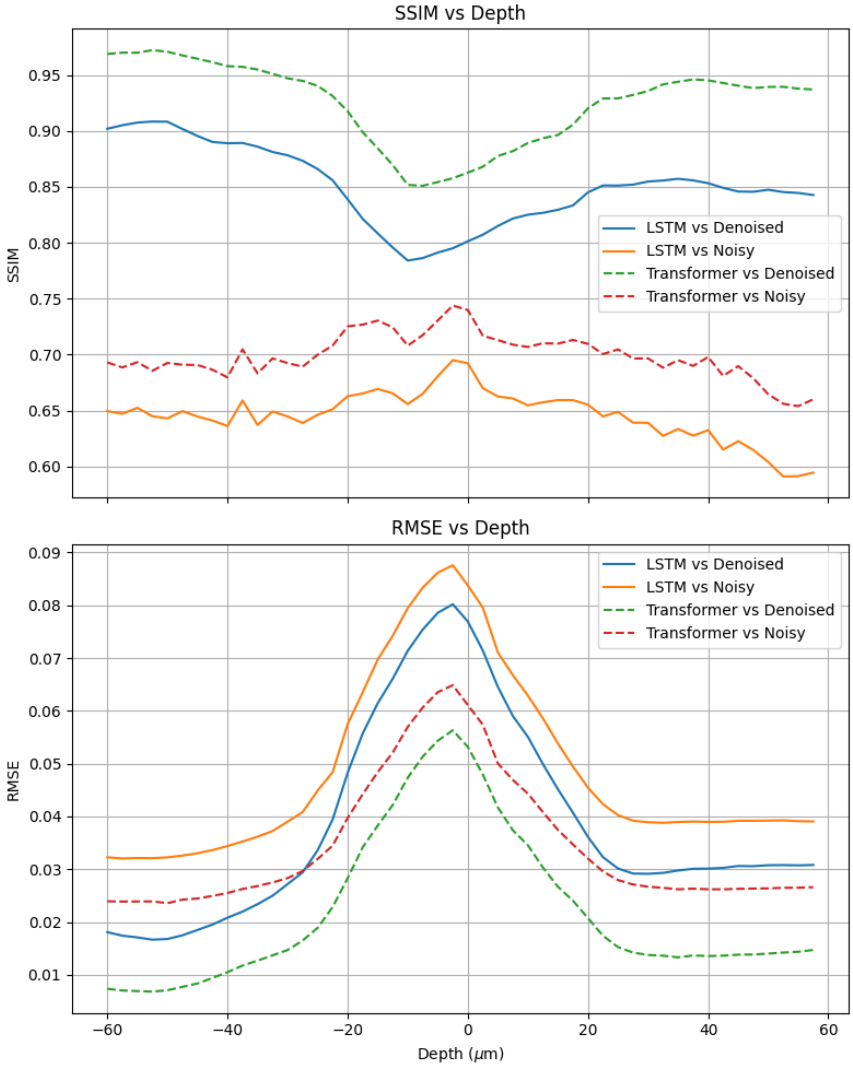

Figure 8.7: *Neural Networks Evaluated against Noisy and Denoised Ground Truth Data.* The plots illustrate how reconstruction accuracy varies across depth planes, highlighting the impact of noise on model performance. Higher SSIM values and lower RMSE values indicate better structural preservation and lower reconstruction error, respectively. The results demonstrate that both TDNN and TINN achieve higher SSIM and lower RMSE when compared to denoised ground truth, suggesting that a significant portion of the reconstruction error arises from the networks' implicit denoising processes. Additionally, TINN consistently outperforms TDNN across all depths, reinforcing its superior ability to maintain structural integrity while minimizing noise-induced artifacts.

CHAPTER 9

Discussion and Conclusion

The results presented in this work demonstrate that TranSIM represents a significant advancement in volumetric imaging, addressing longstanding limitations in traditional microscopy systems. Transverse-Sheet Illumination Microscopy provides a framework for rapid and parallel data acquisition. Typically, fluorescent signals in microscopy are captured with a single sensor, limiting the total bandwidth of the system. Here, I have demonstrated that it is possible to spatially parallelize data acquisition on the detection side while simultaneously designing an efficient illumination scheme capable of volumetric imaging. Furthermore, I have shown that TranSIM operates at or near diffraction-limited resolutions, indicating that there is no significant loss of spatial information, as is often the case with fast volumetric scanners. The ability to maintain high spatial fidelity while achieving large-scale volumetric imaging highlights TranSIM's potential for expanding fast 4D microscopy.

Beyond demonstrating the feasibility of parallelized sensors, I also explored the technical scalability of TranSIM. The system was configured to spatially separate the focal planes and translate them into focus across multiple sensors. By utilizing Hamamatsu Flash 4.0 V2 sensors, which feature a maximum rolling shutter speed of 100 kHz at a 2048-pixel width, I have shown that TranSIM is already operating at high bandwidths. However, the system still holds potential for further optimization with newer-generation sensors such as the Hamamatsu ORCA-Quest 2 qCMOS, which offers a 120 kHz line rate at 4096 pixels and a much quieter noise floor. This upgrade alone could increase the overall bandwidth by a factor of 2.4, raising the 0.38 GHz pixel rate to 0.912 GHz. Additionally, since TranSIM inherently de-scans the image planes before entering the realignment cycle, there is the possibility of integrating line sensors instead of two-dimensional sensors. This alternative

could significantly reduce costs and system complexity—an avenue I plan to explore in future iterations.

In addition to its standalone capabilities, TranSIM's design can be coupled with other imaging modalities, particularly two-photon fluorescence microscopy techniques such as Light Bead Microscopy and Reverberation 2P Microscopy [**?**]. These techniques are already optimized for temporal multiplexing, making them strong candidates for integration with TranSIM's multi-sensor detection approach. By combining these strategies, I anticipate that TranSIM can help overcome existing bandwidth limitations and further push the boundaries of high-speed volumetric imaging. More broadly, I expect TranSIM to pave the way toward parallelized data acquisition on the detection side, providing rapid, high-fidelity volumetric data that enables new investigations into neural connectivity and spatiotemporal dynamics.

While TranSIM significantly improves volumetric imaging on the hardware level, it can and was further enhanced by computational reconstruction techniques to achieve full volumetric continuity. In this work, I investigated deep learning-based approaches to interpolate the sparse imaging planes. I explored both time-independent and time-dependent neural network architectures to assess their efficacy in reconstructing volumetric fluorescence data. First, I implemented three-dimensional convolutional neural networks enhanced with self-attention mechanisms. These networks successfully reconstructed static volumetric data by capturing spatial dependencies across imaging planes. However, when applied to dynamic processes, this type of network exhibited limitations in maintaining temporal coherence. This prompted an exploration of four-dimensional recurrent neural networks, which integrate temporal dependencies alongside spatial information. These models improved temporal stability in volumetric reconstructions, demonstrating enhanced preservation of neural activity patterns or physiologically dynamic processes.

The results presented in this work firmly establish TranSIM as a groundbreaking advancement in high-speed volumetric imaging. The system successfully overcomes the bandwidth and spatial limitations of traditional microscopy by leveraging parallelized detection and neural network-assisted reconstructions. TranSIM consistently delivers high-fidelity volumetric data at unprecedented speeds while maintaining diffraction-limited resolution, making it an

unparalleled tool for capturing dynamic biological processes.

Neural network-assisted reconstructions further enhance the system's capabilities, enabling seamless volumetric interpolation from sparsely acquired image planes. Both 3D-CNNs and 4D-RNNs demonstrated excellent performance, producing high-quality reconstructions across multiple biological models, including larval *Drosophila* and zebrafish larvae. While further refinements in neural network selection and data processing scalability will optimize these results, the current performance already shows the power of combining TranSIM's hardware framework with cutting-edge computational techniques. For example, one minute of raw imaging data generates approximately 0.7 GB/s, while neural network-interpolated data expands this range to 5.1–10.2 GB/s, depending on the specific reconstruction parameters. This shows that I achieved a 10-fold increase in my already parallelized data acquisition technique, all while maintaining spatial sampling awareness.

TranSIM's ability to image large-scale volumes at 100–200 Hz with exceptional spatial fidelity opens new frontiers in neuroscience, physiology, and beyond. Its adaptability and scalability ensure that it will remain at the forefront of high-speed imaging innovations. As computational techniques advance and hardware constraints are further mitigated, TranSIM is poised to redefine ULTRA-FAST volumetric microscopy, unlocking deeper insights into the complexities of neural activity and biological dynamics.

REFERENCES

[1] Misha B. Ahrens, Michael B. Orger, Drew N. Robson, Jennifer M. Li, and Philipp J. Keller. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature Methods*, 10(5):413–420, May 2013. Number: 5 Publisher: Nature Publishing Group.

[2] Venkatakaushik Voleti, Kripa B. Patel, Wenze Li, Citlali Perez Campos, Srinidhi Bharadwaj, Hang Yu, Caitlin Ford, Malte J. Casper, Richard Wenwei Yan, Wenxuan Liang, Chentao Wen, Koutarou D. Kimura, Kimara L. Targoff, and Elizabeth M. C. Hillman. Real-time volumetric microscopy of in vivo dynamics and large-scale samples with SCAPE 2.0. *Nature Methods*, 16(10):1054–1062, October 2019. Number: 10 Publisher: Nature Publishing Group.

[3] Bi-Chang Chen, Wesley R. Legant, Kai Wang, Lin Shao, Daniel E. Milkie, Michael W. Davidson, Chris Janetopoulos, Xufeng S. Wu, John A. Hammer, Zhe Liu, Brian P. English, Yuko Mimori-Kiyosue, Daniel P. Romero, Alex T. Ritter, Jennifer Lippincott-Schwartz, Lillian Fritz-Laylin, R. Dyche Mullins, Diana M. Mitchell, Joshua N. Bembenek, Anne-Cecile Reymann, Ralph Böhme, Stephan W. Grill, Jennifer T. Wang, Geraldine Seydoux, U. Serdar Tulu, Daniel P. Kiehart, and Eric Betzig. Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science*, 346(6208), October 2014. Publisher: American Association for the Advancement of Science Section: Research Article.

[4] Blake Madruga. *Design and Construction of a Rapidly-Reconfigurable SLM-Based Light Sheet Microscope*. PhD thesis, UCLA, 2018.

[5] Katsushi Arisaka and Laurent Bentolila. Trans-Sheet Illumination Microscopy (TranSIM) for decoding whole brain activity at submillisecond temporal resolution.

[6] Owen Randlett, Caroline L. Wee, Eva A. Naumann, Onyeka Nnaemeka, David Schoppik, James E. Fitzgerald, Ruben Portugues, Alix M. B. Lacoste, Clemens Riegler, Florian Engert, and Alexander F. Schier. Whole-brain activity mapping onto a zebrafish brain atlas. *Nature Methods*, 12(11):1039–1046, November 2015. Number: 11 Publisher: Nature Publishing Group.

[7] Rainer Friedrich, Christel Genoud, and Adrian A. Wanner. Analyzing the structure and function of neuronal circuits in zebrafish. *Frontiers in Neural Circuits*, 7, 2013. Publisher: Frontiers.

[8] Sven Dorkenwald, Arie Matsliah, Amy R. Sterling, Philipp Schlegel, Szi-chieh Yu, Claire E. McKellar, Albert Lin, Marta Costa, Katharina Eichler, Yijie Yin, Will Silversmith, Casey Schneider-Mizell, Chris S. Jordan, Derrick Brittain, Akhilesh Halageri, Kai Kuehner, Oluwaseun Ogedengbe, Ryan Morey, Jay Gager, Krzysztof Kruk, Eric Perlman, Runzhe Yang, David Deutsch, Doug Bland, Marissa Sorek, Ran Lu, Thomas Macrina, Kisuk Lee, J. Alexander Bae, Shang Mu, Barak Nehoran, Eric Mitchell, Sergiy

Popovych, Jingpeng Wu, Zhen Jia, Manuel A. Castro, Nico Kemnitz, Dodam Ih, Alexander Shakeel Bates, Nils Eckstein, Jan Funke, Forrest Collman, Davi D. Bock, Gregory S. X. E. Jefferis, H. Sebastian Seung, and Mala Murthy. Neuronal wiring diagram of an adult brain. *Nature*, 634(8032):124–138, October 2024. Publisher: Nature Publishing Group.

[9] Hanchuan Peng, Phuong Chung, Fuhui Long, Lei Qu, Arnim Jenett, Andrew M. Seeds, Eugene W. Myers, and Julie H. Simpson. BrainAligner: 3D Registration Atlases of Drosophila Brains. *Nature methods*, 8(6):493–500, June 2011.

[10] Louis K Scheffer, C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth J Hayworth, Gary B Huang, Kazunori Shinomiya, Jeremy Maitlin-Shepard, Stuart Berg, Jody Clements, Philip M Hubbard, William T Katz, Lowell Umayam, Ting Zhao, David Ackerman, Tim Blakely, John Bogovic, Tom Dolafi, Dagmar Kainmueller, Takashi Kawase, Khaled A Khairy, Laramie Leavitt, Peter H Li, Larry Lindsey, Nicole Neubarth, Donald J Olbris, Hideo Otsuna, Eric T Trautman, Masayoshi Ito, Alexander S Bates, Jens Goldammer, Tanya Wolff, Robert Svirskas, Philipp Schlegel, Erika Neace, Christopher J Knecht, Chelsea X Alvarado, Dennis A Bailey, Samantha Ballinger, Jolanta A Borycz, Brandon S Canino, Natasha Cheatham, Michael Cook, Marisa Dreher, Octave Duclos, Bryon Eubanks, Kelli Fairbanks, Samantha Finley, Nora Forknall, Audrey Francis, Gary Patrick Hopkins, Emily M Joyce, SungJin Kim, Nicole A Kirk, Julie Kovalyak, Shirley A Lauchie, Alanna Lohff, Charli Maldonado, Emily A Manley, Sari McLin, Caroline Mooney, Miatta Ndama, Omotara Ogundeyi, Nneoma Okeoma, Christopher Ordish, Nicholas Padilla, Christopher M Patrick, Tyler Paterson, Elliott E Phillips, Emily M Phillips, Neha Rampally, Caitlin Ribeiro, Madelaine K Robertson, Jon Thomson Rymer, Sean M Ryan, Megan Sammons, Anne K Scott, Ashley L Scott, Aya Shinomiya, Claire Smith, Kelsey Smith, Natalie L Smith, Margaret A Sobeski, Alia Suleiman, Jackie Swift, Satoko Takemura, Iris Talebi, Dorota Tarnogorska, Emily Tenshaw, Temour Tokhi, John J Walsh, Tansy Yang, Jane Anne Horne, Feng Li, Ruchi Parekh, Patricia K Rivlin, Vivek Jayaraman, Marta Costa, Gregory SXE Jefferis, Kei Ito, Stephan Saalfeld, Reed George, Ian A Meinertzhagen, Gerald M Rubin, Harald F Hess, Viren Jain, and Stephen M Plaza. A connectome and analysis of the adult Drosophila central brain. *eLife*, 9:e57443, September 2020. Publisher: eLife Sciences Publications, Ltd.

[11] Zhihao Zheng. A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster. 2018.

[12] Javier Carmona. *Transverse-Sheet Illumination Microscopy.* PhD thesis, UCLA, 2021.

[13] Yichen Wu, Yair Rivenson, Hongda Wang, Yilin Luo, Eyal Ben-David, Laurent A. Bentolila, Christian Pritz, and Aydogan Ozcan. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nature Methods*, 16(12):1323–1331, December 2019. Number: 12 Publisher: Nature Publishing Group.

[14] Luzhe Huang, Hanlong Chen, Yilin Luo, Yair Rivenson, and Aydogan Ozcan. Recurrent neural network-based volumetric fluorescence microscopy. *Light: Science &*

Applications, 10(1):62, March 2021. Publisher: Nature Publishing Group.

[15] Xilin Yang, Luzhe Huang, Yilin Luo, Yichen Wu, Hongda Wang, Yair Rivenson, and Aydogan Ozcan. Deep-Learning-Based Virtual Refocusing of Images Using an Engineered Point-Spread Function. *ACS Photonics*, 8(7):2174–2182, July 2021. Publisher: American Chemical Society.

[16] George Gabriel Stokes. XXX. On the change of refrangibility of light. *Philosophical Transactions of the Royal Society of London*, 142:463–562, January 1852. Publisher: Royal Society.

[17] *Phenomena of Optical Metamaterials*. Elsevier, 2019.

[18] Tatjana Gric and Ortwin Hess. Chapter 8 - Active Optical Metamaterials. In Tatjana Gric and Ortwin Hess, editors, *Phenomena of Optical Metamaterials*, Micro and Nano Technologies, pages 187–261. Elsevier, January 2019.

[19] Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea - Shimomura - 1962 - Journal of Cellular and Comparative Physiology - Wiley Online Library.

[20] Atsushi Miyawaki, Juan Llopis, Roger Heim, J. Michael McCaffery, Joseph A. Adams, Mitsuhiko Ikura, and Roger Y. Tsien. Fluorescent indicators for Ca 2+ based on green fluorescent proteins and calmodulin. *Nature*, 388(6645):882–887, August 1997. Number: 6645 Publisher: Nature Publishing Group.

[21] Tsai-Wen Chen, Trevor J. Wardill, Yi Sun, Stefan R. Pulver, Sabine L. Renninger, Amy Baohan, Eric R. Schreiter, Rex A. Kerr, Michael B. Orger, Vivek Jayaraman, Loren L. Looger, Karel Svoboda, and Douglas S. Kim. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300, July 2013. Publisher: Nature Publishing Group.

[22] Junichi Nakai, Masamichi Ohkura, and Keiji Imoto. A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. *Nature Biotechnology*, 19(2):137–141, February 2001. Publisher: Nature Publishing Group.

[23] Hod Dana. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. 2019.

[24] Vincent Daria. Spatio-temporal parameters for optical probing of neuronal activity. 2021.

[25] Principles of Optics | ScienceDirect.

[26] Peter M. Eimon, Mostafa Ghannad-Rezaie, Gianluca De Rienzo, Amin Allalou, Yuelong Wu, Mu Gao, Ambrish Roy, Jeffrey Skolnick, and Mehmet Fatih Yanik. Brain activity patterns in high-throughput electrophysiology screen predict both drug efficacies and side effects. *Nature Communications*, 9(1):219, January 2018. Number: 1 Publisher: Nature Publishing Group.

[27] Bonifasius Putera Sampurna, Gilbert Audira, Stevhen Juniardi, Yu-Heng Lai, and Chung-Der Hsiao. A Simple ImageJ-Based Method to Measure Cardiac Rhythm in Zebrafish Embryos. *Inventions*, 3(2):21, June 2018. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[28] Minsky Marvin. Microscopy apparatus, December 1961.

[29] Mojmír Petráň, Milan Hadravský, M. David Egger, and Robert Galambos. Tandem-Scanning Reflected-Light Microscope*. *JOSA*, 58(5):661–664, May 1968. Publisher: Optical Society of America.

[30] High Speed | Yokogawa America.

[31] Kang-Bin Im, Sumin Han, Hwajoon Park, Dongsun Kim, and Beop-Min Kim. Simple high-speed confocal line-scanning microscope. *Optics Express*, 13(13):5151–5156, June 2005. Publisher: Optical Society of America.

[32] Jinwoo Lee, Yukihiro Miyanaga, Masahiro Ueda, and Sungchul Hohng. Video-Rate Confocal Microscopy for Single-Molecule Imaging in Live Cells and Superresolution Fluorescence Imaging. *Biophysical Journal*, 103(8):1691–1697, October 2012.

[33] ORCA-Fusion Digital CMOS camera C14440-20UP.

[34] ORCA-Lightning Digital CMOS camera C14120-20P.

[35] Tom Vettenburg, Heather I. C. Dalgarno, Jonathan Nylk, Clara Coll-Lladó, David E. K. Ferrier, Tomáš Čižmár, Frank J. Gunn-Moore, and Kishan Dholakia. Light-sheet microscopy using an Airy beam. *Nature Methods*, 11(5):541–544, May 2014. Number: 5 Publisher: Nature Publishing Group.

[36] Eric Betzig. Excitation strategies for optical lattice microscopy. *Optics Express*, 13(8):3021–3036, April 2005. Publisher: Optical Society of America.

[37] Marc Levoy Levoy, Ren Ng, Andrew Adams, Mathew Footer, and Mark Horowitz. Light field microscopy | ACM SIGGRAPH 2006 Papers.

[38] M. Levoy, Z. Zhang, and I. Mcdowall. Recording and controlling the 4D light field in a microscope using microlens arrays. *Journal of Microscopy*, 235(2):144–162, 2009. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2818.2009.03195.x.

[39] Robert Prevedel, Young-Gyu Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrödel, Ramesh Raskar, Manuel Zimmer, Edward S. Boyden, and Alipasha Vaziri. Simultaneous whole-animal 3D-imaging of neuronal activity using light-field microscopy. *Nature methods*, 11(7):727–730, July 2014.

[40] Robert Prevedel, Aart J. Verhoef, Alejandro J. Pernía-Andrade, Siegfried Weisenburger, Ben S. Huang, Tobias Nöbauer, Alma Fernández, Jeroen E. Delcour, Peyman Golshani, Andrius Baltuska, and Alipasha Vaziri. Fast volumetric calcium imaging across multiple cortical layers using sculpted light. *Nature Methods*, 13(12):1021–1028, December 2016. Publisher: Nature Publishing Group.

[41] Nicolas C. Pégard, Hsiou-Yuan Liu, Nick Antipa, Maximillian Gerlock, Hillel Adesnik, and Laura Waller. Compressive light-field microscopy for 3D neural activity recording. *Optica*, 3(5):517–524, May 2016. Publisher: Optica Publishing Group.

[42] Zhi Lu, Jiamin Wu, Hui Qiao, You Zhou, Tao Yan, Zijing Zhou, Xu Zhang, Jingtao Fan, and Qionghai Dai. Phase-space deconvolution for light field microscopy. *Optics Express*, 27(13):18131–18145, June 2019. Publisher: Optica Publishing Group.

[43] Changliang Guo, Wenhao Liu, Xuanwen Hua, Haoyu Li, and Shu Jia. Fourier light-field microscopy. *Optics Express*, 27(18):25573–25594, September 2019. Publisher: Optica Publishing Group.

[44] Lu Bai, Lin Cong, Ziqi Shi, Yuchen Zhao, Yujie Zhang, Bin Lu, Jing Zhang, Zhi-Qi Xiong, Ninglong Xu, Yu Mu, and Kai Wang. Volumetric voltage imaging of neuronal populations in the mouse brain by confocal light-field microscopy. *Nature Methods*, 21(11):2160–2170, November 2024. Publisher: Nature Publishing Group.

[45] C. Dunsby. Optically sectioned imaging by oblique plane microscopy. *Optics Express*, 16(25):20306–20316, December 2008. Publisher: Optica Publishing Group.

[46] Lei Zhang, Amalia Capilla, Weiye Song, Gustavo Mostoslavsky, and Ji Yi. Oblique scanning laser microscopy for simultaneously volumetric structural and molecular imaging using only one raster scan. *Scientific Reports*, 7(1):8591, August 2017. Publisher: Nature Publishing Group.

[47] Jeongmin Kim. Recent advances in oblique plane microscopy. *Nanophotonics*, 12(13):2317–2334, June 2023. Publisher: De Gruyter.

[48] Matthew B. Bouchard, Venkatakaushik Voleti, César S. Mendes, Clay Lacefield, Wesley B. Grueber, Richard S. Mann, Randy M. Bruno, and Elizabeth M. C. Hillman. Swept confocally-aligned planar excitation (SCAPE) microscopy for high-speed volumetric imaging of behaving organisms. *Nature Photonics*, 9(2):113–119, February 2015. Number: 2 Publisher: Nature Publishing Group.

[49] Hyoungjun Park, Myeongsu Na, Bumju Kim, Soohyun Park, Ki Hean Kim, Sunghoe Chang, and Jong Chul Ye. Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy. *Nature Communications*, 13(1):3297, June 2022. Publisher: Nature Publishing Group.

[50] Wei Ouyang, Andrey Aristov, Mickaël Lelek, Xian Hao, and Christophe Zimmer. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology*, 36(5):460–468, May 2018. Publisher: Nature Publishing Group.

[51] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2Void - Learning Denoising from Single Noisy Images, April 2019. arXiv:1811.10980 [cs].

[52] Joshua Batson and Loic Royer. Noise2Self: Blind Denoising by Self-Supervision, June 2019. arXiv:1901.11365 [cs].

[53] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, January 2021. Publisher: Nature Publishing Group.

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

[55] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Mauricio Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug, and Eugene W. Myers. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, 15(12):1090–1097, December 2018. Publisher: Nature Publishing Group.

[56] Xinyang Li, Yixin Li, Yiliang Zhou, Jiamin Wu, Zhifeng Zhao, Jiaqi Fan, Fei Deng, Zhaofa Wu, Guihua Xiao, Jing He, Yuanlong Zhang, Guoxun Zhang, Xiaowan Hu, Xingye Chen, Yi Zhang, Hui Qiao, Hao Xie, Yulong Li, Haoqian Wang, Lu Fang, and Qionghai Dai. Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nature Biotechnology*, 41(2):282–292, February 2023. Publisher: Nature Publishing Group.

[57] Changyeop Shin, Hyun Ryu, Eun-Seo Cho, Seungjae Han, Kang-Han Lee, Cheol-Hee Kim, and Young-Gyu Yoon. Three-dimensional fluorescence microscopy through virtual refocusing using a recursive light propagation network. *Medical Image Analysis*, 82:102600, November 2022.

[58] Eun-Seo Cho, Seungjae Han, Kang-Han Lee, Cheol-Hee Kim, and Young-Gyu Yoon. 3DM: deep decomposition and deconvolution microscopy for rapid neural activity imaging. *Optics Express*, 29(20):32700–32711, September 2021. Publisher: Optica Publishing Group.

[59] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, November 2017. Publisher: Optica Publishing Group.

[60] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, August 2023. arXiv:1706.03762 [cs].

[61] Yoav Levine, Noam Wies, Or Sharir, Nadav Cohen, and Amnon Shashua. Chapter 7 - Tensors for deep learning theory: Analyzing deep learning architectures via tensorization. In Yipeng Liu, editor, *Tensors for Data Processing*, pages 215–248. Academic Press, January 2022.

[62] Cong Chen, Kim Batselier, and Ngai Wong. Chapter 8 - Tensor network algorithms for image classification. In Yipeng Liu, editor, *Tensors for Data Processing*, pages 249–291. Academic Press, January 2022.

[63] Courtney J. Spoerer, Patrick McClure, and Nikolaus Kriegeskorte. Recurrent Convolutional Neural Networks: A Better Model of Biological Object Recognition. *Frontiers in Psychology*, 8, September 2017. Publisher: Frontiers.

[64] Gil Keren and Björn Schuller. Convolutional RNN: an Enhanced Model for Extracting Features from Sequential Data, July 2017. arXiv:1602.05875 [stat].

[65] Runpeng Cui, Hu Liu, and Changshui Zhang. Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1610–1618, Honolulu, HI, July 2017. IEEE.

[66] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, November 1997.

[67] Dong et. al. Reversible and long-term immobilization in a hydrogel-microbead matrix for high-resolution imaging of Caenorhabditis elegans and other small organisms. 2018.

[68] Feng et. al. A MODIFIED MINIMAL HEMOLYMPH-LIKE SOLUTION, HL3.1, FOR PHYSIOLOGICAL RECORDINGS AT THE NEUROMUSCULAR JUNCTIONS OF NORMAL AND MUTANT DROSOPHILA LARVAE. 2004.

[69] Parisa Kakanj et. al. Long-term in vivo imaging of Drosophila larvae. 2020.

[70] Mateusz Mołoń. Effects of Temperature on Lifespan of Drosophila melanogaster from Different Genetic Backgrounds: Links between Metabolic Rate and Longevity. 2020.