Technische Universität Dresden

Computational Methods for Visualization, Simulation, and Restoration of Fluorescence Microscopy Data

Dissertation

zur Erlangung des akademischen Grades
Doktor rerum naturalium (Dr. rer. nat.)

vorgelegt an der
Technischen Universität Dresden
Fakultät Informatik

eingereicht von

Martin Weigert
geboren am 29.01.1983 in Karl-Marx-Stadt

Gutachter:
Prof. Dr. Eugene W. Myers, MPI-CBG, Technische Universität Dresden
Prof. Dr. Rainer Heintzmann, Friedrich-Schiller-Universität Jena
Tag der Verteidigung: 07. Mai 2019

Dresden, den 24. Oktober 2019

Abstract

Fluorescence microscopy is an indispensable tool for biology to study the spatio-temporal dynamics of cells, tissues, and developing organisms. Modern imaging modalities, such as light-sheet microscopy, are able to acquire large three-dimensional volumes with high spatio-temporal resolution for many hours or days, thereby routinely generating Terabytes of image data in a single experiment. The quality of these images, however, is limited by the optics of the microscope, the signal-to-noise ratio of acquisitions, the photo-toxic effects of illumination, and the distortion of light by the sample. Additionally, the serial operation mode of most microscopy experiments, where large data sets are first acquired and only afterwards inspected and analyzed, excludes the possibility to optimize image quality during acquisition by automatically adapting the microscope parameters. These limits make certain observations difficult or impossible, forcing trade-offs between imaging speed, spatial resolution, light exposure, and imaging depth. This thesis is concerned with addressing several of these challenges with computational methods. First, I present methods for visualizing and processing the volumetric data from a microscope in real-time, *i.e.* at the acquisition rate of typical experiments, which is a prerequisite for the development of adaptive microscopes. I propose a low-discrepancy sampling strategy that enables the seamless display of large data sets during acquisition, investigate real-time compatible denoising, convolution, and deconvolution methods, and introduce a low-rank decomposition strategy for common deblurring tasks. Secondly, I propose a computational tractable method to simulate the interaction of light with realistically large biological tissues by combining a GPU-accelerated beam propagation method with a novel multiplexing scheme. I demonstrate that this approach enables to rigorously simulate the wave-optical image formation in light-sheet microscopes, to numerically investigate correlative effects in scattering tissues, and to elucidate the optical properties of the inverted mouse retina. Finally, I propose a data-driven restoration approach for fluorescence microscopy images based on convolutional neural networks (CARE) that leverages sample and imaging specific prior knowledge. By demonstrating the superiority of this approach when compared to classical methods on a variety of problems, ranging from restoration of high quality images from low signal-to-noise-ratio acquisitions, to projection of noisy developing surface, isotropic recovery from anisotropic volumes, and to the recovery of diffraction-limited structures from widefield images alone, I show that CARE is a flexible and general method to solve fundamental restoration problems in fluorescence microscopy.

Acknowledgements

This thesis, which is the result of my work at the group of Gene Myers at the Max Planck Institute for Molecular Cell Biology Dresden (MPI-CBG), would not have been possible without the help of many people.

First and foremost, I thank my supervisor *Gene Myers* for the opportunity to pursue this project, for providing guidance while giving me the freedom to explore many different ideas, and for energetic encouragements on numerous occasions. I would like to thank the members of my thesis advisory committee, *Ivo Sbalzarini* and *Frank Jülicher*, for their support and insightful comments. Furthermore, I am grateful to *Pavel Tomancak* and *Moritz Kreysing* for many fruitful discussions, and *Rainer Heintzmann* for agreeing to review my thesis.

During my time at the Myers Lab, I was surrounded by many amazing colleagues. I am particularly grateful to *Loic Royer*, who first introduced me to the world of modern microscopy and who has been an exuberant source of advice, ideas, wisdom, and dropping stick sounds. I thank *Nicola Maghelli*, whose optical wizardry and Italian dishes saved the day many times early on, and *Florian Jug*, whose enthusiasm and pun generating ability has been unparalleled. It was a great pleasure to meet and work with *Uwe Schmidt*, from whom I learned many things about computer vision, machine learning, and Indonesian culture. I thank current and past members of the Myers lab *Laurent Abouchar*, *Fabio Cunial*, *Corinna Blasse*, *German Tischler*, *David Chen*, *David Richmond*, *Dagmar Kainmüller*, *Debayan Saha*, *Coleman Broaddus*, *Alexandr Dibrov*, *Robert Haase*: it was a pleasure working with all of you.

Finally, I thank *Resi Unsinn*, *Marc Rohkost*, all my friends, and my family for their constant support throughout the years.

Contents

List of Figures

List of Tables

1 Introduction

In the first half of the 17th century advances in optical lens making and precision mechanics spawned two twinned devices that were destined to transcend the bounds of human perception: The telescope and the microscope. In the skillful hands of natural scientists of that time such as Galileo, Cassini, Hooke, and van Leeuwenhoek these instruments provided the first glimpses into the hitherto unknown world of the very large and very small: the *macroscopic* realm of the starry sky soon discovered to be inhabited by extended objects surrounded by rings of gleaming dust and with moons bound by their gravitational spell; and the *microscopic* domain of living matter soon found to be composed of complex arrangements of cellular compartments and populated by single-cellular animals too small for the naked eye to see. In the four centuries that followed, microscopes (and telescopes) came a long way: the advent of adequate physical theories of light (Newton, Fresnel, Huygens), the theoretical foundation of optical resolution (Rayleigh, Abbe), experimental ways to increase imaging contrast via staining and fluorescent labeling (Gram, Golgi, Chalfie), the invention of strong coherent light sources (Townes, Gould), and advanced optical designs (Köhler, Siedentopf, Minsky, Hell, Betzig) led to tremendous progress in their applicability in particular for medicine and biology and resulted in a vastly improved understanding of pathogen-mediated diseases, and of the structural properties of tissue and organs.

Today, light microscopy is an essential tool for *developmental biology*, where one studies the remarkable process by which a single cell - the fertilized egg - grows into a complex multi-cellular organization of layered tissue that develops into an organism capable of autonomously interacting with its environment, all within days or weeks. A standard tool in biology to investigate these dynamic processes is *fluorescence microscopy*. Here, *fluorescent* labels are chemically or genetically attached to specific structures of a specimen, of which then many volumetric images over the course of multiple hours or days are acquired. In Fig. 1.1 we depict several examples of common biological samples imaged in that fashion: a) The early embryo of the fruit fly *Drosophila melanogaster* at the *blastoderm stage*, where several thousand fluorescently labeled cells are crowding on the periphery of an egg half a millimeter in length, b) a small segment of the developing wing of a *Drosophila* pupea with labeled cell membranes, showing the hexagonal arrangement of cells in a dynamic sheet, and c) an early embryo of the roundworm *C. elegans* at 2-cell stage with histone-labeling, showing the condensation of chromatin in preparation for the next cell division. In recent years, advances in hardware control and optical instrumentation

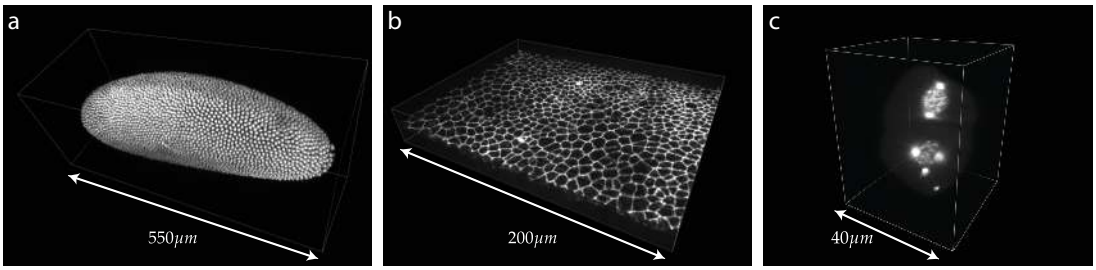


Figure 1.1: Examples of volumetric acquisitions of fluorescence light microscopes: a) An embryo of the fruit fly *Drosophila melanogaster* (nuclei labeling, His-YFP) , b) a section of the developing wing of a *Drosophila* pupea (membranes labeling, Ecad-GFP), and c) early cell divisions in a *C. elegans* (roundworm) embryo (histone labeling, His-YFP).

have led to the development of more and more complex fluorescence microscopy designs - most prominently light-sheet microscopy - that enable the observation of developing organisms at unprecedented spatio-temporal resolution, and produce long time-lapses of large volumetric signal measurements that can constitute several Terabytes of data.

Yet, fundamental imperfections of the acquisitions process impair the unobstructed view on many biological phenomena: the limited optical resolution of any microscope, noise in the detection process, or the distortive interaction of light with the sample, often extensively degrade the quality of acquired images and necessitate the *post-hoc* inference of the correct structures from imperfect observations. As a consequence, computational restoration and inference routines are becoming increasingly important in fluorescence microscopy, and today form the basis of super resolution microscopy (Heintzmann et al. 1999; M. G. Gustafsson 2000; Betzig et al. 2006), deconvolution (Richardson 1972; Preibisch et al. 2014), surface projection algorithms (Blasse et al. 2017; Shihavuddin et al. 2017), and denoising methods (Buades et al. 2005; Morales-Navarrete et al. 2015). All of these allow one to *computationally* recover important biological information that otherwise would stay hidden in the raw images. Still, most of these computational methods are only applied long after the data acquisition, and often ignore the experimental specifics of the biological sample and the image formation process.

This thesis is primarily concerned with addressing via computational means these challenges of fluorescence microscopes that acquire volumetric images of biological samples. It is guided by the belief, that the augmentation of optical hardware and software control with powerful on-board computational methods will transform classical microscopes into potent “discovery machines”. These devices will be capable of autonomously deciding on imaging conditions, able to extract biological information during acquisition, and will be driven by the increasing prevalence of fast computing platforms available to ordinary microscopes, *e.g.* via massively parallel GPU accelerators. With this long-term goal in mind, this thesis seeks to investigate several fundamental aspects pertaining to the computational augmentation of the

imaging process in microscopy: First, we will ask whether *visualization and low-level image processing* of volumetric acquisitions can be made compatible with *real-time* application on a microscope. We next will study whether *fast simulation methods for light propagation* through scattering biological tissue can be used to form a computational understanding of the image formation process in microscopy. Finally, we will consider how *machine-learning* based approaches can leverage prior structural information of biological structures to yield superior image restoration methods. This might eventually result in the design of microscopes that understand how to extract structural biological information from measured images by directly learning from large amounts of previously acquired data.

1.1 Fluorescence Microscopy

A microscope is an optical device that produces a magnified image of an object by optically refracting light originating from the sample by a cascade of multiple lenses and projecting the enlarged image onto a photon detector (camera or eye). In *fluorescence microscopy*, an external light source is used to excite fluorescent molecules within the specimen and to detect the subsequently emitted light. As the excitation and emission light can easily be optically separated[1] and different biological structures (*e.g.* nuclei, membranes) can be labeled by different fluorescent dyes or genetically expressed fluorescent proteins this results in an imaging method of high structural specificity. The excitation light is commonly provided by a laser of appropriate wavelength. A fundamental physical limitation of any optical microscope is its *finite resolution*, which is the smallest length-scale of objects that can be unambiguously discerned by the imaging process. A spherical wave emitted by a perfect point-source in the sample domain can only be partly captured by the *objective* - the primary lens closest to the sample - and will thus be refocused in the imaging domain as an extended light-field which is called the *point spread function* (Fig. 1.2a). The maximal attainable spatial resolution is given by $\Delta x \approx \frac{0.61\lambda}{NA}$ laterally, and $\Delta z \approx 2\frac{n_0\lambda}{NA^2}$ axially[2], where λ is the wavelength of emitted light, n_0 is the refractive index of the medium, and $NA = n_0 \sin\alpha$ is the numerical aperture of the objective[3] that admits rays under a maximal half-angle α (Born et al. 1999). A different way to look at this phenomenon, originating from the seminal work by Abbe 1873, is to consider an microscope as an imaging device with an effective cut-off in the frequency domain, which only admits the passage of spatial frequencies below the resolution limit (Goodman 1996). As a result the resolution of a microscope operating with visible light is generally limited to $180nm - 350nm$ laterally and $500nm - 900nm$ axially, which is called the *diffraction limit*.

[1] The wavelength of the emitted light is bigger than that of the excitation light by the molecule specific *Stokes shift*.

[2] As is common, we denote by x/y the *lateral* dimensions orthogonal to the optical path, and by z the *axial* dimension parallel to the line of light propagation.

[3] *I.e.* Objectives with higher NA result in better resolution.

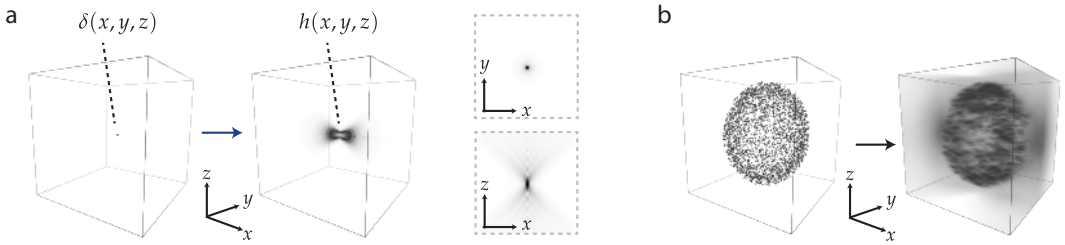


Figure 1.2: Fundamental resolution limit in microscopy: a) For any optical microscope, the image of a point source $\delta(x,y,z)$ is an axially elongated point spread function $h(x,y,z)$. b) This leads to anisotropic image blur and places a fundamental limit to the optical resolution.

When illuminating the extended sample volume with a uniform light distribution (*epifluorescence microscopy*), only the emitted signal from structures near the *focal plane* of the objective will be imaged with maximal sharpness, whereas light from out-of-focus planes will be blurred (Fig. 1.2b). By either moving the sample or focal plane along the optical axis z, a series of two dimensional images can be acquired that form a three dimensional image (*z-stack*) of the volume. To reduce the intrusive effect of out-of-focus blur, each point in the volume can be selectively illuminated and its signal detected in a scanning fashion, forming the basis of *confocal laser-scanning microscopy* (Minsky 1961; J. Pawley 2010) and *spinning-disk microscopy* (Petráň et al. 1968; Gräf et al. 2005) that both became a mainstay for imaging in biology.

In the last two decades, several approaches to circumvent the *diffraction limit* have been introduced, marking the beginning of *super-resolution microscopy*. *E.g.* S. Hell et al. 1992 demonstrated with *4π-microscopy* that axial resolution can be improved to match lateral resolution by coherently exciting/detecting the specimen with two opposing objectives. Another class of super-resolution techniques uses the chemical property of certain fluorophores to exhibit different *photo-switchable* (on/off) states. For instance in STED/RESOLFT (S. W. Hell et al. 1994; Grotjohann et al. 2012), fluorophores are depleted via stimulated emission with a second laser in a small doughnut-shaped region that is then scanned across the sample, delivering lateral resolution on the order of $20 - 50nm$ (Westphal et al. 2005). In *single-molecule localization microscopy* (SMLM) the same properties of photo-switchable fluorophores is used to stochastically convert the plurality of fluorophores into the off state, resulting in only a sparse subset of all fluorophores to emit photons which can be detected and mapped to the precise location. This way, a sequence of (typical thousands) of widefield frames can be converted to a super-resolved image, as first demonstrated with PALM (Betzig et al. 2006) and STORM (Rust et al. 2006). A third class of super-resolution method, *structured illumination* (SIM), uses the *Moiré effect* to combine different frequency information from several non-redundant acquisitions of the same structures illuminated with different light pattern, thus achieving a doubling of lateral resolution (Heintzmann et al. 1999; M. G. Gustafsson 2000). Still super-resolution microscopy has yet to gain traction in the study of long-term volumetric imaging of large, living organisms, due to its often high excitation intensities being incompatible with sample health, the reliance on special fluorophores, the typically

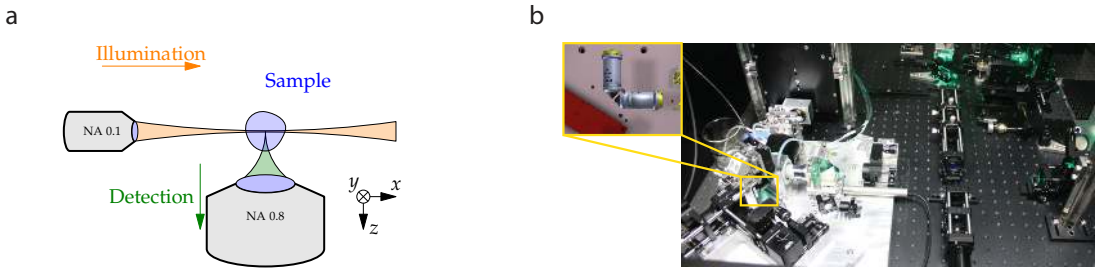


Figure 1.3: Principle of light-sheet microscopy: a) Illumination of the sample with a thin light-sheet and orthogonally detecting the excited fluorophores results optical sectioning and vastly reduces exposure of the sample to the destructive effects of light. b) Typical setup of a custom-built microscope. Inset shows the orthogonally arranged illumination and detection objectives.

slow acquisition speed, and its inefficiency deep inside tissues.

In recent years, *light-sheet fluorescence microscopy* (LSFM) has emerged as the imaging method of choice for time-lapse acquisitions of developing organisms on the whole tissue scale. In LSFM, a thin sheet of light is created by either a cylindrical lens (*selective plane illumination microscopy*, SPIM) or by scanning a focused beam (*digitally scanned light-sheet microscopy*, DSLM) that illuminates the sample only within a confined volume (Voie et al. 1993; Huisken et al. 2004; Keller et al. 2008b). By placing the detection objective *orthogonal* to the illumination objective with its focal plane parallel to the excited volume, the whole field of view can be imaged at once resulting in exceptionally fast acquisition speeds (Fig. 1.3a). By either moving the sample or translating the illumination sheet and focal plane simultaneously, a whole volumetric image can then be acquired. Importantly, the axially confined width of the light sheet not only provides intrinsic *optical sectioning* and vastly reduces out-of-focus signal, but as well massively lowers the amount of photo-damage inflicted on the specimen, paramount for long-term imaging of living samples. Although the principle of LSFM was known for a long time (*Ultramicroscope*, Siedentopf et al. 1902), the combination of fluorescent labeling techniques, sophisticated hardware control, and powerful reconstruction algorithms has led to a rapid development in the field (Tomer et al. 2012; Royer et al. 2016; Power et al. 2017). The relative ease with which custom optical configurations can be assembled and extended (Fig. 1.3b), resulted in the availability of open microscopy platforms (OpenSPIM, Pitrone et al. 2013), multi-view illumination strategies (Kumar et al. 2014; Preibisch et al. 2014), sophisticated illumination modes as in *lattice light-sheet* (Chen et al. 2014), adaptive acquisition control (Royer et al. 2016), or successful application of *adaptive optics* that can correct for tissue induced optical distortions (K. Wang et al. 2014; Liu et al. 2018). For these reasons, light-sheet microscopy has become an indispensable tool in the life sciences for investigating the dynamics of cells, tissues, and developing organisms at sub-cellular resolution over long imaging times (hours, days) with high acquisition rate, with experiments routinely producing Gigabytes of data per minute and several Terabytes per day (Amat et al. 2015).

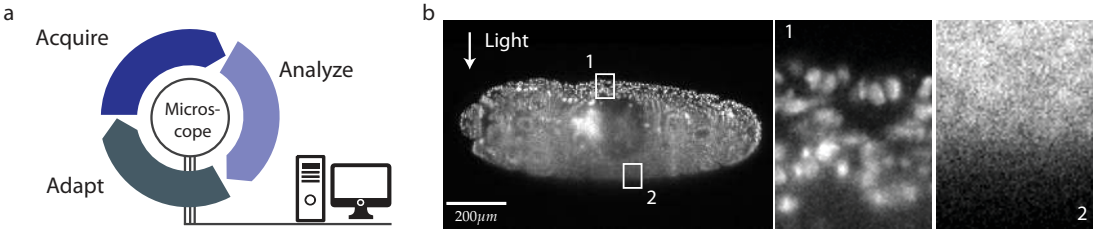


Figure 1.4: Challenges in fluorescence microscopy: a) Enabling smart, adaptive microscopes that would integrate acquisition, analysis, and adaption of imaging parameters within a tight feedback loop. b) Interaction of light with tissue leads to sever loss of image contrast deep inside tissues. Depicted is a light sheet illuminating the midsection of a *Drosophila* embryo, with a region unaffected by prior light-tissue interaction (1) and a region with poor signal due to scattering of the light on its way through the sample (2).

1.2 Challenges

Apart from the described advances in optical designs, fluorescence microscopy exceedingly relies on computational methods that can adaptively control image acquisition parameters, restore the latent fluorescent signal from imperfect observations, and interpret the resulting images in biological meaningful ways. This emerging trend of *computational microscopy* and *smart microscopy* (Scherf et al. 2015; Royer et al. 2016) calls for the tight interplay of classical optical techniques with powerful computational control and restoration pipelines, aiming to engage with the following key challenges of fluorescence microscopy:

Challenge 1 *How to enable smart, adaptive microscopy*

Most long-time microscopy experiments are executed in a sequential mode, where data is first acquired, and only afterwards visualized and processed. This makes it impossible to automatically adapt imaging parameters and perform experimental interventions during the course of acquisition, which would improve the extraction of meaningful biological information. The goal of *smart, adaptive microscopy* is to integrate acquisition, analysis, and adaption of imaging parameters within a tight feedback loop, where the analysis of acquired data can provide immediate feedback to the microscope control software (Fig. 1.4a). It is therefore necessary to investigate fast and efficient methods to analyze the constant stream of data acquired by the microscope *in real-time*, *i.e.* at a rate compatible with the typical acquisition time. This includes visualization methods, low-level image processing pipelines, and quality assessment of microscopy images.

Challenge 2 *How to image deeper into tissues*

Interaction of light with the biological sample itself causes severe image degradation in whole tissue microscopy, where the optical inhomogeneities of the specimen result in refraction, absorption, and scattering of the illumination and excitation light fields. This is particularly pronounced for regions inside deep inside of large tissues, where the attainable image quality is therefore vastly reduced (Fig. 1.4b). Consequently, classical microscopy is currently limited to

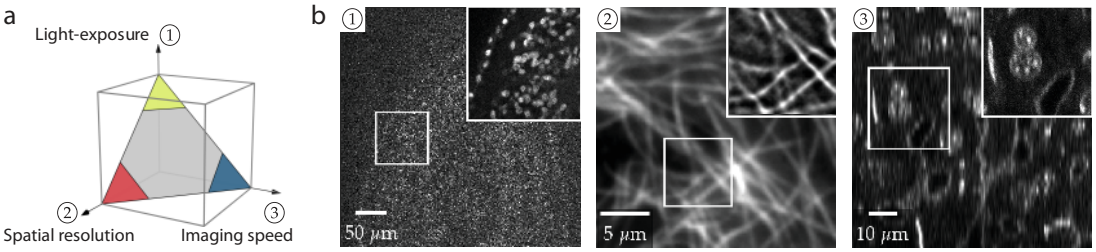


Figure 1.5: Challenges in fluorescence microscopy *(continued)*: a) Imaging trade-offs in live-cell microscopy: spatial resolution, imaging speed, and light-exposure (signal-to-noise ratio, SNR) cannot be maximized at the same time, enforcing trade-off of the imaging parameters. b) Common image degradations that result from these trade-offs: Low SNR images (left), loss of spatial resolution (middle), and anisotropic image blur (right). Insets show regions acquired under perfect imaging conditions.

imaging depths of $50\mu m - 150\mu m$ even for relatively transparent specimen. Approaches like adaptive optics or wavefront shaping promise to meliorate this situation. Complementary to these optical approaches a the proper theoretical understanding of light-tissue interaction especially deep inside tissue would help to computationally address this problem.

Challenge 3 *How to extract information from imperfect observations*

In fluorescence microscopy of living specimen, the best image quality at which biological processes can be faithfully recorded is not only determined by the spatial optical resolution, but similarly by the total duration of an experiment, the desired temporal resolution, and the adverse effects of light exposure on the fluorophores (photo-bleaching) and sample health (photo-toxicity) (Icha et al. 2017; Laissue et al. 2017). These aspects cannot all be optimized at the same time – one must make trade-offs, for example, sacrificing signal-to-noise ratio by reducing exposure time in order to gain imaging speed. Such trade-offs are often depicted by a *design-space* that has resolution, speed, light-exposure, and imaging depth as its dimensions (Fig. 1.5a) with the volume being limited by the maximal photon budget compatible with sample health (J. B. Pawley 2006; Scherf et al. 2015). The acquired volumetric images therefore often exhibit low signal-to-noise ratios, image blur of small structures, or anisotropically worse axial resolution (Fig. 1.5b). A fundamental problem in microscopy is to computationally restore the latent true signal from these imperfect observations.

1.3 Overview of Thesis and Contributions

In this thesis, we will try to address all described challenges by focusing on their computational aspects. In Chapter 2, we will investigate how to visualize, analyze and process the constant stream of volumetric data in *real-time*, *i.e.* within a time period that is compatible with the typical acquisition rate of *e.g.* light-sheet microscopy experiments. This includes fast rendering pipelines for large microscopy volumes

including a novel *low-discrepancy sampling* strategy, low-level image restoration methods such as GPU-accelerated denoising and deconvolution with *low-rank PSF decompositions*, and image sharpness measures for instant image quality feedback on the microscope. In Chapter 3, we will demonstrate how fast and accurate methods that simulate interaction of light fields with the sample allow to understand tissue induced distortion in scattering media and to reconstitute *in-silico* the wave-optical image formation process in microscopy for realistically large volumes. We validate our method on several analytical and experimental examples, investigate scattering correlations relevant for adaptive optics, and finally demonstrate is utility for biological questions by simulating the light propagation through large tissue models of retinal tissue, confirming earlier biological findings. In Chapter 4, we show how image restoration based on *deep learning* extends the range of biological phenomena observable by microscopy. We introduce *content-aware image restoration* (CARE) that incorporates both the imaging process and the complex statistical structure of images. We demonstrate that this provides a flexible and general method to solve fundamental restoration problems in fluorescence microscopy, namely restoration of high quality images from low signal-to-noise-ratio acquisitions, projection of noisy developing surface, isotropic recovery from anisotropic volumes, and resolving of diffraction-limited structures from widefield images alone.

The specific contribution of this thesis can be briefly summarized as follows:

Real-time Processing and Visualization of Microscopy Data (Chapter 2)

- Fast rendering schemes with low discrepancy sequences and density attenuation projection
- Real-time denoising by adapting Non-Local-Means (NLM) to GPU
- Efficient convolution algorithms via low-rank approximation of the PSF
- Real-time deconvolution via low-rank approximation of the PSF
- Open-source implementations in Java (*ClearVolume*) and Python (*spimagine*, *gputools*)

Parts of this chapter are published in (Royer et al. 2015).

Fast Multiplexed Models of Light Propagation in Biological Tissues (Chapter 3)

- Investigation and validation of a GPU-based beam propagation method for simulation of light propagation in biological, weakly scattering tissue
- Devising a novel multiplexing scheme for parallel PSF calculations inside of scattering media
- Providing the first rigorous simulations of wave-optical image formation for light-sheet microscopes
- Identifying a new correlative effect between scattering components in tissue (*extended memory effect*)

- Investigating *in-silico* the optical properties of the inverted nuclei architecture in mouse retina

Parts of this chapter are published in (Weigert et al. 2018b).

Content-Aware Image Restoration (Chapter 4)

- Introducing a deep learning based image restoration method for fluorescence microscopy via *content-aware image restoration* networks (CARE)
- Proposing different strategies to create training data *physically*, *semi-synthetically*, or *fully synthetically*
- Recovery of extremely low signal-to-noise ratio with up to 60 fold reduced exposure/laser-power, enabling live-cell imaging of the light sensitive species *Schmidtea mediterranea*
- Joint denoising and surface projection of *Drosophila* epithelia tissue that allows to image 6-10 times faster
- Proposing a novel recovery strategy (ISONET) for the typically poor axial resolution of microscopes enabling up to 10 fold reduction of axial acquisitions
- Demonstrating the real-time ability of a neural network based whole tissue restoration pipeline running on a custom light-sheet microscope
- Resolving diffraction-limited structures such as microtubules and granules from wide-field images at 20-times higher frame-rates compared to state-of-the-art methods

Parts of this chapter are published in (Weigert et al. 2017), (Weigert et al. 2018a), and (Sui et al. 2018).

In the text of this thesis, the pronoun *"we"* generally refers to the author and the reader. In case certain parts involve the collaboration with other individuals, they are named accordingly.

2 Real-time Processing and Visualization of Microscopy Data

Overview

Owing to advances in hardware control, camera recording speed and sensitivity, modern light-sheet microscopes are capable of acquiring exceedingly large volumes at an increasingly high temporal resolution, and multiple views and colors (Chhetri et al. 2015; De Medeiros et al. 2015; McConnell et al. 2016; Royer et al. 2016). The rate at which volumetric data is produced by these microscopes is on the order of 0.1-1GB per second, and several TB of image data are generated during the course of a typical 24h recording. Commonly, an experiment starts with the adjustment of imaging parameters (such as camera exposure, focus, field of view), after which an automated process acquires the volumetric time-lapse recording which is then saved to disk concurrently. It is thus only *after* the completion of the entire experiment, that the resulting volumes are fully visualized, processed and analyzed (Tomer et al. 2012; Chen et al. 2014), making it difficult to adapt imaging parameters and perform experimental interventions during the course of acquisition. In contrast to such a *post-hoc* approach of analyzing microscopy data, an ideal way would be to perform as many of these steps concurrently, *i.e.* in real-time alongside the microscope acquisition, and while the experiment is performed and imaging specifications can still be adjusted. This would have the following advantages: *i)* Visualizing the data in its entirety allows the live assessment of image quality and effect of imaging parameters

at every timepoint of the experiment, thus providing immediate feedback to the experimenter and the potential to adjust the imaging setup. *ii)* It minimizes the amount of necessary postprocessing steps that would otherwise have to be carried out after the experiment has been done. Finally, *iii)* such an approach would constitute an important step towards developing *smart microscopes*, that would integrate on-the-fly image processing, quality assessment, and selection of acquisition parameters in a semi-autonomous way (Scherf et al. 2015; Royer et al. 2016).

In this chapter, I will investigate how to visualize, analyze and process this constant stream of data in *real-time*, *i.e.* within a time period that is compatible with the constraints imposed by the data throughput of modern light-sheet microscopes, while using only commonly available microscopy hardware (*i.e.* a single workstation). The chapter is structured as follows: We first discuss related work in Section 2.1, and then describe some general considerations of parallelized computation paradigms in Section 2.2. In Section 2.3 we will discuss several approaches to visualization of microscopy data and show how real-time rendering can be achieved even for large microscopy volumes with a novel *low-discrepancy sampling* strategy. In Section 2.4 we then investigate the merits of several low-level image processing tasks for real-time application. Specifically, we will show that a classical denoising approach (Non-local-means) can be accelerated via GPU processing to serve as an adequately fast strategy for on-the-fly denoising of microscopy volumes. Furthermore, we discuss several approaches of fast image convolution and introduce the usage of *low-rank decompositions* to this problem. We then show how this enables the real-time application of classical deconvolution methods. After briefly discussing a real-time image sharpness measure, we finally describe all implementations in Section 2.5 and close with a summary in Section 2.6.

Contributions

1. Fast rendering schemes with low discrepancy sequences and density attenuation projection
2. Real-time denoising by adapting Non-Local-Means (NLM) to GPU
3. Efficient convolution algorithms via low-rank approximation of the PSF
4. Real-time deconvolution via low-rank approximation of the PSF
5. Open-source implementations in Java (*ClearVolume*) and Python (*spimagine*, *gputools*)

Parts of this chapter are published in Royer et al. 2015.

2.1 Related Work

Many software packages provide the ability to visualize static three-dimensional data. Among them are commercial tools such as Imaris (Bitplane AG) or Amira

(Thermofisher), which are expensive to acquire and difficult to extend. Open-source alternatives include ParaView (Ahrens et al. 2005), the slice-based BigDataViewer (Pietzsch et al. 2015), and the ImageJ 3D Viewer that provides 3D rendering primitives for the Java/Fiji ecosystem (Schmid et al. 2010). All of these are either ill-suited to time-lapse data or cannot interface with microscopy software directly, thus making their application impossible for our purposes.
Recently, real-time compatible image processing of 3D microscopy data has been explored by several software packages, mostly driven by performance increase offered of Graphic Processing Units (GPUs). GPU-implementations of popular *denoising* algorithms were designed and implemented by many groups (*e.g.* Goossens et al. 2010; Márques et al. 2013), yet the absence of accessible code and available implementations resulted in a lack of adoption by the microscopy community. For instance for *deconvolution*, *i.e.* the restoration of images blurred by the microscope PSF, Bruce et al. 2013 showed that GPUs enabled the deconvolution of 3D stacks at the acquisition-rate of spinning-disk microscopes. Schmid et al. 2015 demonstrated fast GPU-based multi-view deconvolution of light-sheet data by ignoring the lateral extent of the PSF. To address the challenges of exceedingly growing data sizes in microscopy Balazs et al. 2017 provided a real-time compression method for large microscopy data and Cheeseman et al. 2018 proposed a particle based representation of images. Going beyond pixel-level image restoration, Stegmaier et al. 2016 demonstrated real-time nuclei segmentation in developing *Drosophila* embryos, and Afshar et al. 2016 showed the utility of distributed particle methods for segmentation of large images.

2.2 GPU-based Image Processing

CPUs and GPUs: Traditionally, the processing of large images relied on the CPU (Central Processing Unit) as the underlying program execution platform. Although historically the raw processing speed of CPUs has seen exponential growth over the second half of the 20th century (Moore 1965; Brock et al. 2006), the numerical clock-rate of single-core CPUs has stalled since ~2004, mostly due to the constraints on power consumption and heat dissipation, as well as current leakage problems related to the (still exponentially) increased miniaturization of integrated circuits (Sutter 2005). The dominant source of performance gains can since then be attributed to *i)* the shift from single-core to multi-core architectures , and *ii)* the utilization of increasingly complex execution flow optimizations, such as instruction pipelining, out-of-order execution, and data level parallelism. An example of *i)* is the of use of multi-threaded programs with shared-memory (*e.g.* via `OpenMP`, Dagum et al. 1998) or distributed programming across several physical machines (*e.g.* via `MPI`, Gropp et al. 1996). An example of *ii)* is the addition of *single data multiple instruction* (SIMD) primitives to modern CPUs, which allow several operations on several elements of data processed within a single clock cycle (such as with the Intel AVX/AVX2 extensions or fused multiply-add operations FMA, Firasta et al. 2008).

(a)

Mode	GFLOP/s
CPU (naive, 1 core)	3.4
CPU (SIMD, 1 core)	37.7
CPU (SIMD +opt, 1 core)	98.1
CPU (SIMD +opt, 40 cores)	1230.8
GPU (naive)	9317.0

(b)

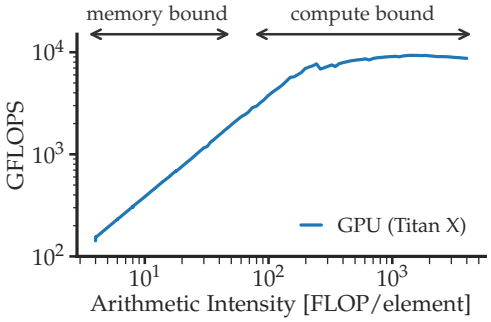


Figure 2.1: Performance considerations for an *embarrassingly parallel* problem: a) Peak performance for different CPU and GPU implementations. See main text for detailed explanation. b) Different performance regimes of a GPU program depending on the arithmetic intensity (operations per memory access).

On the other hand, hardware accelerators such as GPUs (Graphics Processing Unit) or FPGAs (Field-Programmable Gate Arrays) have in recent years outgrown their original confined application domain. Specifically GPUs have evolved from their primary scope of hardware-accelerating rendering in computer graphics to general-purpose computing devices (Pharr et al. 2005), with the two dominant programming models currently being CUDA (Nickolls et al. 2008) and OpenCL (Stone et al. 2010). A distinctive property of GPUs is their emphasis on a massively parallel SIMD execution model, often boasting thousands of compute-cores within a single device that are able to run the same program on different data elements in parallel. It is this efficiency of massive parallel operations, which make GPUs specifically attractive for image processing tasks (Alglave et al. 2015).

A simple experiment: To demonstrate the performance difference of CPUs and GPUs for pixel independent operations, let's consider the following simple experiment: Given an input array of numbers a_i we compute for every element $b_i = f_n(a_i)$, where f_n is a scalar operation consisting of n single-precision floating point operations. As this function is evaluated independently for every input value, this is the most simple instance of an *embarrassingly parallel* problem. A computer program implementing it will be compute-bound for large enough values of n (*i.e.* when the single memory access per element becomes negligible) and the overall run-time will give a rough estimate of the peak compute performance of a given implementation (calculated in Flop/s). We compare the following implementations:

- CPU naive: C++ code using a simple for-loop over all elements
- CPU SIMD: C++ code using AVX2 and fused-multiply-add (FMA) primitives
- CPU SIMD+opt: C++ code using AVX2, FMA and interleaved operations.
- CPU SIMD+opt, 40cores: Multi-threaded version via OpenMP on 40 cores.
- GPU naive: OpenCL code running with one thread per element.

As CPU we used a Intel Xeon CPU (Broadwell) E5-2630 (clock-rate 2.2GHz/3.1Ghz Turbo Boost) and as GPU a Nvidia Titan X (Pascal, 3584 cores at 1.5 Ghz boosted

clock rate), both located on the same workstation. The C++ code was compiled with aggressive optimization and auto-vectorization enabled[1]. The numerical peak performance in both cases are 99 GFlop/s for the single CPU core[2] and 11 TFlop/s for the GPU[3]. The measured peak performance achieved for each implementation is shown in Figure 2.1a. An interesting observations is that the naive CPU implementation excessively under-utilizes the CPU, and achieves only 3.4 GFlop/s, which is 3% of the numerical CPU peak performance. Changing the code to use 8-wide SIMD and FMA instructions already provides 37.7 GFlop/s, and interleaving instructions delivers 98.1 GFlop/s, or 38% and 99% of peak CPU performance respectively, which however comes with the increased burden of writing hardware specific code. Further parallelization over 40 cores results in 1.2 TFlop/s, indicating diminishing returns due to threading overhead[4]. Finally, the GPU implementation achieves 9.3 TFLOP/s, even when implemented with a naive for-loop. It is this superior performance, together with the ease of deployment (OpenCL is supported by most GPUs and even CPUs), which makes the GPU an attractive execution device for such tasks. Another important consideration for a specific image processing task is the proportion of computational effort spend per pixel in relation to the amount of memory accesses. Memory bandwidth is generally order of magnitudes lower than compute throughput, on both CPUs and GPUs. Thus, the performance on tasks that have few operations per memory access is often solely limited by memory access time, which is then said to be *memory bound*, in contrast to tasks that show many operations per memory access which are *compute bound*. This is illustrated in Figure 2.1b, where the GPU performance (for the same experiment as before) is shown as a function of the *arithmetic complexity*, *i.e.* the number n of floating point operations per memory access. As can be seen, memory access dominates for values $n < 100$ resulting in sub-peak performance, whereas only for $n > 100$ peak performance is achieved. For this reason, GPUs typically possess several hierarchies of dedicated fast memory caches, which includes *constant memory*, *shared/local memory* that is only addressable by group of threads, and *texture memory* that caches pixel-values in the vicinity of the recently accessed pixel of a multidimensional image.

2.3 Visualization of Microscopy Data

In the previous section, the utility of GPU-based implementations for pixel-parallel image operations was demonstrated. As a first application, we now consider the visualization of microscopy volumes during acquisition, *i.e.* the rendering of three-dimensional data on a microscope. Such real-time 3D visualization would be beneficial in several ways: it would *i)* allow one to instantly inspect and analyze the time-resolved (3D+t) data while being acquired, and virtually rotate and zoom into

[1] `g++-6 -O3 -march=native -mavx`
[2] 3.1 GHz $\times$ 32 single precision float operations (two 8-wide FMA instructions) per cycle
[3] 3584 cores $\times$ 1.53 GHz $\times$ 2 single precision float operations per FMA instructions
[4] Additionally, Turbo-Boost can only be enabled for a small number of cores simultaneously

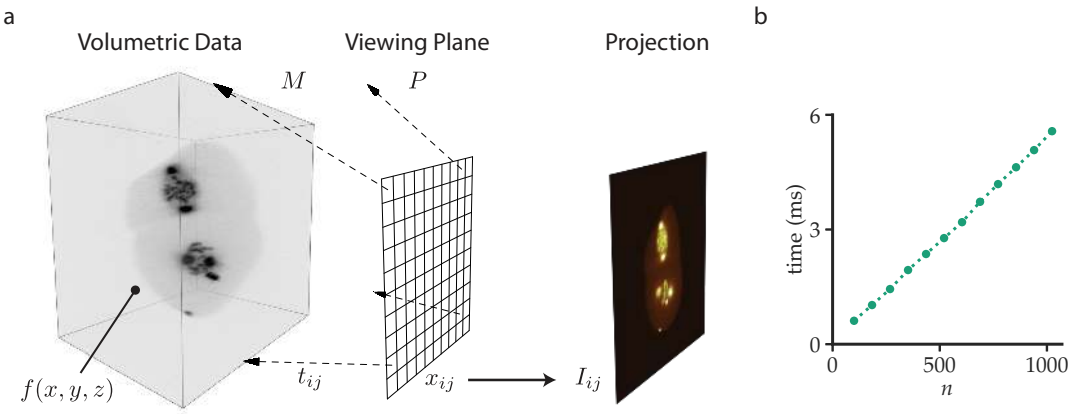


Figure 2.2: a) Direct volume rendering by maximal-intensity ray casting. For every pixel x_{ij} of the viewing plane the 3D volume intensities $f(x,y,z)$ are sampled along a ray t_{ij}, whose maximal value I_{ij} is rendered with a given colormap to produce the final image. b) Rendering time for a cubic volume of size $n \times n \times n$ pixels on a Titan X(Pascal) GPU. Note that the size of the rendering target was fixed, which results in linear increase of rendering time.

the structures, *ii)* provide an instantaneous overview of a specimens viability and health, and *iii)* facilitate the process of alignment and image parameter tuning (*e.g.* focus, laser power). Still, most currently employed advanced light-sheet microscopes only display the latest image plane acquired during acquisition (Tomer et al. 2012; Chen et al. 2014) and only in a second step this data is processed and visualized offline with commercial[5] or open-source tools (Schmid et al. 2010).

Due to the abundance of volumetric data in many scientific fields, visualization methods have a long history, going back to early work in rendering CT scans and MRI data in bio-medical image analysis (Mazziotta et al. 1976; Drebin et al. 1988). Generally, one can distinguish between *surface rendering*, where a specific isosurface of the dataset is converted into a polygonal mesh (Lorensen et al. 1987), and *volume rendering*, where information at every pixel is incorporated into the final rendering (Drebin et al. 1988; Q. Zhang et al. 2011). In the following, I describe how both can be efficiently implemented on GPUs and how this results in robust and fast visualization modules that allow to render volumetric microscopy data in real-time.

2.3.1 Volume Rendering

We first consider *volume rendering*, *i.e.* the conversion of a three dimensional volume to a two dimensional image (the rendered output) depending on a geometric transformation given by the user input (*e.g.* rotation, zoom). The standard method to achieve this is to perform *volume ray casting* (Roth 1982), whose basic concepts are depicted in Fig. 2.2a. First, for every pixel position (i,j) of the final image a ray with position x_{ij} and direction $t_{i,j}$ is constructed in 3D world space according to a specified (camera) viewpoint. This viewpoint is generally encoded in a *model-view matrix* M and a *projection matrix* P that both describe an affine transformation of the

[5] *e.g.* Imaris (Bitplane AG) or Amira (Thermofisher)

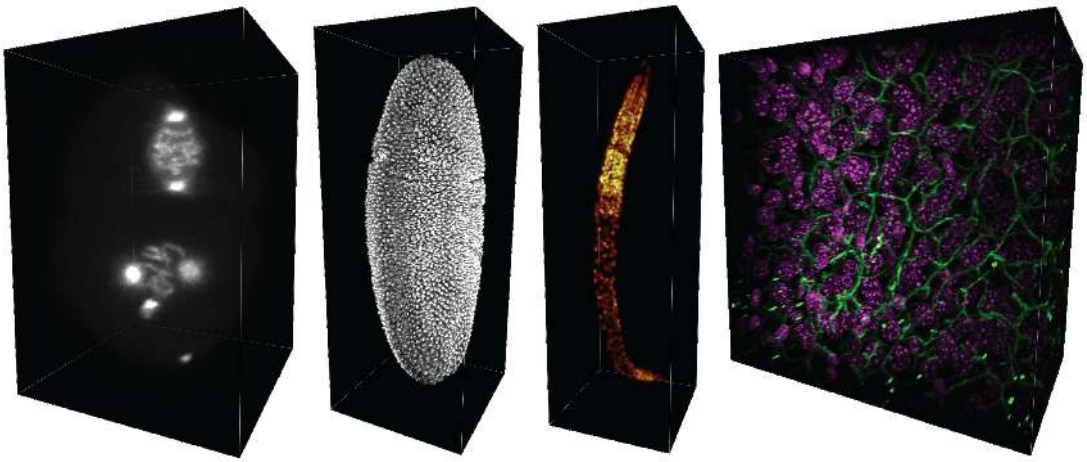


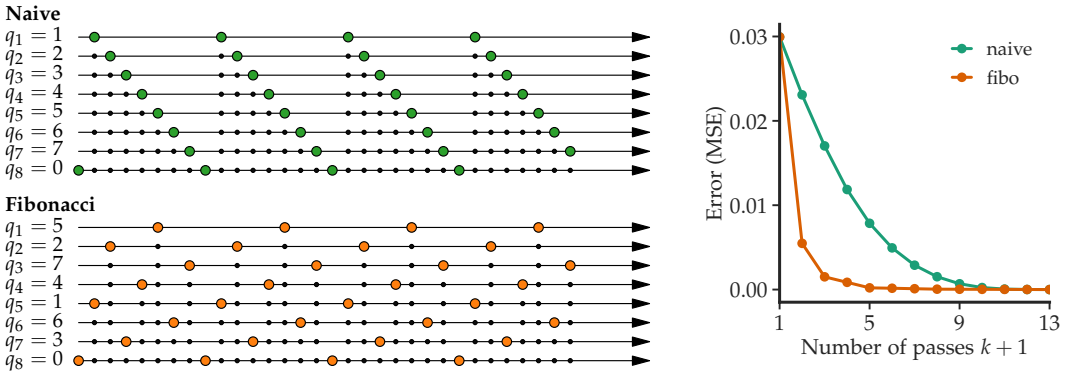
Figure 2.3: Example renderings of microscopy acquisitions. From left to right: *C. elegans* early cell divisions, early *Drosophila* embryo, *C. elegans* L1-stage worm, and two-channel liver tissue (sinusoidal network/green and hepatocytes/magenta).

4-dimensional homogeneous coordinate space[6]. Next, the volumetric intensity distribution $f(x, y, z)$ is sampled along this ray at discrete (equidistant) points p_n and projected to a single value $\tilde{p}$. A simple projection function is the maximum function $\tilde{p} = \max_n f(p_n)$ resulting in the commonly used *maximum intensity projection* (MIP). The final color value at that point is then calculated by applying a transfer function (colormap) to the projected value. Although all the above outlined operations (transformation, sampling, projection) could be implemented directly from within OpenGL (Shreiner 2009), I chose to implement them as dedicated OpenCL kernels, as this allows for greater flexibility and future extensibility. In this way, for every pixel of the output image an individual thread can be used and each output intensity pixel I_{ij} can be calculated in parallel. For multi-channel images, all rendering steps are carried out sequentially for every channel, producing the final rendered image as an overlay of each channels individual renderings. As the whole process is memory bound due to the sampling operation, the usage of texture memory for the input volume is paramount, as its caching of pixel neighborhood values is well suited for the ray sampling operation. To test the overall performance, I measured the wall time to render volumes of increasing size to an output texture of constant size ($w/h = 1024/800$), whose results are displayed in Fig. 2.2b: Even large single-channel volumes of pixel dimension 1024^3 can be rendered in $\approx$ 6ms, resulting in rendering speeds that are compatible with real-time applications. In Fig. 2.3, we show renderings of several biological samples.

Ray Casting with Low-discrepancy Sequences

For exceedingly large (>2 GB) volumes with several channels, the full ray traversal might still cause a slight latency between user interaction requests and the required rendering updates. It is therefore advantageous to employ a *multi-pass strategy*, *i.e.* to

[6] 4D homogeneous coordinates is a standard way in computer graphics to allow affine transformations (*e.g.* translations) to be expressed with a single matrix operation



(a) Illustration of multipass strategies

(b) Rendering error

Figure 2.4: Rendering with low-discrepancy sequences. a) Illustration of different multi-pass strategies, including using naive/linear offsets (green) and low-discrepancy Fibonacci numbers (orange). b) Error (MSE) of intermediate renderings (using maximum projection) at different multi-pass steps for $m = 13$ and a *Drosophila* volume. (Illustration adapted from Royer et al. 2015)

break down the full traversal into several, faster sub-operations, whose output each approximate the desired final rendering and which can be interrupted in case of a quick succession of changed viewpoints or other user defined rendering parameters. In the case of maximum projection, an easy strategy uses the associativity of the maximum operation: One consecutively finds the maximum along a ray with a m-fold increased step-size and different offsets $q_0, \dots, q_{m-1} \in \mathbb{Z}_m$ and successively updates the overall maximum with the current sub-sampled maximum after each such pass. This will result in a succession of approximative rendering results $\{I_k\}_{0 \le k < m}$ whose last image I_{m-1} will agree with the overall maximum given that all possible offsets are represented, *i.e.* $\cup q_k = \mathbb{Z}_m$. To avoid perceivable rendering artifacts (*e.g.* flashes) the consecutive passes should sample the full ray as uniformly as possible, which asks for a certain choice of offsets q_k. A *naive* multi-pass strategy would simply consider consecutive offsets, *i.e.* choosing $q_k = k$. Such a choice, however, leaves large regions undersampled during the first passes as can be seen in the upper row of Fig. 2.4a. This property can be formalized by defining the discrepancy $\mathcal{D}$ of a set of numbers $A = \{a_1, \dots a_N\} \subset [0, 1]$ as

$$\mathcal{D} = \max_J \left| \frac{\#(a_i \in J)}{N} - |J| \right| \quad , J \subset [0, 1] \tag{2.1}$$

which measures the uniform-ness of the set A and sequences with low $\mathcal{D}$ are therefore called *low-discrepancy sequences* (Kuipers et al. 2012). Such low-discrepancy sequences have been used for ray tracing of 2d textures in (Wong et al. 1997) and (Pharr et al. 2016) provides many applications to Monte-Carlo integration that are relevant for physically realistic rendering. Clearly, choosing offsets $\{q_k\}$ with that property would be equally advantageous for the sub-sampling ray-cast in our case. Indeed, this can be achieved by selecting a number p that is coprime to m and choosing $q_k = k \cdot p \bmod m$. Further, an ideal choice of p to achieve a uniform and balanced sampling would be $p = \varphi^{-1} m$ with the golden ratio $\varphi = \frac{\sqrt{5}+1}{2}$. Being restricted

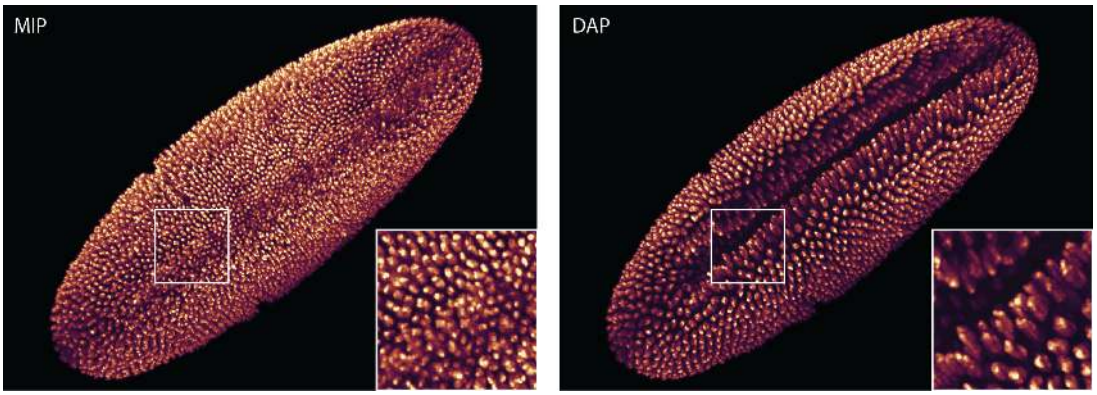


Figure 2.5: Intensity projection of a *Drosophila* stack with Maximum Intensity Projection (MIP, left) and with Density Attenuated Projection (DAP, right). The difference in structural perception is particularly visible around the ventral furrow (inset, Data from Royer et al. 2016).

to integer valued p and m, we can further use the well-known fact that consecutive *Fibonacci numbers* F_p are coprime and their ratios are the best rational approximants to the golden ratio[7] and therefore will always choose $(p, m) = (F_p, F_{p+1})$ [8]. Both the naive and the Fibonacci multi-pass strategy are illustrated in Fig. 2.4a where the latter samples the interval in a much more balanced and uniform way. This fact is further demonstrated in Fig. 2.4b showing the error of the intermediate maximum projections compared to the full maximum projection for each consecutive step and $m = 13$. As can bee seen, choosing the Fibonacci multi-pass strategy results in substantial faster decrease of the error compared to the naive strategy.

Density Attenuated Projections

Often using maximum intensity projection (MIP) as the projection function leads to suboptimal results. Specifically, it can be a poor indicator of the correct spatial relationship of structures residing at different depth: bright points that are far away can be projected to nearby locations, thus confounding their relative distances. A typical example can be seen in Fig. 2.5a, where the maximum projection of a *Drosophila* embryo during gastrulation is depicted. Although this developmental stage is characterized by the folding of the cellular surface into several furrows, the "inter-bleeding" of intensities from opposing sides clearly obstructs the appreciation of the embryos changing geometry. This can be remedied by introducing density depending attenuation, *i.e.* instead simply finding the maximum $\text{MIP} = \max_n f(p_n)$ of the signal along the ray, one finds the maximum of a weighted version

$$\text{DAP} = f(p_{\tilde{n}}) \text{ with } \tilde{n} = \arg\max_n f(p_n) w_n \quad , w_n = e^{-\sum_0^{n-1} f(p_n)}. \tag{2.2}$$

[7] In the sense that F_{p+1}/F_p is the p-th rational truncation of the continued fraction expansion $\varphi = [1, 1, 1, \dots]$

[8] Interestingly, the low-discrepancy property of this sequences is prominent in other areas, such as in quasi Monte Carlo methods, the ideal packing of sunflower seeds (Atela et al. 2002) and the break up behavior of the last regular phase space tori in nonlinear chaotic mappings (MacKay et al. 1984)

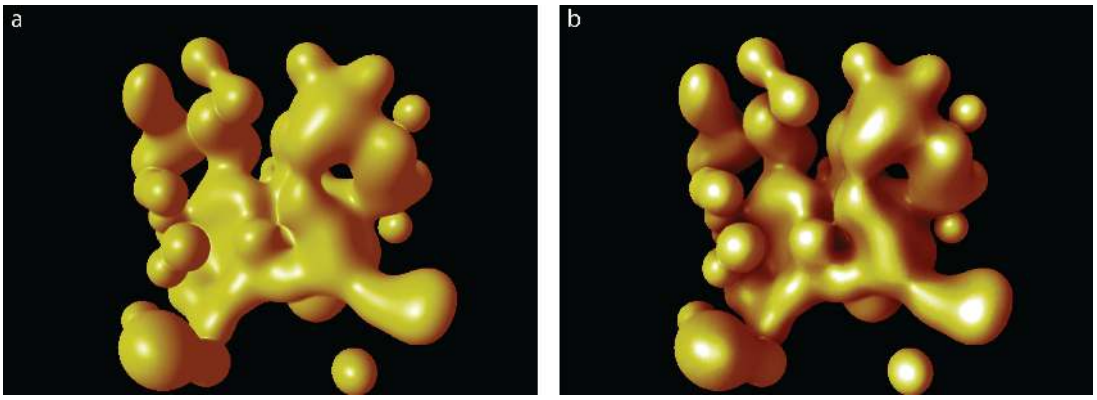


Figure 2.6: Isosurface rendering of synthetic blobs with a) normal Phong shading, and b) additionally with ambient occlusion.

Namely, the weights w_n get attenuated proportional to the amount of already traversed signal, much akin to being attenuated by an absorbing optical medium. Note that w_n can be efficiently computed while traversing the ray $w_{n+1} = w_n e^{-f(p_n)}$. A comparison of these two ways of projection is given in Fig. 2.5, where the improvement in surface perception can clearly be seen, especially in the heavily folded region around the ventral furrow.

2.3.2 Isosurface Rendering

Finally, we consider *isosurface rendering*, *i.e.* the selection and rendering of a surface of constant intensity $f(x, y, z) = c$ with a user defined intensity c. A simple way of finding this surface is to (as before) simply traverse the ray until the function $g(p_n) = f(p_n) - c$ first changes its sign. Afterwards, one computes the normal ∇f at this point via finite differences and calculates the final rendered color via a global illumination model, *e.g.* Phong-shading (Phong 1975). An example of this simple approach is shown in Fig. 2.6a, where a isosurface of synthetically created Gaussian blobs is rendered. I observed that sufficient rendering detail can be achieved without converting the surface into a polygonal mesh, as could have been done with the *marching cubes* algorithm (Lorensen et al. 1987). Furthermore, I incorporated more advanced shading models such as *ambient occlusion* (Miller 1994), where the shading model is modified to allow for illumination occlusion of nearby surface points. An example is shown in Fig. 2.6b, where one can observe that especially points near saddle points and surface grooves are easier to discern.

2.4 Low-level Image Processing

After having discussed how visualization of microscopy data can be efficiently achieved, we now consider the task of processing this data within the temporal constraints of the acquisition process. In the following, we will concentrate on low-level operations that are typically applied to all acquired images regardless of the further biological question. These tasks include *denoising* (removing the inevitable noise

inflicted by the camera detection process), *deconvolution* (reverting the image blur resulting from the diffraction limited resolution of microscopes), and the calculation of image *sharpness measures* (to be used for calibration and autofocus).

2.4.1 Denoising

Fluorescence microscopy images often exhibit a considerable amount of noise, that can be attributed to the following processes: First, the fluorescence signal arrives at the camera in the form of discrete photons resulting in *shot-noise*, *i.e.* the number of photons incident on a pixel follow a Poisson distribution. Second, the conversion of photons to a digital signal as counted in the camera is subject to *read-out noise*, *thermal noise*, and *dark current*, all of which is specific for the used camera type. For instance the later type of noise contribution is generally smaller for lower frame-rate EM-CCD[9] cameras used in Spinning-Disk microscopy, and slightly larger for sCMOS[10] cameras used in high-speed light-sheet microscopy (H. T. Beier et al. 2014). Fig. 2.7 shows raw images of different sample acquisitions that highlight the different kinds of noise corruption that can be commonly observed. As can bee seen, both types of noise contributions lead to low *signal-noise-ratio* images especially in the common situation of live-imaging developing biological samples, since camera exposure time and laser power have to be reduced to achieve the desired time resolution and to ensure the viability of the living specimen (Icha et al. 2017). Here, sCMOS cameras [11] are – due to their speed (up to 1 GB/s) and large field of view (2048×2048) – now the dominant choice for fast volumetric images (Chen et al. 2014; Royer et al. 2016).

Characterization of fluorescence microscopy noise: The most simplistic model that relates the noisy observation y at a specific image pixel to the true signal x is to assume additive Gaussian noise, *i.e.* y and x are related via

$$y = x + \eta, \quad \eta \sim \mathcal{N}(\mu, \sigma) \tag{2.3}$$

with $\mathcal{N}(\mu, \sigma)$ being a Gaussian (normal) distribution with mean μ and standard deviation σ. A more realistic model that recognizes the Poissonian nature of detected photons is that of Poisson-Gaussian noise (Mäkitalo et al. 2012), *i.e.*

$$y = g \cdot \tilde{x} + \eta, \quad \tilde{x} \sim \mathcal{P}(x), \eta \sim \mathcal{N}(\mu, \sigma) \tag{2.4}$$

where g is a (pixel-dependent) gain, $\tilde{x}$ is Poisson-distributed, and η is the background/read-noise contribution. Note, that this can be transformed into a distribution of unit variance with the generalized Anscombe transform (Starck et al. 1998; Mäkitalo et al. 2012). To validate whether such assumptions are appropriate in our case a light-sheet microscope with a sCMOS camera was used to collected several time-series at a single focal plane at two conditions: *i) dark-frames*,

[9] Electron-Multiplying Charged Coupled Device (Robbins et al. 2003)
[10] Scientific Complementary Metal–Oxide–Semiconductor (Z.-L. Huang et al. 2011)
[11] Such as the Hamamatsu Orca Flash 4.0 v2

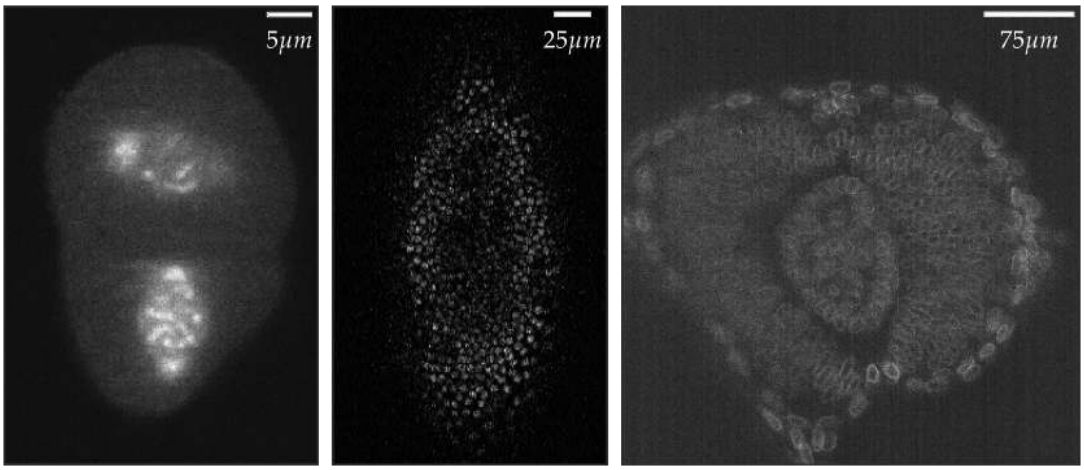


Figure 2.7: Examples of camera noise in live-cell imaging, from left to right: *C. elegans* embryo at 2-cell stage (histone/β-Tubulin labeling/light-sheet microscope), *Tribolium castaneum* embryo (nGFP+Ecad/spinning-disk) and developing *Danio rerio* retina (DRAQ5/light-sheet). Note the different background noise distribution of sCMOS (light-sheet) and EMCCD (spinning-disk) cameras (Raw data from Loic Royer, Akanksha Jain and Mauricio Rocha-Martins).

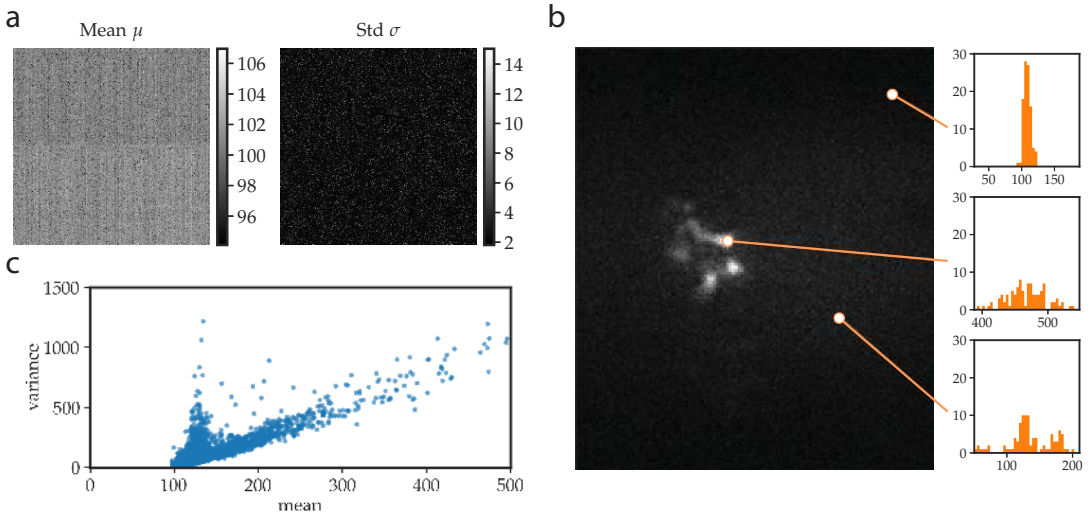


Figure 2.8: Noise characterization of a commonly used sCMOS camera (Hamamatsu Orca Flash 4.0 v2). a) Mean μ and standard deviation σ of acquired dark frames exhibiting great variability throughout the sensor (outliers/hot pixels were clipped in the image to help the visualization). b) Temporal histogram of the recorded intensity at 3 different pixels for histone labeled *C. elegans* nuclei. Shown is a background pixel with predominantly Gaussian noise statistics (top), a foreground pixel exhibiting Poisson-Gaussian noise (middle), and a pixel with low mean and high variance (bottom). c) Scatter plot of pixel mean and variance of b). For each single experiment 100 time-points were acquired per stack, and the camera was water-cooled to prevent hot pixels.

i.e. frames acquired without incident laser illumination and background light, and *ii)* frames acquired with normal laser exposure showing labeled structures of different intensity in the field of view. The mean intensity and standard deviation of dark-frames is shown in Fig. 2.8a, demonstrating that the background intensity and read-noise is variable across the sensor and poorly described by the parameters of a single normal distribution: background pixels of similar mean can exhibit greatly different variance as shown in Fig. 2.8b. Further, if following a Poisson-Gaussian distribution Eq. (2.4), each pixels variance would depend linearly on its mean, $\langle y^2 \rangle - \langle y \rangle^2 = g\langle x \rangle + \mu$, which interestingly is only true for a subset of pixels, as can be inferred from Fig. 2.8c.

Denoising methods: Since camera based imaging systems are ubiquitous within almost all sciences, the field of denoising methods has a long history. The most basic methods encompass filtering methods, where each pixel is modified depending on its local neighborhood, such as median filters (Loupas et al. 1989), low-pass filters, or bilateral filters (Tomasi et al. 1998). Next, variational methods aim to minimize a global energy[12], such as anisotropic diffusion (Perona et al. 1990), total variation denoising (Rudin et al. 1992) or curvature filters (Sbalzarini 2016; Gong et al. 2017). Another class of *patch-based* methods leverage the self-similarity of image structures to adaptively generate averages of similar patches, such non-local-means (NLM) denoising (Buades et al. 2005) and its extension (Boulanger et al. 2010; Carlton et al. 2010), or learned dictionary methods like EPLL (Zoran et al. 2011). Grouping similar patches into blocks and filtering them in a suitable (*e.g.* DCT-) domain is the basis of *collaborative filtering methods*, with the most popular example being BM3D (Dabov et al. 2009) and BM4D (Maggioni et al. 2013), for 2D and 3D images respectively. Finally, machine learning approaches that interpret denoising as a regression task were recently shown to out-perform classical methods. Examples include *shrinkage fields* (U. Schmidt et al. 2014), or *neural networks* (Jain et al. 2008; Burger et al. 2012; Mao et al. 2016).

Suitable methods for real-time applications: To identify suitable methods for their potential in real-time application, I chose a subset of these to be tested on a biological data set with known ground-truth. The methods compared were median filter, bilateral filter, total variation (TV), non-local-means (NLM) and BM3D. All methods were tested on a corpus of 2D fluorescence microscopy images from the Broad Bioimage Benchmark Collection[13] (Ljosa et al. 2012) showing nuclei stained cells (*cf.* Fig. 2.9a). To synthetically create realistic camera noise, we created noise-corrupted images from the (practically noise-free) ground-truth images by first scaling them to an intensity range $[0, 100]$ and adding Poisson noise. Finally, a Gaussian noise background was added with pixel-wise dependent means and variances as measured from real sCMOS dark-frames (*cf.* Fig. 2.9a). For each method, we tuned their parameters (*e.g.* filter size, smoothing factor etc.) on a subset of 72 input images, and finally measured

[12] Typically consisting of a data-term and a regularization term
[13] I used images from the BBBC006 and BBBC008 data set

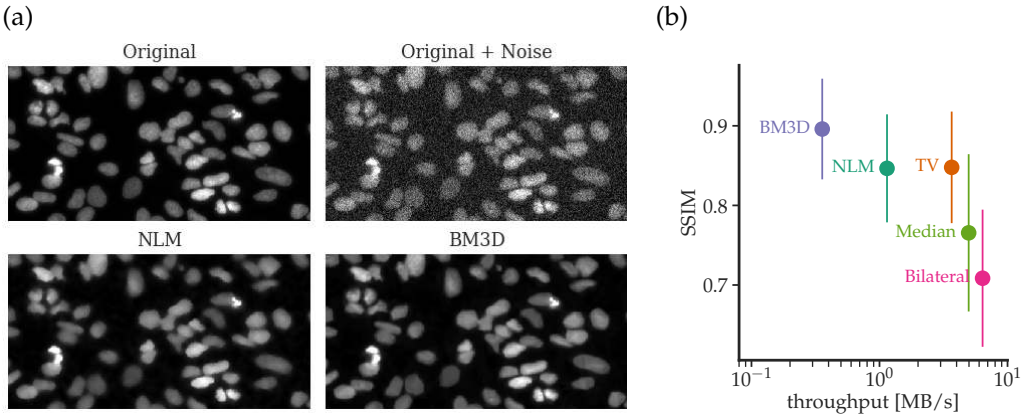


Figure 2.9: Overview of compared denoising methods. a) Example image from the used data set of cell nuclei images, alongside the corrupted version and two denoising results (BM3D and NLM). b) Restoration quality (SSIM, higher is better) *vs*. throughput for compared methods. Error bars designate standard deviation.

the average restoration quality as given by the SSIM[14] and run-time on a hold-out test set of 24 images. All reference implementations were run as single-core CPU processes to allow for fair comparison of their run-time. As can be seen in Fig. 2.9b, best restoration quality was achieved by BM3D (mean SSIM $= 0.90$), followed by NLM and TV (both 0.85), median filter (0.77) and finally bilateral filter (0.71). On the other hand, the achievable throughput is largest for bilateral (6.4 MB/s) and median filter (4.9 MB/s), and lowest for BM3D (0.4 MB/s), showing the general trade-off between restoration quality and run-time. Importantly, all considered methods fall short of the throughput necessary for real-time denoising which is on the order of ≈ 100 MB/s (*e.g.* a 16-bit stack of size $1024 \times 1024 \times 50$ per second).

In the next section, we will show that a GPU implementation can indeed decrease denoising runtime to a level suitable for real-time applications. Specifically we pick non-local-means (NLM) denoising among the considered methods, hypothesizing that this patch based method could benefit the most from the parallelized execution model.

2.4.2 GPU-accelerated Non-Local Means

The basic idea of non-local means (NLM) denoising (Buades et al. 2005) is to consider small patches of the corrupted image and estimate the denoising results as an weighted average of patches that are similar to each other. Let $x \in \mathbb{R}^n$ be the corrupted input image and $y \in \mathbb{R}^n$ the restoration result (both given as a discrete array of intensities), then

$$y_i = \sum_{j \in S} w(P_i, P_j) x_j \tag{2.5}$$

where y_i is the value of a specific pixel i, x_j is the value of pixel j in the input image, P_i and P_j are patches centered at pixel i and j, and S is the subset of the input pixels

[14] Structural similarity index (Z. Wang et al. 2004, higher is better).

that is searched for similar patches (see Fig. 2.10a). The weight between patches $w(P_i, P_j)$ is chosen as

$$w(P_i, P_j) = \frac{1}{Z_i} e^{-\frac{d(P_i,P_j)}{\sigma^2}}, \quad \text{with} \quad d(P_i, P_j) = \left\| P_i - P_j \right\|^2, \quad Z_i = \sum_{j \in S} w(P_i, P_j) \tag{2.6}$$

which is bigger for similar patches and smaller for patches with different intensities. In that way NLM denoising not only considers intensities at a single pixel, but many self-similar patches within the search range in an image. The parameters of the methods are the width p of considered patches[15], the width s of the search space[16], and the Gaussian width σ that controls the amount of exponential weighting and therefore the amount of overall denoising. NLM-denoising in its original form (Eq. (2.5)) has time complexity $\mathcal{O}(Np^d s^d)$ with d being the dimension of the image ($d = 2, 3$) and N the overall number of pixels. Later, Darbon et al. 2008 showed that by using integral images for the weight calculation, this complexity can be reduced to $\mathcal{O}(Ns^d 2^d)$, which serves as the basis of almost all current implementations (Walt et al. 2014).

Efficient GPU implementation: To adapt NLM denoising to the GPU, we use a similar strategy as Darbon et al. 2008, whilst employing texture-memory for the distance and weight calculations to ensure fast neighborhood access patterns. Note that the calculation of integral images is not immediately parallelizable due to the inherently sequential cumulative summation, thus we opted for direct parallelization over neighborhood patches instead. A similar approach has been already taken by L. Li et al. 2013, where the (CUDA) implementation was however not made publicly available. As before, we chose OpenCL as the GPU backend, ensuring its applicability on various platforms.

Runtime comparison: To assess the utility of our GPU implementation for the desired real-time application purposes, we compared the runtime of several implementations: a single-core CPU implementation[17], a multi-threaded CPU implementation[18], and the proposed GPU implementation in OpenCL. Additionally, we measured the runtime of the GPU implementation without memory transfer time between CPU to GPU, to estimate the incurred overhead of host-device transfer. All measurements were performed on a machine with a 40-core Intel Xeon CPU (Broadwell) E5-2630 and a Nvidia Titan X (Pascal) GPU. In Fig. 2.11a we show both the runtime and the speedup achieved by the GPU version for 2D images of increasing size (up to 8192×8192), with a patch size of $p = 7$ and search size of $s = 35$.

[15] *i.e.* patches of size $p \times p$ in 2D, or $p \times p \times p$ in 3D
[16] Same as before, *e.g.* $s \times s$ in 2D
[17] `scikit-image` (Walt et al. 2014)
[18] https://imagej.net/Non_Local_Means_Denoise, ImageJ/Fiji (Schindelin et al. 2012)

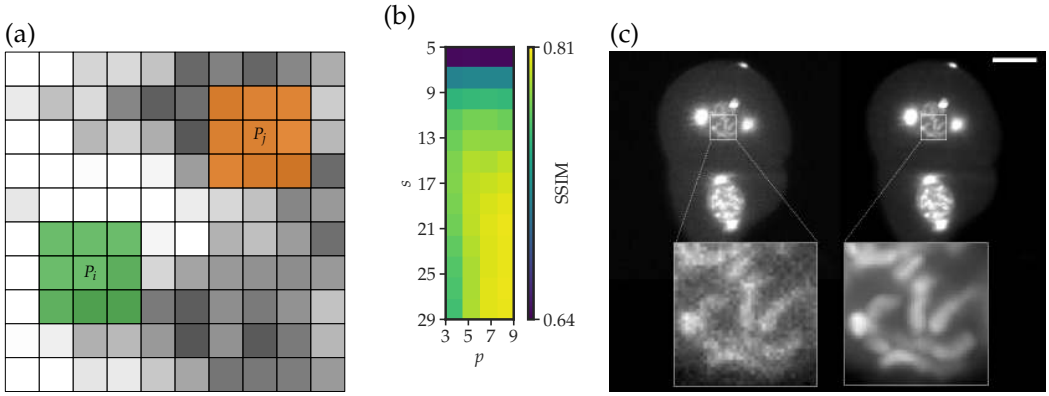


Figure 2.10: a) Schematic overview of non-local-means denoising. b) Restoration quality (SSIM) as function of patch size p and search size s on the BBC test data set. c) Results of NLM-3D(GPU) on light-sheet acquisitions of developing *C. elegans* (Histone/β-Tubulin labeling, Scalebar 10μm).

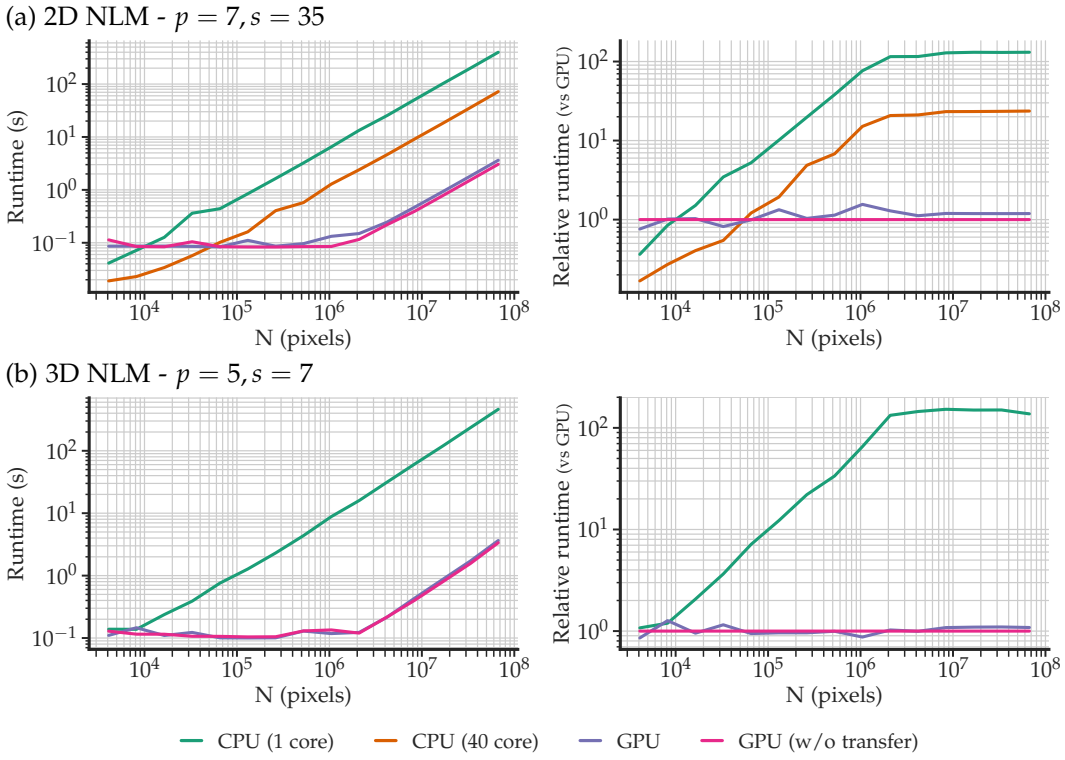


Figure 2.11: Runtime of 2D and 3D NLM implementations on CPU and GPU for a) 2D images and large patches/search-sizes ($p = 7, s = 35$), and b) 3D images and small patches/search-sizes ($p = 3, s = 7$). For both a) and b) total runtime (left) and relative runtime compared to the GPU implementation (right) are shown. Note that for the 3D case a multi-threaded CPU implementation was not available. CPU: Intel Xeon (Broadwell) E5-2630, GPU: Nvidia Titan X CPU.

Expectedly, the GPU version is considerable faster both than the single core CPU version (speedup factor ≈ 102) and the multi-threaded CPU version running on 40 cores (factor ≈ 21) for larger images ($> 10^6$ pixels). Additionally, the host/device memory transfer time played only a negligible role. The same effect can be seen for 3D images (with $p = 5$ and $s = 7$), whose results are shown in Fig. 2.11b. and which indicate a similar increase in performance (speedup ≈ 105). Note that the multi-threaded CPU version only supports 2D images and hence could not be compared with in the 3D case. In all cases, we see the runtime grow linearly with the number of input pixels, as expected. Furthermore, in the case of 3D, the GPU accelerated NLM implementation enables the denoising of a stack of size $1024 \times 1024 \times 64$ in approximately $3s$, which satisfies the requirements of real-time applications. As an example we show in Fig. 2.10c a single time-point of a developing *C. elegans* embryo during the first cell divisions. Here a custom light-sheet microscope was used to acquire stacks of size $320 \times 320 \times 81$ at a rate of 1 Hz. Using the GPU implementation, we were able to denoise these volumes on a common microscope workstation faster than they were acquired, thus enabling real-time denoising (*cf.* Fig. 2.10c).

2.4.3 Convolutions

Many important image processing methods, such as image filtering or deconvolution, incorporate *convolutions* as basic building blocks (Richardson 1972; Lucy 1974; Sbalzarini 2016). Additionally, in microscopy, the image formation process is often described as a convolution of the object signal with the point spread function (PSF) of the microscope. The *convolution* of a *kernel* $h(x)$ with a function $f(x)$ is defined as

$$g(x) = h(x) \otimes f(x) = \int dx'\, h(x - x') f(x) \tag{2.7}$$

which is a bi-linear and symmetric operator of both f and h. Furthermore, with h being fixed, it acts as a shift-invariant filter on functions f. When considered as a linear operator on the real vector space $g \in R^n$, *i.e.* $g_i = \sum_j h_{i-j} f_j$, it is represented by a *Toeplitz matrix*. Assuming circular boundary conditions, the *Fourier convolution theorem* states that the convolution operator becomes diagonal in the Fourier domain, specifically that the Fourier transform of the convolution between h and f is simply the product between the Fourier transform of each of both functions:

$$\mathcal{F}[h \otimes f] = \mathcal{F}[h] \cdot \mathcal{F}[f], \quad \mathcal{F}[f](k) = \int dx\, f(x) e^{-2\pi i k x}. \tag{2.8}$$

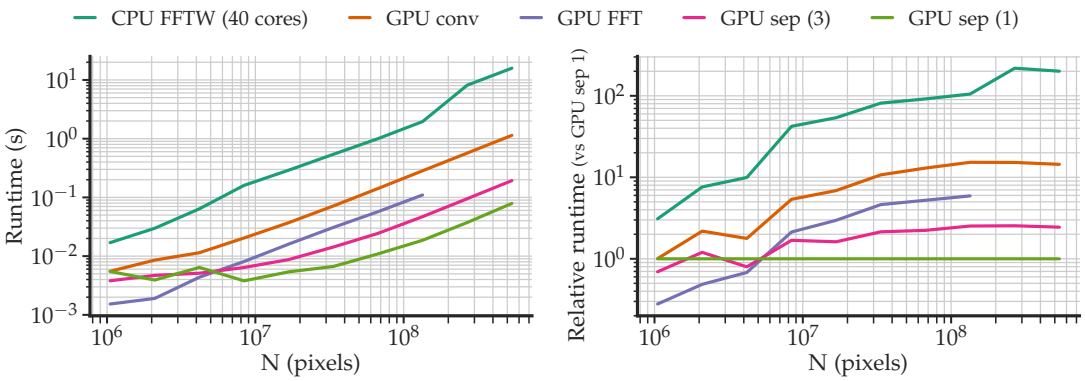
This relation is often used to implement convolutions via the Fast Fourier Transform (FFT), particularly in common deconvolution approaches (Karas et al. 2011; Bruce et al. 2013; Preibisch et al. 2014; Schmid et al. 2015). In Fig. 2.12a we compare the performance of several different 3D convolution implementations for different sizes of input volumes, specifically: FFT based, multi-threaded convolution on the CPU (40 cores, via FFTW), FFT based convolution on the GPU (via OpenCL), and a spatial implementation (*e.g.* directly computing Eq. (2.7), using texture memory). A

kernel of size $7\times7\times11$ was chosen, comparable to the sizes of typical PSFs present in microscopy images, which - due to the sampling being close to the Nyquist limit - typically exhibit a lateral width of approximately $2-5$ pixels. As can be seen in Fig. 2.12b, FFT convolutions can be sped up by a factor of approx. 17 on the GPU compared to 40 CPU cores, allowing the convolution of a kernel of size $7\times7\times11$ with a stack of size $1024\times1024\times512$ in $\approx 1.3\ s$. Computing the same operation in a purely spatial way is roughly a factor 2.6 slower than the GPU FFT-based version. Note that all dimensions of the input image were chosen to be a power of 2, as the performance of the FFT is known to considerably suffer for non power-of-2 dimensions (Steinbach et al. 2017).

2.4.4 Low-Rank (Separable) Approximations

Although FFT based convolutions are the common choice *e.g.* within common deconvolution methods (Bruce et al. 2013; Preibisch et al. 2014; Schmid et al. 2015), they bear a number of disadvantages. First, to be efficient they demand certain input shapes, often power-of-2 or multiples of small prime numbers[19], and the input is therefore enlarged appropriately. Second, the Fourier Convolution Theorem assumes circular boundary conditions, thus the input has to be padded to avoid boundary artefacts. Finally, the convolution kernel used *e.g.* in deconvolution of microscopy images have often small support, corresponding to the size of the PSF typically on the order of only several pixels, and thus lead to extremely sparse convolution matrices. We therefore hypothesize that finding a fast way of implementing convolutions with small kernels in a purely spatial way would offer a more efficient approach. Specifically, we will use the fact that *i)* the blur kernel/PSF encountered

[19] Radix-2 Cooley-Tuckey (Cooley et al. 1965), or Bluestein's algorithm (Bluestein 1970) respectively



(a) Runtime (b) Relative runtime

Figure 2.12: Performance comparison of 3D real-to-real convolutions when implemented with different methods. a) Runtime for compared methods and different 3D input sizes. b) Relative runtime compared to GPU sep (1) shows the relative performance gains when using a separable kernel. The same convolution kernel of size $7\times7\times11$ was used for all methods. All image dimension were power-of-two. CPU: Intel Xeon (Broadwell) E5-2630, GPU: Nvidia Titan X CPU. FFTW version 3.3.4 real-to-real, inplace, with `FFTW_MEASURE` plan creation. GPU methods were implemented in Python/OpenCL.

in microscopy can be approximated by a series of separable kernels, *i.e.* by a *low-rank decomposition*, and *ii) separable convolution* can be efficiently implemented on the GPU (Rigamonti et al. 2013). A 3-dimensional kernel $h(x, y, z)$ admits a tensor decomposition of rank S if it is expressible as

$$h(x, y, z) = \sum_{s=1}^{S} u^s(x)v^s(y)w^s(z) = \sum_{s=1}^{S} h^s(x, y, z) \tag{2.9}$$

or equivalently for a discrete kernel h_{ijk}

$$h_{ijk} = \sum_{s=1}^{S} u_i^s v_j^s w_k^s = \sum_{s=1}^{S} h_{ijk}^s. \tag{2.10}$$

This decomposition is also known as canonical polyadic decomposition (CDP) or CanDecomp/ParaFac[20] decomposition (Kolda et al. 2009). In the same way, for any general kernel h_{ijk} and order S one can define its *rank-S approximation* $\tilde{h}_{ijk}^S$ as the tensor of rank S that is closest to h_{ijk} according to some tensor norm[21]. Note that this decomposition readily generalizes to n-dimensional tensors, and coincides with the usual SVD decomposition in the case of 2-dimensional matrices. Importantly, being able to express a kernel of size $K \times K \times K$ with its low-rank approximation $\tilde{h}_{ijk}^S$ reduces the time complexity of a convolution from $\mathcal{O}(K^3)$ to $\mathcal{O}(3SK)$.The performance gain resulting from using separable convolutions can be seen from in Fig. 2.12a, where we (additionally to the methods before) measured the time to perform GPU-based convolutions with a kernel of $7 \times 7 \times 11$ using its low-rank approximations (with $S = 1$ and $S = 3$). As can be seen in Fig. 2.12b, this approach is $6x$ and $2.5x$ faster than the next fastest method (GPU-FFT) when decomposing the kernel with rank $S = 1$ or $S = 3$ respectively. Additionally to being faster for small kernels, performing the convolution in the spatial domain does not impose any restriction on the input dimensions (and the necessary padding) and does not enforce circular boundary conditions.

Low-rank Decomposition of Microscopy PSFs

After having demonstrated the benefits of performing convolutions via the low-rank approximation of kernels, we now investigate the question how well such representations are appropriate for the special case of PSFs arising in the analysis of microscopy images. Specifically, we consider theoretical PSFs $h(x, y, z)$ of widefield and light-sheet microscopes, which are given by

$$h(x, y, z) = h_{illum}(x, y, z) \cdot h_{detect}(x, y, z) \tag{2.11}$$

as the product of the detection and illumination PSF (with $h_{illum}(x, y, z) = 1$ for the widefield case). To that end, we generate theoretical PSFs $h(x, y, z)$ according

[20] CANonical DECOMPosition (Carroll et al. 1970), PARAllel FACtors (Harshman 1970)
[21] *e.g.* the Frobenius (L_2) norm.

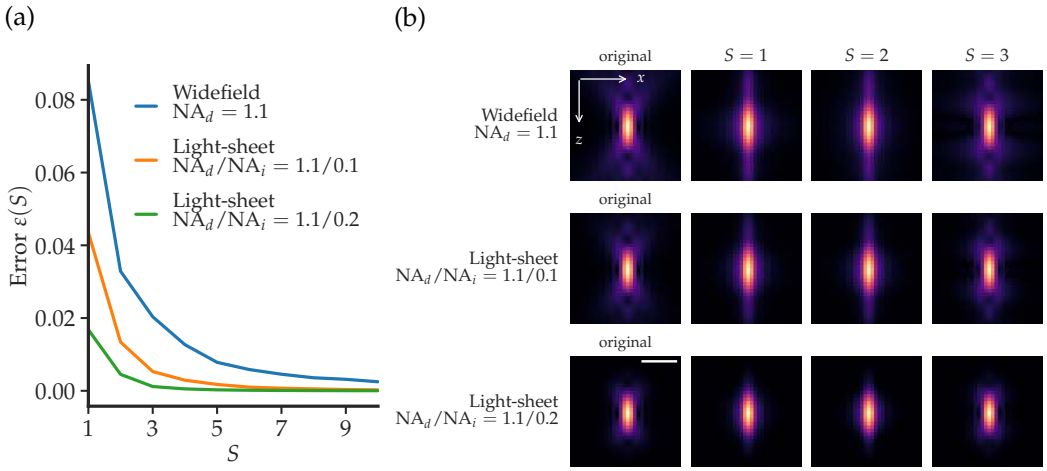


Figure 2.13: Low-rank approximation of microscopy PSFs. a) Normalized overlap error ε as a function of the approximation order S for three different theoretical PSFs, discretized on a grid with $\Delta x = 0.1\mu m$ and assumed wavelength $\lambda = 0.5\mu m$ (using the same for excitation/emission to keep things simple). See main text for further details. b) Axial (xz) slices of each original PSF and its low-rank approximation of increasing order S. Scalebar is $1\mu m$ and $\gamma = 0.5$ was applied to increase visibility of smaller features.

to Eq. (2.11) on a discrete grid with $\Delta x = 0.1\mu m$ and different detection (NA_d) and illumination (NA_i) numerical apertures. For each PSF, we calculate the series of low-rank approximations $h^S(x,y,z)$ of order S and compute the approximation error $\varepsilon(S)$ as the deficiency of the normalized overlap integral:

$$\varepsilon(S) = 1 - \frac{\langle h, h^S \rangle}{\langle h, h \rangle^{1/2} \langle h^S, h^S \rangle^{1/2}}, \quad \text{with } \langle f, g \rangle = \int dx dy dz f(x,y,z) g(x,y,z). \tag{2.12}$$

In Fig. 2.13a the resulting approximation error is shown for *i)* a widefield PSF with $NA_d = 1.1$, *ii)* a light-sheet PSF with $NA_d = 1.1$ and $NA_i = 0.1$, and *iii)* a light-sheet PSF with $NA_d = 1.1$ and $NA_i = 0.2$. As expected, the error decreases with increasing approximation order S in all cases. Importantly, for any given order S the rank-S approximation of both light-sheet PSFs exhibit substantially less error compared to the widefield PSF, seemingly caused by the axial confinement of the PSF due to the light-sheet illumination. We note, that for both light-sheet PSFs already order $S = 3$ results in well approximating h^S. This is further corroborated by Fig. 2.13b, where xz-slices are depicted for each PSF and different approximation orders: At $S = 3$, the low-rank approximations are virtually indistinguishable from the original PSF.

Fast Deconvolution with Low-Rank Approximations

As an important application of the previous section we now consider a classical deconvolution algorithm, specifically *Lucy-Richardson deconvolution* (Richardson 1972; Lucy 1974): Given a blurry input image f and a PSF h, this method computes the

restored original image g via the following iterative update steps:

$$g_{n+1} = g_n \cdot \left[h^\dagger \otimes \frac{f}{g_n \otimes h}\right], \tag{2.13}$$

which can be derived as a fixed point equation when maximizing the likelihood under the assumption of Poisson noise. Here, $h^\dagger(x,y,z) = h(-x,-y,-z)$ is simply the mirrored PSF. Despite - or perhaps because - of its simplicity, Lucy-Richardson deconvolution is still the basis for currently employed deconvolution methods (Preibisch et al. 2014; Schmid et al. 2015). As is apparent from Eq. (2.13), the computational most intensive part of the method are convolutions with the PSF (and its mirrored version). We now investigate the implications of approximating these convolutions with their low-rank decomposition, as outlined in the last section. To that end, we generated synthetic ground-truth (hollow cylinders), applied blur with a realistic light-sheet PSF, and added Poisson noise (Fig. 2.14a). We then computed the result of $N = 50$ iterations of Eq. (2.13), while *i)* using a low-rank decomposition of h and $h^\dagger$ with $S = 1$, *ii)* the same with $S = 3$, and *iii)* with using the full PSF in the convolutions during the iterative steps. As can be seen in Fig. 2.14a, the results of using only a rank-1 approximation ($S = 1$) is - particularly in the axial view - visibly different from using a rank-3 approximation or the full PSF, which appear visually comparable. This observation is confirmed when computing the error of the reconstruction *vs*. ground-truth for the different PSF approximations as depicted in Fig. 2.14b.

Here, both low rank approximations with $S = 2$ and $S = 3$ exhibit approximately the same error during iterations as when using the full PSF, whereas using only the $S = 1$ approximation results in a substantially larger error. This indicates, that a $S = 3$ decomposition is likely to be sufficient for typical deconvolution tasks. The performance benefits of this approach are depicted in Fig. 2.14c, where we compute the time to perform 50 iterations of the Lucy-Richardson algorithm on stacks of different sizes, showing that *e.g*. deconvolution of a volume of size $1024 \times 1024 \times 256$ can be performed in roughly 10 seconds. This shows that the low-rank decomposition of PSFs enables to deconvolve even large volumes within times compatible with real-time applications. We note, that highly asymmetric PSFs might be less well approximated by a low-rank decomposition and thus the outlined approach might not be effective for *e.g*. highly aberrated and tilted PSFs such as they appear deep inside aberrating tissues. Additional future work will be needed to investigate such situations.

2.4.5 Image Sharpness Filter

As a final application, we demonstrate how a GPU based *image sharpness filter* provides a fast and simple way to measure microscopy image sharpness in real-time during acquisition. Doing so is important for optimizing imaging parameters during microscope calibration or for automatic focusing during acquisition (Royer et al.

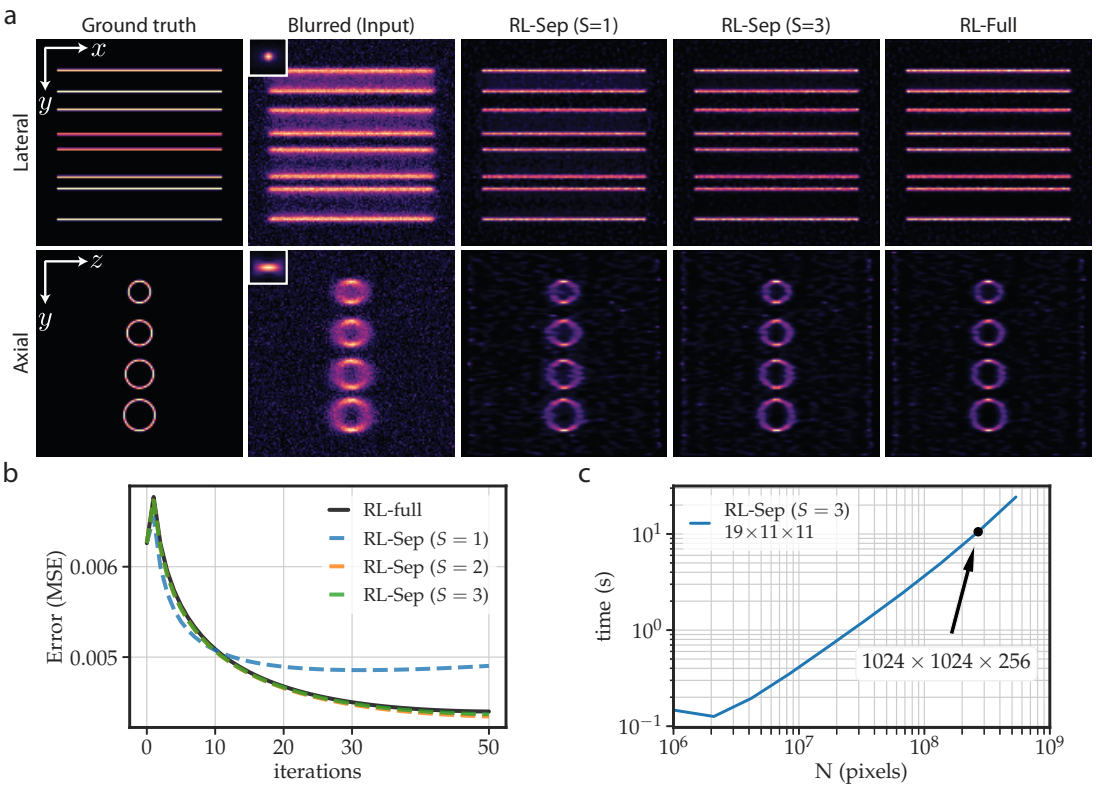


Figure 2.14: Fast deconvolution with low-rank approximations. a) Richardson-Lucy deconvolution of a synthetic stack of hollow cylinders (Ground-truth) blurred by a light-sheet PSF with $NA_d = 0.8$ and $NA_i = 0.15$ (Blurred). Shown are deconvolution results after 50 iterations when using a $S = 1$ and $S = 3$ low rank approximation, and the full, exact PSF. b) Reconstruction error for same experiment as a function of number of iterations. c) Runtime for Richardson-Lucy deconvolution with 50 iterations and $S = 3$ for 3D input stacks of different size.

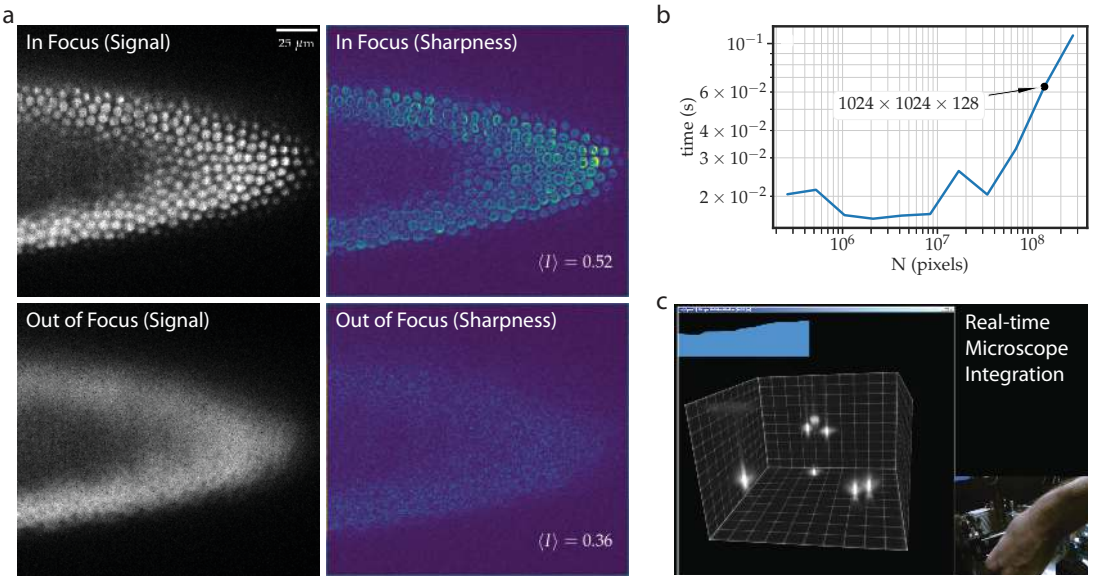


Figure 2.15: Fast sharpness filter. a) Input slice (left) and filter response and normalized mean sharpness measure $\langle I \rangle$ (right) for in-focus (top) and out-of-focus (bottom) image of the same *Drosophila* slice. b) Runtime (without CPU/GPU transfer) of sharpness estimation for 3D input stacks of different sizes. c) Example of microscope integration. Real-time feedback during manual PSF optimization via instant display of the sharpness measure over time (blue chart, left corner). (Panel c reproduced with permission from Royer et al. 2015. Copyright 2015, Springer Nature)

2016). Motivated by exhaustive search from (Royer et al. 2016) we use the popular Tenengrad filter (Krotkov 1988) and compute the following image sharpness measure I as the normalized vector norm

$$\langle I \rangle = \frac{1}{|vol|} \sum_{vol} I(x,y,z) \quad , I(x,y,z) = \sqrt{G_x^2 + G_y^2 + G_z^2} \tag{2.14}$$

on the volumetric intensity $f(x,y,z)$. Here, the values G_i are the Sobel filter responses along the respective dimension:

$$G_x(x,y,z) = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}_x \otimes \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}_y \otimes \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}_x f(x,y,z) \tag{2.15}$$

As the Sobel filters are separable, they can again be efficient calculated and reduced to I on the GPU using several successive convolution and reduction passes employing shared memory. In Fig. 2.15a we depict a slice through *Drosophila* stack, once acquired with the detection object being in focus (upper), the other while being voluntarily defocused (lower). As can be seen, the Tenengrad responses as defined above are able to highlight the difference and with the focus measure being substantially higher for the in-focus slice compared to its out-of-focus counterpart ($\langle I \rangle = 0.52$ *vs.* $\langle I \rangle = 0.36$). Importantly, the separability of Eq. (2.15) again results in run-times compatible with real-time application: the full calculation on a stack of size $1024 \times 1024 \times 128$ takes about 70ms (*cf.* Fig. 2.15b). It can thus be calculated during *e.g.* bead calibration, where the microscope is adjusted to produce well-formed microscope PSFs, and displayed in real-time as the different imaging parameters are manually tuned (*cf.* Fig. 2.15c & Supp. Video 4).

Video 4

2.5 Implementations and Availability

All methods describe in this section have been implemented and been made available as open-source software with a strong emphasize to be readily accessible by microscope operators and ease of integration into existing microscope controlling software (Supp. Video 1).

Video 1

ClearVolume

Specifically, the visualization and most of the processing methods are part of the software suite *ClearVolume*, which constituted joint work with Loic Royer (lead author), Ulrik Günther, and Florian Jug (Royer et al. 2015). *ClearVolume* was designed to augment custom-built light sheet microscope with real-time visualization and processing capabilities, and provides interfaces to all major programming languages/frameworks prevalent in microscopy: Java, C/C++, Python, LabVIEW, and OpenSPIM/μManager. Next to multi-view and multi-color visualization of the acquired volumes, it provides real-time GPU-based image processing modules, such

Video 3

as denoising and image sharpness estimation (Supp. Video 3). Finally, it enables remote live streaming of 3D data across the network via a server-client architecture (via *pyclearvolume*). This allows, for example, in the case of long-term developmental imaging to remotely inspect the health of the embryo and the quality of the recorded volumes. In Fig. 2.16 we provide some usage examples of *ClearVolume* for live visualization of *C.elegans* cell divisions (left) and real-time denoising and deconvolution of HeLa-cells (right). The code and documentation can be found at https://clearvolume.github.io/.

Spimagine and gputools

Apart from being able to integrate real-time visualization and processing into the microscope software as done by *ClearVolume*, it is valuable to expose the same functionality to an interactive and dynamic language and thus create a useful research tool for rapid prototyping. We thus developed two packages *spimagine* and *gputools* that aim to provide these capabilities within Python[22] as the dynamic language of choice for scientific image processing. *spimagine* is designed to interactively visualize and process time-lapse volumetric data, providing volumetric and isosurface rendering of 3D+t data both via a standalone viewer as well as interactively[23]. Animations can be easily created via key frame editing and exported to image sequences (*cf.* Fig. 2.17), thus providing a powerful way to create movies of time-lapse datasets with complex transitions, viewpoint changes, and varying zoom levels. On the other hand, *gputools* provides implementations of mentioned denoising and fast convolution algorithms. As before, all processing methods are provided as fast GPU based implementations via pyopencl and can additionally incorporated into *spimagine* as *processing plugins* that make it possible to study their impact on the data interactively (Supp. Video 2). Code and documentation can be found at https://github.com/

Video 2

[22] https://www.python.org/

[23] All renderings of 3D data in this thesis were done with *spimagine*.

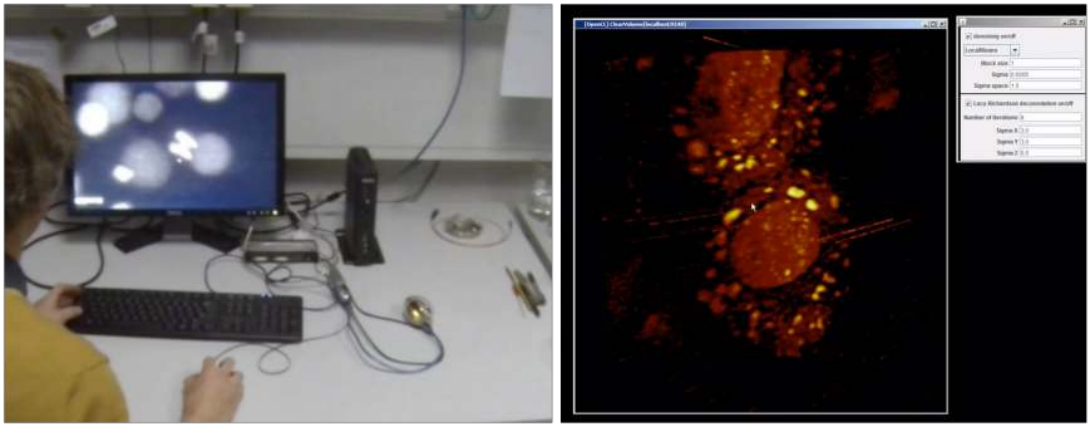


Figure 2.16: Real-time visualization and processing with *ClearVolume*. During acquisition, the volume can be rendered and interactively inspected (left, *C. elegans* embryo) and low-level processing such as denoising and deconvolution can be applied (right, *HeLa* cells). (Reproduced with permission from Royer et al. 2015. Copyright 2015, Springer Nature)

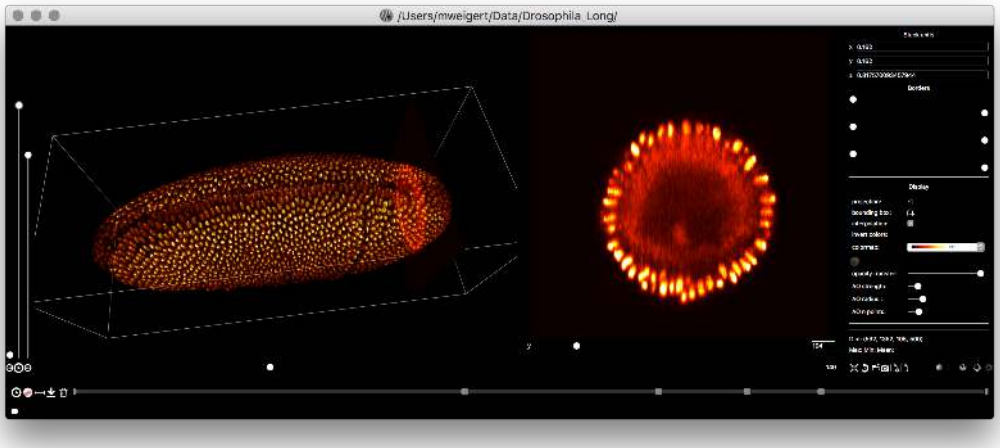


Figure 2.17: Real-time visualization and processing with *spimagine*. Shown is the keyframe-animation panel.

maweigert/spimagine and https://github.com/maweigert/gputools.

2.6 Summary

In this chapter, we addressed the question how to visualize, analyze and process the volumetric data arising in fluorescence microscopy within the time constraints of the typical acquisition rate of typical microscopy experiments. To that end, we described fast rendering pipelines for large microscopy volumes including a novel *low-discrepancy sampling* strategy. We further demonstrated that the design of GPU-parallelizable low-level image restoration methods, including the use of *low-rank PSF decompositions*, allows acquisition-rate denoising and deconvolution of microscopy volumes, and provides for image sharpness measures that yield instant image quality feedback on the microscope. Most described methods are provided as open-source packages (*ClearVolume*, *spimagine*, and *gputools*) which already have found many users within the research community.

3 Fast Multiplexed Models of Light Propagation in Biological Tissues

Overview

In the previous chapter we discussed several sources of image quality loss that are induced by the imperfection of the microscopy detection process such as camera noise and the signal blur inflicted by the limited optical resolution of the microscope. A different, yet equally significant, contribution to image degradation in whole tissue microscopy is caused by the interaction of light with the biological sample itself. In light-sheet microscopy, for example, the illumination sheet experiences several optical perturbations while traversing the specimen: it is bent into out-of-focus planes via refraction at layered structures of the sample (*e.g.* the eggshell of an embryo, Royer et al. 2016), its thickness is increased by diffusive scattering, and it is partially attenuated or even blocked by absorbing tissue components (Huisken et al. 2007). This results in a loss of optical sectioning, subsequent decrease in axial resolution, shadowing and striping artifacts. In the same way, the light emitted from excited fluorophores towards the camera is refracted and scattered by sample parts between the focal plane and the detection objective, leading to

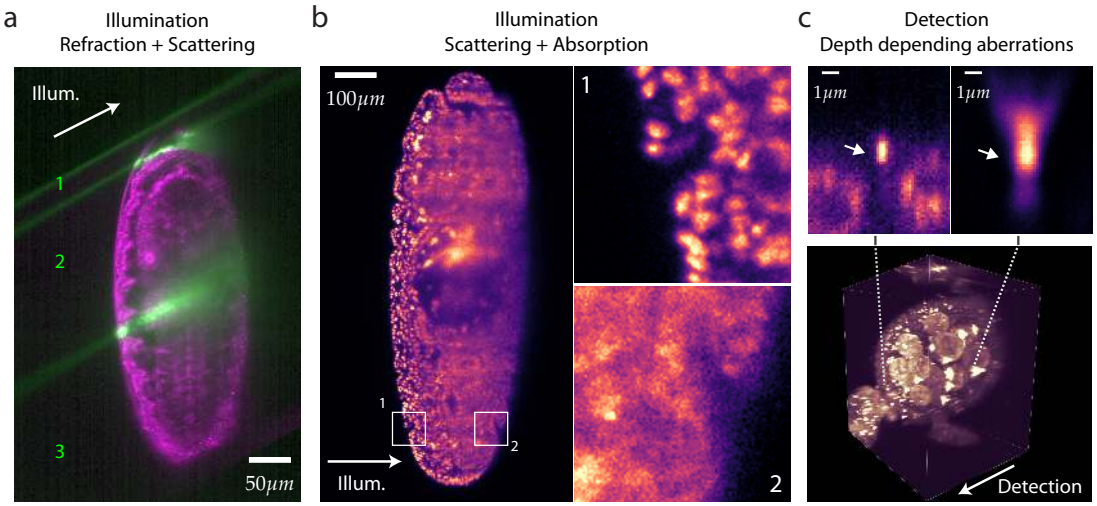


Figure 3.1: Adverse effects of light tissue interactions: a) Intensity of single illumination beams (green) of a DSLM light-sheet microscope when traversing the midsection of a histone labeled *Drosophila* embryo (magenta). Whereas outside the sample the beam assumes a pencil like Gaussian profile (1), it gets refracted and deflected at the outer embryo layers (2), and diffusively scattered by the embryo interior (3). Image is shown with $\gamma = 0.5$ to highlight low intensity parts. b) Light-sheet image of the middle section of a histone labeled *Drosophila* embryo. Increase in scattering and absorption of the light sheet along the propagation direction leads to severe image blur and striping artifacts on the right side of the embryo. c) Depth depending aberrations in *Drosophila* ovary, WGA/mCherry membrane staining. Embedded sub-diffraction beads at different depth result in a compact, diffraction limited PSF close to the camera (left upper) and highly aberrated PSF far from the camera (right upper). Data from Helena Jambor.

aberrations, broadening of the point spread function (PSF) and an image resolution below the theoretically possible diffraction limit (Ji 2017). Examples of these distortions are depicted in Fig. 3.1, where we show the refraction and bending of single beams within a *Drosophila* embryo (Fig. 3.1a), the absorption of a light sheet and the induced asymmetric image blur (Fig. 3.1b), and deformed PSFs due to spatially changing aberrations within the sample (Fig. 3.1c). The fundamental reason for these distortions is the spatial inhomogeneity of the optical properties of tissue: The propagation speed of light in media is determined by its refractive index distribution and inhomogeneities result in aberrations, *i.e.* deviations of the optical wavefront from its otherwise undisturbed (*e.g.* spherical) reference. Whereas for fixed samples these effects can be mitigated via *sample clearing*[1], *i.e.* the chemical removal of highly scattering cell constituents (*e.g.* lipid compartments), such an approach is impossible for live-cell imaging due to the toxicity of the process. Consequently, without additional optical or computational tools large sample microscopy is limited to imaging depths of $\sim 150\mu m$ in zebrafish (Liu et al. 2018) or $\sim 50\mu m$ in Drosophila embryos (Azucena et al. 2010). Recent optical methods to circumvent these limitations include the employment of non-diffracting illumination modes such as Bessel beams (Fahrbach et al. 2010; Planchon et al. 2011; Gao et al. 2014), the use of infrared illumination wavelengths that exhibit less scattering in biological tissues as utilized

[1] Popular clearing protocols are 3Disco (Ertürk et al. 2012) or Clarity (Chung et al. 2013)

by two-photon microscopy (Helmchen et al. 2005; Truong et al. 2011) or by employing near infra-red dyes (X. Zhang et al. 2012). A purely optical approach is to actively alter the detection wavefront via *adaptive optics* (Kubby 2013; Booth 2014). Here, the wavefront aberrations at different points within the sample are first measured[2] and then corrected using, for instance, a deformable mirror (Liu et al. 2018). Nearly diffraction limited imaging at a depth of $400 - 700\mu m$ has been demonstrated, albeit at the cost of decreased imaging speed and greater optical setup complexity (K. Wang et al. 2015; Ji 2017). Another related approach is *wavefront-shaping* (Vellekoop et al. 2007), which allows to create a focus behind even extremely scattering media (*e.g.* layers of opaque titanium white paint) by precise conjugation of the speckle wavefront via high resolution spatial light modulators (SLM, Vellekoop et al. 2010).

Apart from microscopy, a proper understanding of light-tissue interaction is important for the study of the biological role played by the optical properties of tissues. An important example is the mammalian eye. Here, due to the inverted architecture of the retina, incident light has to pass through several layers of refracting cell nuclei *before* it reaches the photoreceptors (Solovei et al. 2009). As the severity of light scattering within these layers could influence the visual sensitivity especially in low light conditions it has been hypothesized that the optical properties of these layers are evolutionary adapted in specific species (*e.g.* nocturnal animals) to minimize these adverse effects (Kreysing et al. 2010). Quantitatively understanding the consequences of optical changes on the scattering properties is evidently important for validating such biological propositions.

Given the importance of understanding how light travels through complex tissue for imaging applications in microscopy, study of optical distortions in scattering samples, and for biological questions regarding the optical properties of mammalian eye, we seek in this chapter to investigate whether a *fast yet accurate computational model of light propagation in biological tissues* can help to understand these processes. This is motivated by several reasons: Faithfully simulating the wave-optical imaging process in microscopy would constitute a powerful generative (forward) model of microscopy necessary for advanced (*e.g.* machine learning based) image restoration and segmentation methods, potentially allowing incorporation of optical image simulation into the microscope acquisition strategies. Next, it would allow to computationally explore efficient acquisition strategies for adaptive optics (K. Wang et al. 2014; Judkewitz et al. 2015). Finally, it would enable the *in-silico* validation of biological hypothesis that are related to optical properties of *e.g.* retinal tissue as described above. The desired simulation method has to meet several requirements to be practical for whole tissue imaging and propagation simulations through large samples: *i)* It has to be compatible with the typically weakly scattering yet highly in-homogeneous distributions of refractive indices in biological tissue (Choi et al.

[2] *E.g.* by inducing fluorescent guide stars and using a wavefront sensor (K. Wang et al. 2014)

2007), *ii)* it has to be applicable for the large computational geometries necessary for whole tissue simulations ($\sim 500 \mu m^3$), and *iii)* it has to perform fast enough to allow rapid experimentation with optical parameters.

Predicting light-tissue interactions is demanding when leaving the regime of single scattering (Choi et al. 2007) or strictly diffusive transport (C. Zhu et al. 2013), which is the case in geometries of highly clustered cells. Despite significant computational advances (Osnabrugge et al. 2016), generally applicable solutions (Oskooi et al. 2010) remain computationally costly, effectively prohibiting the simulation of image formation in microscopy or light propagation through realistically large sample geometries. Although biological samples are composed of predominantly forward scattering tissue (Jacques 2013; Judkewitz et al. 2015), individual PSF calculations still require multiple seconds (Glaser et al. 2016) on the whole tissue scale. As aberrations deep inside tissues are unique for virtually each point in the sample (Choi et al. 2007; Judkewitz et al. 2015) a realistically large biological specimen, *i.e.* an embryo with a volume $\sim 100 \mu m^3$ would therefore require $10^5 - 10^6$ volumetric PSF calculations in order to faithfully mimic the wave-optical imaging process, which with current methods would take several weeks. As a consequence, attempts to simulate microscopic imaging have been limited to ray optics (Uddin et al. 2011; Young et al. 2018) or convolution with a constant PSF (Venkataramani et al. 2016), approaches that do not reflect the wave-optical nature of light interaction with the optically heterogeneous biological samples.

In the following, we will demonstrate a fast GPU implementation of a classical beam propagation method that in combination with judicious multiplexing of the aberrations map inside of biological samples enables to go beyond these limitations and provide an efficient simulation platform for wave-optical *in-silico* experiments in microscopy. The chapter is organized as follows: In Section 3.1 we first discuss related work, after which we describe the used propagation method, its fast implementation on the GPU, introduce the PSF multiplexing scheme, and compare its performance with competing methods in Section 3.2. We validate our method by considering a multitude of analytical and experimental examples relevant for biological tissue imaging in Section 3.3. We next demonstrate in Section 3.4.1 that this enables to simulate the whole wave-optical image formation process in light-sheet microscopy, including the spatially varying degradations caused by the optical inhomogeneity of the sample. In Section 3.4.2 we investigate correlations of the scattered wave-front when translating the focus across the sample (*memory effect*) and propose a novel focusing scheme based on *in-silico* experiments, that we call *extended memory effect*. Finally we consider the scattering properties of the inverted architecture of the outer nuclear layer of the mouse retina in Section 3.4.3. Here, by simulating for the first time the light propagation through realistically large tissue models extracted from microscopy, we show that the distribution of nuclear DNA in retinal mouse

cells entails a reduction in side scattering and therefore increase of visual sensitivity, as has been hypothesized earlier.

Contributions

1. Investigation and validation of a GPU-based beam propagation method for simulation of light propagation in biological, weakly scattering tissue.
2. Devising a novel multiplexing scheme for parallel PSF calculations inside of scattering media.
3. Providing the first rigorous simulations of wave-optical image formation for light-sheet microscopes.
4. Identifying a new correlative effect between scattering components in tissue (*extended memory effect*).
5. Investigating *in-silico* the optical properties of the inverted nuclei architecture in mouse retina.

Substantial parts of this chapter are published in Weigert et al. 2018b.

3.1 Related Work

Due to its fundamental importance, the simulation of light propagation through general refractive index maps has been multifariously investigated. As is common, a trade-off exists between performance and accuracy since certain refractive index distributions allow for approximations that enable faster methods. In the following, we give examples of these in ascending order of computational complexity and accuracy.

For spherical symmetric particles, such as spheres or core-shell geometries, *Mie theory* allows the analytic calculation of the field as an asymptotic series (Bohren et al. 2008), with open source implementations provided *e.g.* by GmmField (Ringler 2008). When wave-optical effects like diffraction or coherent interference can be ignored, optical simulations can be done efficiently (and in parallel) via geometric *ray tracing* (O'shea et al. 1985). For light-sheet microscopy, ray tracing has been for example used to restore uneven image illumination (Uddin et al. 2011), or to simulate the image formation process (Abdellah et al. 2015). Furthermore, ray tracing allows to calculate geometric aberrations *e.g.* due to refractive index mismatch (Booth et al. 1998) or large spherical interfaces (Schwertner et al. 2004). The *scalar beam propagation* method (BPM) is a classical algorithm to simulate the light propagation for more general refractive fields and entails a scalar approximation of the vectorial wave field valid for moderate inclination of wave-fronts, small refractive index differences, and non-polarizing material (Feit et al. 1978; Van Roey et al. 1981). Scalar BPM and its vectorial generalization (W. Huang et al. 1993) have seen many applications *e.g.* in fiber optics (Kawano et al. 2001) or in the investigation of scattering properties of Bessel beams through biological cells (Fahrbach et al. 2010). There is, however, no general available implementation that is fast and flexible enough to

be used for the applications above. Going beyond the scalar approximation, several alternative approaches such as the *vector wave propagation* method (Brenner et al. 1993; Fertig et al. 2010) or *convergent Born series* (Osnabrugge et al. 2016) have been introduced that bypass the requirements for low refractive index contrasts, at the cost of computational complexity and runtime. Recently, S. Schmidt et al. 2016 demonstrated that for refractive fields consisting of only a small number of different discrete refractive phases an accurate yet still efficient wave propagation method can be derived. Lastly, *finite difference time domain* (FDTD) methods (Yee 1966) allow for the most general treatment of light propagation by numerically solving the time-dependent Maxwell's equation via finite-differences, with MEEP (Oskooi et al. 2010) being a popular open-source implementations. However, FDTD is computationally extremely costly, effectively prohibiting the simulation of light propagation for realistically large geometries. Hence, we sought to implement a fast, yet accurate light propagation method that is able to address these shortcomings of all currently available methods.

3.2 Method

The simulation of light propagation through tissue amounts to solving for the electrical field $\vec{E}(x, y, z)$ given a refractive index distribution $n(x, y, z)$ and given boundary conditions. This in general requires the numerical treatment of the time dependent vectorial *Maxwell's equations* (Born et al. 1999). For monochromatic (wavelength λ) illumination and ignoring polarization effects however, a simpler description in terms of a complex scalar field $u(x, y, z)$ becomes applicable, and the problem reduces to solving the scalar *Helmholtz equation* (Born et al. 1999):

$$\Delta u(\mathbf{r}) + n(\mathbf{r})^2 k_0^2 u(\mathbf{r}) = 0, \qquad k_0 = \frac{2\pi}{\lambda} \tag{3.1}$$

Assuming the field is propagating along the optical axis z - and thus ignoring back scattering - Eq. (3.1) can be solved in the spectral domain by applying the Fourier transform F_{xy} with respect to the coordinates orthogonal to the propagation direction:

$$\mathcal{F}_{xy}[u](k_x, k_y, z) = \frac{1}{2\pi} \int dx dy \, u(x, y, z) e^{-i(k_x x + k_y y)} \tag{3.2}$$

$$u(x, y, z) = \frac{1}{2\pi} \int dk_x dk_y \; \mathcal{F}_{xy}[u](k_x, k_y, z) e^{i(k_x x + k_y y)} \tag{3.3}$$

$\mathcal{F}_{xy}[u]$ represents $u(x, y, z)$ as a superposition of plane waves with different angles and is therefore called the *angular spectrum* at position z (Goodman 1996). The field at position $z + \Delta z$ can now be found by propagating each plane wave component in the angular spectrum and transforming back into the spatial domain

$$u(x, y, z + \Delta z) = \frac{1}{2\pi} \int dk_x dk_y \; \mathcal{F}_{xy}[u] H(x, y, z, k_x, k_y) e^{i(k_x x + k_y y)} \tag{3.4}$$

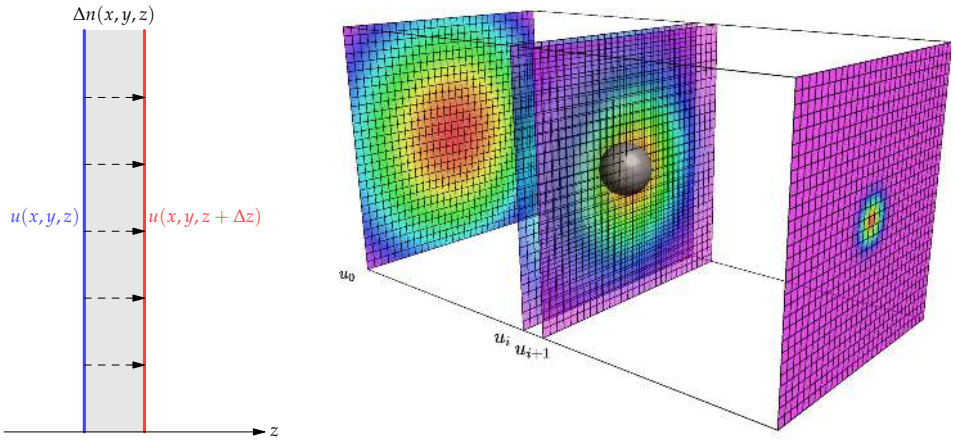


Figure 3.2: Principle of scalar beam propagation: each complex field $u(x,y,z)$ is first propagated freely to the next plane $z+\Delta z$ after which a space dependent phase shift $\Delta n(x,y,z)$ is applied. (Reproduced with permission from Weigert et al. 2018b, Creative Commons license CC BY 4.0)

where

$$H = e^{i\Delta z\sqrt{n(x,y,z)^2k_0^2-k_x^2-k_y^2}} \tag{3.5}$$

is the *propagator* in the Fourier domain (*cf.* Fig. 3.2). In a further approximation the refractive index is assumed to be a small variation around a constant $n(x,y,z) = n_0(z) + \Delta n(x,y,z)$ so that the final approximation gives the *scalar beam propagation*:

$$u(x,y,z+\Delta z) \approx \mathcal{F}_{xy}^{-1}\left[\,\mathcal{F}_{xy}[u(x,y,z)]e^{i\Delta z\sqrt{n_0(z)^2k_0^2-k_x^2-k_y^2}}\right]e^{i\Delta z\Delta n k_0} \tag{3.6}$$

which can be efficiently implemented with the Fast-Fourier-Transform (FFT). It has to be noted, that the assumption of low refractive index contrast is justified for most biological tissues, where the refractive index of most cellular compartments lies between $n = 1.33 - 1.46$ (Johnsen et al. 1999; Choi et al. 2007), often embedded in water $n = 1.33$. Furthermore, *back-scattering* of light, *e.g.* on the way from the specimen to the lens, is entirely neglected with this approach. This is, however, justified for simulation of the image formation process in tissues, as *i)* this light would only contribute to the final image when changing direction a second time, and *ii)* adaptive optics aberration correction requires the forward scattered photons only. Furthermore, the computational geometry has to be discretized with a sufficiently small pixel-size, typically $\Delta x \sim \lambda/5$, to avoid aliasing artifacts. Crucial for the performance of BPM is an appropriate choice of the refractive index floor $n_0(z)$ while propagating the fields, as the approximation in Eq. (3.6) otherwise quickly becomes inaccurate for non-paraxial light fields. This is caused by at least two phenomena: *i)* phase shifts at large angles are under-represented when refractive index contrasts are high, since the increasing path length with angle is not represented. *ii)* inaccuracy could also be introduced in the propagator if the implementation does not match the refractive index representation to calculate phase shifts. To minimize inaccuracies for high angles in regions of large refractive index contrasts, the following representations

increase accuracy step-wise:

- *constant*: sets $n_0 = \text{const}$, typically $n_0 = \langle n \rangle_{xyz}$ *i.e.* to the full average of n. This results in refractive index contrast for typical biological tissue to be on the order of ± 0.06.
- *in-plane average*: sets $n_0(z) = \langle n \rangle_{xy}$, *i.e.* calculating the per-plane average of n. This has two advantages. *i)* it effectively allows the modeling of aberrations that occur at planar interfaces of high refractive index contrast, i.e. a cover glass or implications of a mismatched working distance. *ii)* biological tissues often possess a planar stratification, which implies plane wise varying average refractive indices are represented *exactly*.
- *weighted in-plane average*: sets $n_0(z) = \langle |u| n \rangle_{xy} / \langle |u| \rangle_{xy}$, *i.e.* calculating the per-plane average weighted by the field magnitude. This choice accounts, additionally to stratification averages, for finite lateral dimensions of samples like embryos and typical spatially confined light distributions (as in focused beams).

Already the first strategy largely reduces the problem of inaccuracy of phase projections of high refractive indices at large angles. The two additional tweaks are not generally applicable to any physical light scattering problem, but are of significant practical relevance when dealing with biological samples.

3.2.1 Implementation

Since each operation in Eq. (3.6) – specifically FFTs and point wise multiplications – is parallelizable within the whole propagation process it can be implemented efficiently on the GPU. The same applies to the calculations of mean refractive indices as mentioned above, where plane-wise averaging can be implemented on the GPU by reduction operations. We implemented the described light propagation method in OpenCL within the open source Python software package *biobeam*[3] (Weigert et al. 2018b), where all computationally heavy parts are lifted to the GPU again via OpenCL. We further extensively reused components already designed and implemented within Chapter 2. The package was specifically designed to make wave optical experiments *in-silico* as easy as possible, and additionally provides methods to calculate theoretical PSFs and predefined input fields, as well as PSF/aberration calculations by propagating diffraction limited point sources from within the tissue (described in more detail in the next sections). Furthermore the fast execution of the propagation simulations makes it possible to design intuitive graphical interfaces that enable the real-time simulation and display of different scattered light fields (by sharing OpenCL resources) and allow interactive exploration of optical experiments *in-silico* (*cf.* Fig. 3.3). For a short demonstration see Supp. Video 8.

Video 8

[3] https://github.com/maweigert/biobeam

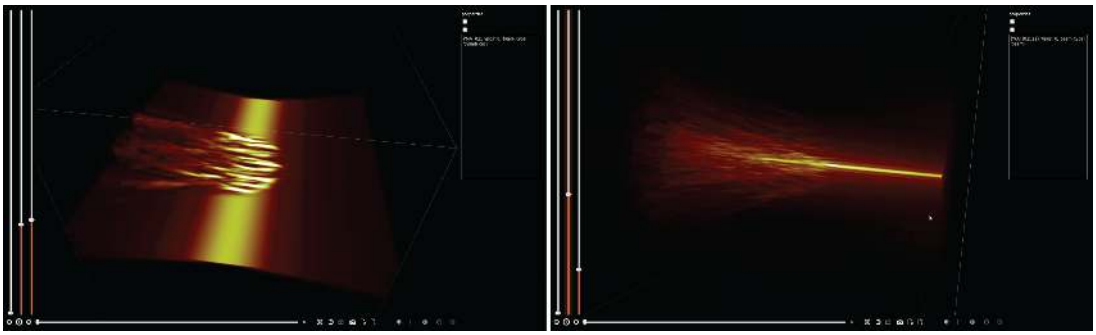


Figure 3.3: Screenshot of an interactive GUI for simultaneous simulation and display of light fields as scattered by refractive index distributions. Example shown for a spherical refracting tissue model. Beam parameters such as illumination type (left: cylindrical sheet, right: Bessel beam) and numerical apertures can be changed interactively and the resulting light intensity can be simulated and displayed within seconds. Implemented with *biobeam* and *spimagine* (*cf.* Section 2.5).

Dimension $x \times y \times z$	Mie code (GMMFIELD)	FDTD (MEEP)	BPM (*biobeam*)
$128 \times 128 \times 128$	1314s	80 s	34 ms
$256 \times 256 \times 256$	10480s	790 s	81 ms
$512 \times 512 \times 512$	–	4800 s	154 ms
$512 \times 512 \times 512$	–	–	440 ms

Table 3.1: Runtimes of plane wave propagation through a refracting sphere embedded in geometries of given dimensions.

3.2.2 Performance and Comparison with Existing Software

We compared the attainable performance of our *biobeam* implementation with available open source programs, specifically with MEEP (Oskooi et al. 2010)[4], which implements the Finite Difference Time Domain (FDTD) method to solve Maxwell's equation, and the Mie code GMMFIELD (Ringler 2008) that allows analytically exact internal field calculations for spherical refractive index distributions. To that end, we measured the runtime to propagate a plane wave through a refracting sphere ($\Delta n = 0.08$) in computational geometries of different size and constant pixel-size of $\Delta x = \lambda/5 = 0.1\mu m$ and list the results in Table 3.1. The runtime differences of several orders of magnitude clearly demonstrates the superiority of the BPM as implemented on GPUs (speedup of $\sim 10^5$ and $\sim 10^4$ compared to GMMFIELD and MEEP respectively). This efficiency results from a linear runtime scaling with respect to the propagation distance compared to Mie and FDTD codes (the latter being quadratic in every dimension), the mentioned approximations of low refractive index contrast, and a further speed up due to the massive parallelization of remaining operations on the GPU. To measure the speedup of the GPU/OPENCL based implementation compared to a pure CPU (in C) and GPU/CUDA implementation we compared the runtime of a plane wave propagating through a constant refractive index distribution for three different basic implementations of the BPM method:

[4] http://ab-initio.mit.edu/wiki/index.php/Meep

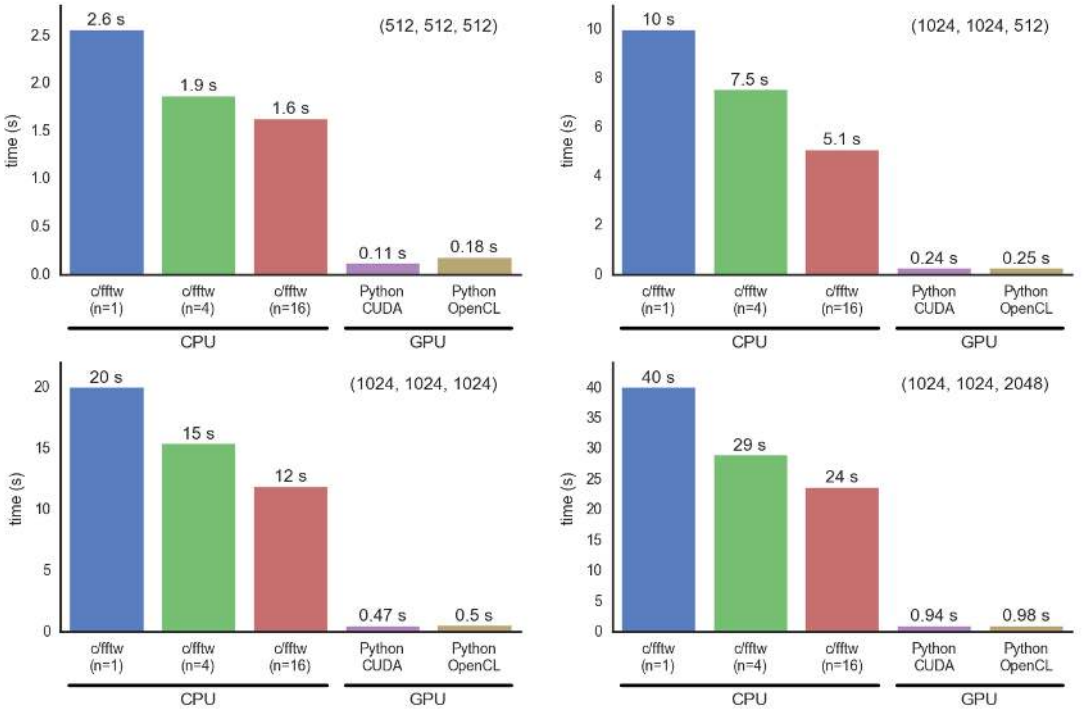


Figure 3.4: Comparison for different BPM implementations on CPU vs GPU. We show the time to propagate a plane wave through a constant refractive index distribution on geometries of increasing dimensions. CPU: FFTW3 with `FFT_MEASURE` with 1, 4 or 16 threads on a 20 core Xeon(R) CPU E5-2660 v3 (2.60GHz, 64GB RAM). GPU: OpenCL via PyOpenCL or CUDA via PyCuda on a NVIDIA GeForce GTX Titan X (12GB RAM).

CPU Single precision implementation in C (FFTW 3.3 with `FFTW_MEASURE` as FFT library), and different levels of multithreading with 1, 4 or 16 threads.

GPU (1) GPU implementation with CUDA in Python

GPU (2) GPU implementation with OpenCL in Python

The runtime for the different methods and sizes are shown in Fig. 3.4. Here, the GPU accelerated methods offered an order-magnitude increase in performance compared to the single- and multi-threaded CPU versions. The speedup of the OpenCL version compared to the single threaded CPU version was between 15–40 fold and to the 16-multithreaded version between 10–25 fold, where the greater speedups were achieved in the relevant scenario of bigger propagation volumes. Interestingly, the performance of both OpenCL and CUDA GPU implementations are comparable for large geometries, vindicating our choice of the hardware independent – yet potentially slower – OpenCL compared to the hardware specific CUDA platform. Importantly, the propagation of an arbitrary input field through a realistically large refractive index distribution of size $1024 \times 1024 \times 2048$ can be achieved in $\sim 1s$, rendering it compatible with iterative algorithms and motivating its incorporation into more complex wave-optical image formation models in microscopy as described in the later sections.

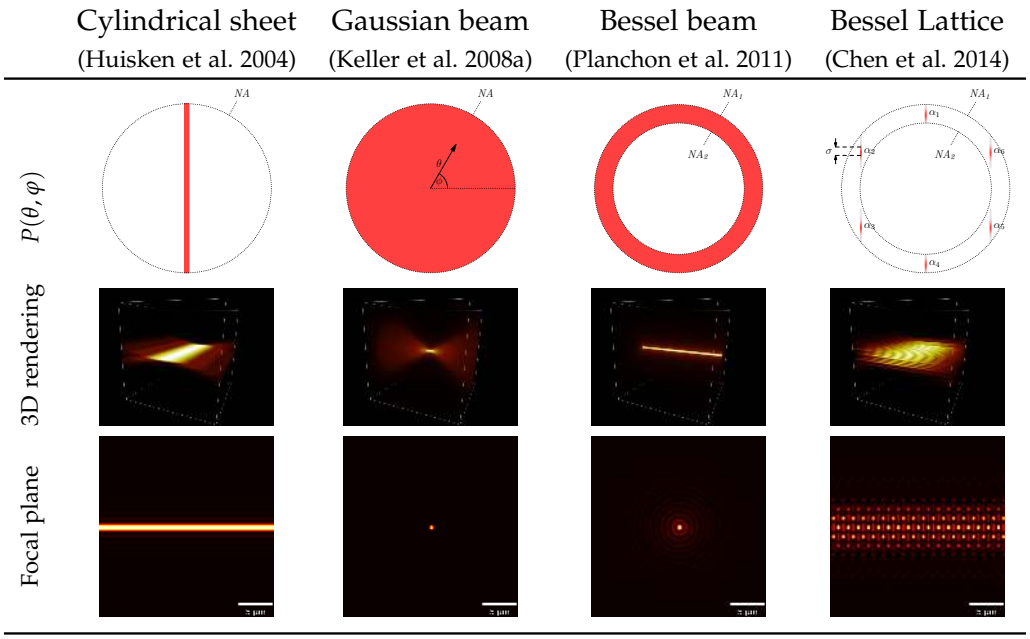


Table 3.2: Different pupil functions and the resulting illumination fields. Scalebar $5\mu m$. See as well Supp. Video 9.

3.2.3 Input Fields

Apart from the three-dimensional refractive index distribution $n(x,y,z)$ the full light field $u(x,y,z)$ is determined by the incident complex field $u(x,y,z=z_0)$, which can be a plane wave, a focused Gaussian beam, a light sheet generated by a cylindrical or other more complex initial distributions. For microscopy, where light is passing through the exit aperture and focused by an objective, this incident focus field at a certain position z_0 on the optical axis can be calculated from the pupil *aperture function* $P(\theta,\varphi)$, *i.e.* the complex field distribution incident on the (circular) exit pupil of the objective given in polar pupil coordinates (θ,φ). An accurate description of the focus field for linear polarized light (with $\vec{E}^0 = (E_x^0, 0, 0)$) can be derived by the vectorial Debye-Wolf integral (*cf.* Foreman et al. 2011, Appendix A.2)[5]:

$$
\begin{aligned}
E_i(\rho,\varphi,z) &= \frac{-ikfE_x^0}{4\pi}\int_0^{\alpha}\int_0^{2\pi} d\theta d\phi\, B_i(\theta,\phi)P(\theta,\phi)\sqrt{\cos\theta}\sin\theta\cdot e^{ik\rho\sin\theta\cos(\phi-\varphi)}e^{ikz\cos\theta}\\
B_x(\theta,\phi) &= (\cos\theta+1)+(\cos\theta-1)\cos 2\phi\\
B_y(\theta,\phi) &= (\cos\theta-1)\sin 2\phi\\
B_z(\theta,\phi) &= -2\sin\theta\cos\phi, \qquad \alpha = \arcsin\frac{NA}{n_0}
\end{aligned}
\tag{3.7}
$$

with the numerical aperture NA, the surrounding refractive index n_0, and E_i being the vectorial field components ($i = x,y,z$) expressed in cylindrical coordinates (ρ,φ,z). We will thus use the accurate $E_x(x,y,z=z_0)$ as input field (with $E_x^0 = 1$) for the different illumination modes (*cf.* Eq. (3.7)). Depending on the aperture function $P(\theta,\varphi)$, these will result in Gaussian or Bessel beams (Keller et al. 2008a; Planchon

[5] For objectives that fulfill Abbe's sine condition, e.g. aplanatic objectives.

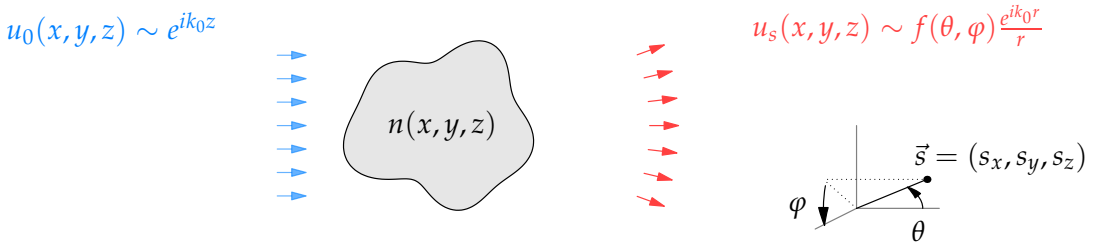


Figure 3.5: Schematic of scattering geometry: Incoming plane wave $u_0(x,y)$ is scattered at an object with refractive index $n(x,y,z)$. The scattered field is assumed to be a spherical wave in the far field modulated by the phase function $f(\theta,\varphi)$.

Video 9

et al. 2011), cylindrical light sheets (Huisken et al. 2004) or Bessel lattices (Chen et al. 2014) as can be seen in Table 3.2 (*cf.* Supp. Video 9). All of these were implemented on the GPU as volumetric diffraction integrals that allow the fast and efficient calculation of relevant input fields as well as volumetric PSFs according to the accurate vectorial focus field model Eq. (3.7).

3.2.4 Scattering Calculations

After having seen how to efficiently calculate the light distribution inside a given refractive index map, we embark now to quantify the amount of *scattering* that a plane wave suffers when propagating through tissue. This will be important later for validating our approach with analytical scattering models, as well as for quantifying the bulk properties of retinal tissue and the effect of nuclear inversion on the scattering properties. In general, the resulting light field $u(x,y,z)$ of a incident plane wave in the presence of a refractive index distribution $n(x,y,z)$ can be described as the superposition of the incident plane wave $u_0(x,y,z) \sim e^{ik_0 z}$ and the scattered field $u_s(x,y,z)$

$$u(x,y,z) = u_0(x,y,z) + u_s(x,y,z), \tag{3.8}$$

where the scattered field in the *far field* is assumed to be a outgoing spherical wave modulated by the *scattering phase function* or *scattering amplitude* $f(\theta,\varphi)$:

$$u_s(x,y,z) \underset{r\to\infty}{\sim} u_s(s_x,s_y,s_z) = f(\theta,\varphi)\frac{e^{ik_0 r}}{r} \tag{3.9}$$

where $\vec{s} = (s_x,s_y,s_z), |\vec{s}| = 1$ is a point on the unit sphere (*cf.* Fig. 3.5, Bohren et al. 2008). The phase function $f(\theta,\varphi)$ is the fundamental quantity of interest that describes the scattering properties. Its square modulus is the *differential cross section*

$$\frac{d\sigma}{d\Omega} = |f(\theta,\varphi)|^2 \tag{3.10}$$

describing how much energy is scattered at a specific solid angle Ω. The total *scattering cross section* σ is the integral over all solid angles

$$\sigma = \int d\varphi d\theta \sin\theta |f(\theta,\varphi)|^2 \tag{3.11}$$

and the *asymmetry parameter* g is given by the average cosine of the polar angle θ:

$$g = \sigma^{-1} \int d\varphi d\theta \sin\theta \cos\theta |f(\theta,\varphi)|^2. \tag{3.12}$$

As these quantities are defined only for the far field of the scattered wave, we now want to derive $f(\theta,\varphi)$ when only the near field distribution $u(x,y,z_0)$ at a propagation position z_0 after the refracting object is given. We again start with the angular spectrum of $u(x,y,z_0)$

$$U(k_x,k_y,z_0) = \frac{1}{2\pi} \int dxdy\, u(x,y,z_0) e^{-i(k_x x + k_y y)} \tag{3.13}$$

from which the far field $u_s(s_x,s_y,s_z)$ can be calculated via the method of stationary phase (Born et al. 1999)[6]

$$u_s(s_x,s_y,s_z) = -ik_0 s_z U(k_0 s_x, k_0 s_y) \frac{e^{ik_0 r}}{r}, \tag{3.14}$$

which results in a scattering phase function

$$f(\theta,\varphi) = -ik_0 \cos\theta U(k_0 \cos\varphi \sin\theta, k_0 \sin\varphi \sin\theta). \tag{3.15}$$

To carry out the integration for the cross section and asymmetry factor in Eqs. (3.11) and (3.12) we further have to perform a coordinate change in the spectral domain from Cartesian to polar coordinates $(k_x,k_y) \to (\theta,\varphi)$ via

$$k_x = k_0 \cos\varphi \sin\theta \qquad\qquad k_y = k_0 \sin\varphi \sin\theta, \tag{3.16}$$

inducing the area element $dA = dk_x dk_y$ to change according to the determinant of the Jacobian of this transformation $dk_x dk_y = k_0^2 \sin\theta \cos\theta d\theta d\phi$. Together with the relation $\cos\theta = \sqrt{1 - k_x^2/k_0^2 - k_y^2/k_0^2}$ we thus find the integrals Eqs. (3.11) and (3.12) to be

$$\begin{aligned} \sigma &= \int dk_x dk_y |U(k_x,k_y)|^2 \sqrt{1 - k_x^2/k_0^2 - k_y^2/k_0^2} \\ g &= \sigma^{-1} \int dk_x dk_y |U(k_x,k_y)|^2 (1 - k_x^2/k_0^2 - k_y^2/k_0^2). \end{aligned} \tag{3.17}$$

This relates all important quantities to the angular spectrum of the near field as propagated through the refractive index distribution.

[6] For more information see as well https://www.photonics.ethz.ch/fileadmin/user_upload/Courses/EM_FieldsAndWaves/AngularSpectrumRepresentation.pdf

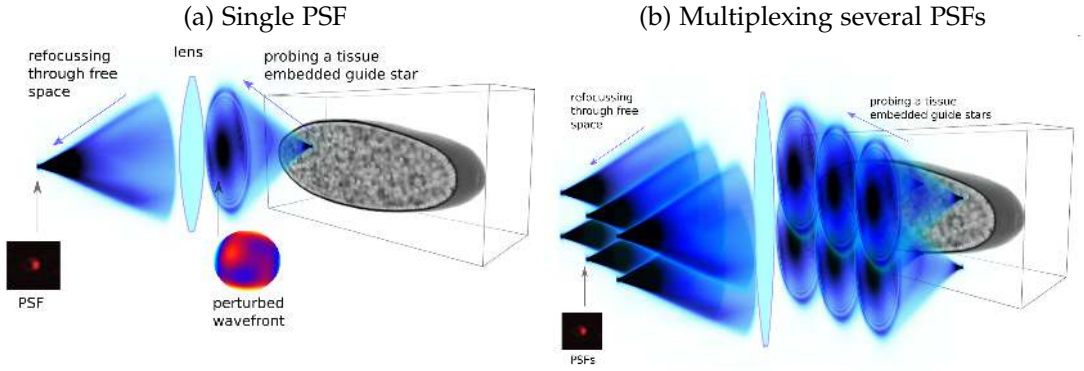


Figure 3.6: Single and multiplexed detection PSF calculations. a) Propagating a diffraction limited input field through parts of the sample and refocusing by an idealized optical system gives the focus field as seen by the detector. b) If the refocus spots are separated for different starting points, the propagation of a complete grid can be carried out in a highly multiplexed manner, accelerating the process for typical microscopy simulations by a factor 100–1000.

3.2.5 Multiplexing of PSF Calculations

To estimate the aberrations inflicted on a PSF at a specific location z_0 within a refractive index distribution we proceed as follows: First, we propagate an input field corresponding to a diffraction-limited focus spot with the given numerical aperture of the detection system towards the direction of the lens for a distance Δz through the refractive index map. The angular spectrum of the resulting field corresponds to the aberrations incurred. We next phase-conjugate this complex field, and propagate it over Δz again, this time through free space. The resulting volumetric field corresponds to the volumetric PSF (*cf.* Fig. 3.6a & Supp. Video 6). In case of light-sheet microscopy simulations these point spread function can furthermore be multiplied in axial direction with the intensity of the orthogonal illumination sheet to result in the single-point PSF.

Video 6

To realistically simulate image formation of the whole volumetric acquisition process, many PSFs at different locations have to be determined owing to the inhomogeneity of refractive index distribution. Importantly, the induced aberrations can change within $\sim 10\mu m - 20\mu m$, such that for a field of view of $300\mu m^3$ approximately $3 \cdot 10^3 - 3 \cdot 10^4$ unique such PSF calculations would be necessary. To circumvent this problem, we exploit the linearity of optics, specifically the property, that if sufficiently spaced point spread functions in a plane do not significantly overlap, they can be propagated and refocused in a *multiplexed* way (*cf.* Fig. 3.6b). We thus define as the input field at z_0 a whole grid of equally spaced, diffraction limited focal distributions which will be simultaneously propagated and refocused in a multiplexed manner. The PSF grid sampling has to be done at a grid-spacing below the smallest *iso-planatic patch* size in the sample, *i.e.* the length scale of aberration variations. Furthermore the grid-spacing has be larger than the typical width of the PSF, ensuring that no substantial coherent overlap arises between neighboring PSFs. In case both requirements are not satisfied, a checkerboard like partition of the grid should be

considered that alternatingly computes neighboring PSFs. This reduces the computational complexity by the number of simultaneously calculated PSFs, *i.e.* typically 100-1000 fold (*i.e.* the number of PSFs that are propagated for each grid at z_0).

3.2.6 Image-formation in Light-sheet Microscopy

We conclude this section with the description of the image formation simulation in light-sheet microscopy. Here, the microscope is composed of two distinct light paths, one for excitation and one for the collection of fluorescent light. Light-sheet microscopy achieves optical sectioning by the orthogonality between the illumination and the detection axis (Huisken et al. 2004). To model the image-formation process let

$f(x,y,z)$, the fluorophore density in the sample
$h(x,y,z)$, the PSF of the detection objective
$j(x,y,z)$, the illumination intensity (PSF) of the light sheet
$g(x,y,z)$, the observed volumetric intensity (image)

where z is the coordinate along the optical axis of the detection objective and y the propagation axis of the illuminating field. Let the current focal position be z_F and fixed. Then the excited intensity distribution of the fluorophore $\tilde{f}(x,y,z)$ is

$$\tilde{f}(x,y,z) = f(x,y,z) \cdot j(x,y,z_F - z) \tag{3.18}$$

For each constant slice at z the intensity $\tilde{f}(x,y,z)$ gets convolved with $h(x,y,z_F - z)$ resulting in

$$a(x,y,z_F,z) = \int dx'dy'h(x',y',z_F - z)\tilde{f}(x - x', y - y', z) \tag{3.19}$$

$$= \int dx'dy'h(x',y',z_F - z)f(x - x', y - y', z)j(x - x', y - y', z_F - z) \tag{3.20}$$

and the overall observed image plane $g(x,y,z_F)$ for a fixed focal position is then the integral of all contributions

$$g(x,y,z_F) = \int dx'dy'dz'h(x',y',z')f(x - x', y - y', z_F - z')j(x - x', y - y', z') \tag{3.21}$$

Note that if the illumination field can be factored as $j(x,y,z) = j_{xy}(x,y) \cdot j_z(z)$, then

$$g(x,y,z_F) = (h \cdot j_z) \otimes (f \cdot j_{xy}) \tag{3.22}$$

and the observed image g is simply the 3 dimensional convolution of the illuminated density $f \cdot j_{xy}$ with an effective PSF $h_{eff} = h \cdot j_z$. For a distorted illumination field, this however is not true in general and the detection PSF typically varies spatially due to the distortions induced by the tissue. For modeling this complete image-formation process we thus execute the following steps at every desired axial position z_F:

1. computing the distorted illumination field j at a given position z_F
2. calculating a fine grid of multiplexed spatially varying detection PSFs at that axial position (as described before), and
3. performing with it a spatially varying convolution of the product of the fluorophore signal and the 3d excitation profile to produce the final image at z_F as seen by the detector.

The resulting PSFs are interpolated between sampling points to result in a quasi-continuum that can be used for an accurate convolution with a spatially varying, wave-optically determined kernel. To convolve the input signal with the spatially varying PSF field we use standard methods (Nagy et al. 1998). A detailed description of a simulation can be found in Section 3.4.1. For a visual exposition of the whole process see as well Supp. Video 5.

Video 5

3.2.7 Limitations

As already indicated in the introductory sections, the approximations employed in the scalar beam propagation method incurs some limitations. This include its restriction to refractive index distributions of low contrast $\Delta n < 0.15$ that is *e.g.* exceeded by structures made of lipids or collagen (with a refractive index of 1.49 and 1.55 respectively, Johnsen et al. 1999). Furthermore simulation of high numerical illumination/detection apertures ($NA > 0.9$) and or refraction events at high angles will lead to lost accuracy. An adaption of more accurate methods *e.g.* (S. Schmidt et al. 2016) to the use-cases as illustrated in this thesis might therefore merit future work. Another aspect is that Eq. (3.6) in its current form assumes periodic boundary conditions, thus any refractive index distributions needs to be padded to avoid artifacts. A potential approach to circumvent this problem is described in (Hadley 1992). Furthermore, the current implementation of fast discrete Fourier transforms in OpenCL only allows for power-of-two inputs, hence the volume has to be appropriately padded along the dimensions orthogonal to the propagation axis. Finally, due to the formulation as a propagation method, back-scattering is completely ignored, which can be problematic for certain applications (*e.g.* effectively prohibiting OCT microscopy simulations, Fujimoto et al. 2000).

3.3 Validation

After describing the main methods and its efficient implementation on GPUs, the purpose of this section is to validate the adequacy of the scalar beam approximation for a multitude of imaging scenarios in microscopy of biological tissue. First, we will compare against analytical models of sphere scattering, after which we compare against several experiments such as knife-edge diffraction, scattering in light-sheet microscopy, and the projection of images through refracting spheres.

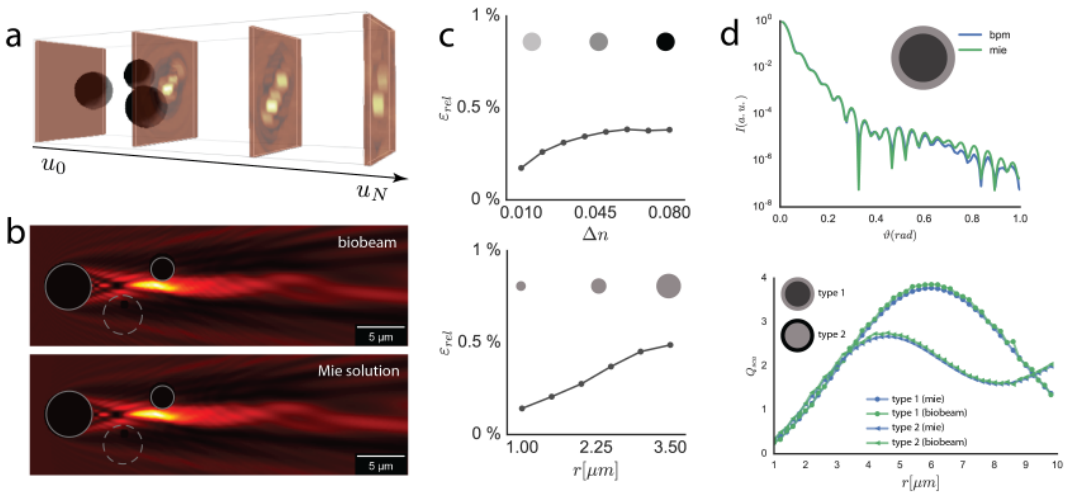


Figure 3.7: Validation with analytical solutions. a) Plane wave scattered by three solid spheres ($\lambda = 500nm$, $r = 2 - 2.5\mu m$, refractive index contrast $\Delta n = 0.05$), b) Comparison of analytical solution (Mie theory) versus simulation. c) Maximal relative error of near field distribution as a function of single sphere radius r ($\Delta n = 0.05$) and refractive index contrast Δn ($r = 2.5\mu m$). d) Top: Phase function of analytically tractable coated spheres as cell models ($\Delta n = 0.02/0.04, r = 2\mu m/2.5\mu m$) shows high accuracy up to approximately 0.5 radians. Bottom: Size dependent scattering efficiency of the same sphere architecture and its inverse. (Reproduced with permission from Weigert et al. 2018b, Creative Commons license CC BY 4.0)

3.3.1 Analytical Validation

To validate the accuracy of the described BPM implementation we first compared simulated field distributions with analytically tractable models. Specifically, we analyzed the case of incident plane waves scattered by solid spheres of a given diameter and refractive index. Both the scattering properties and the near field distribution for this scenario can be analytically calculated by *Mie theory* (Hergert et al. 2012), where the Helmholtz equation can be solved by expanding the field as (vector) spherical harmonics and the scattered field described as a converging series of contribution terms (Gouesbet et al. 2011). Furthermore for both scattering and near field calculations software packages are available (by Mätzler 2002; Bohren et al. 2008[7] and GMMFIELD by Ringler 2008[8] respectively). We first compared the simulation result with the Mie theory prediction for 3 randomly distributed spheres of radius $r = 2 - 2.5\mu m$ and refractive index contrast of $\Delta n = 0.05$ (*cf.* Fig. 3.7a). We chose a computational geometry of size $12.8\mu m \times 12.8\mu m \times 42\mu m$ and an isotropic pixel discretization $\Delta x = 0.1$, and padded the synthetic refractive index map appropriately. We then calculated the intensity distribution of an incident plane wave with $\lambda = 0.5\mu m$ with both *biobeam* and *Gmmfield*. In Fig. 3.7b, we show a slice through the intensity distribution in both cases, where one can see both solution to be virtually indistinguishable. The maximal relative intensity error in this case amounted to $\varepsilon_{rel} = 3\%$. We next quantified the relative error in the same way for single spheres

[7] https://github.com/jleinonen/pymiecoated
[8] http://moritz-ringler.name/dissertation/GMM_FIELD.tar.bz2

of varying refractive index contrast $\Delta n \in (0.01, 0.08)$ and radius $r \in (1\mu m, 3.5\mu m)$. As can be inferred from Fig. 3.7c and Fig. 3.7d the relative error stays below 5% in all cases, being the highest for large spheres with large refractive index contrast, as expected. We then calculated the scattering phase function $f(\theta, \phi)$ from the angular spectrum as derived in Eq. (3.15) for a concentrically coated spheres. This mimics a simple nuclei model, where a refracting spherical cell nucleus is surrounded by a spherical shell of cytoplasm. Again, the results for both the BPM simulations as well as calculated from Mie theory are near identical for small to moderate angles (Fig. 3.7d top). Significant relative errors only arise at high angles at which however little intensity is scattered. We finally calculated the *scattering efficiency* $Q_{sca} = \frac{\sigma}{\pi r^2}$, *i.e.* the scattering cross section relative to the geometric cross section, for coated spheres with various size parameters. The comparison with analytical results again showed good agreement over the whole range of different sizes (Fig. 3.7d bottom).

3.3.2 Experimental Validation

To test the applicability of *biobeam* in settings relevant for microscopy, we designed a number of imaging experiments and compared them with their *in-silico* reconstitution. Specifically, we *i)* acquired the light intensity behind a diffracting edge and compared it with the simulation results, *ii)* recorded the projected image of a test-chart as seen through refracting spheres and simulated the whole corresponding image formation process *in-silico*, and *iii)* measured the light scattered from agarose embedded spheres in a commercial light-sheet microscope and compared it with the results from the simulation pipeline.

Diffraction around a Knife Edge

We first consider the classical experiment of the diffraction pattern of a plane wave at a knife edge. For that, we used a custom microscopy setup that focused a incoherent light source (M470L3 Thorlabs LED, mean wavelength $\lambda_0 = 0.47\mu m$, measured spectral FWHM $\sim 20nm$) at infinity to create an effectively plane incident wave ($NA = 0.001$) at the image plane, where a Thorlabs R1L3S6P variable line grating test chart was positioned to be used as an edge[9]. The resulting intensity distribution was then recorded as a volumetric stack with isotropic pixel-size $\Delta x = 0.29$.
To reconstitute the experiment *in-silico*, we simulated the propagation of an incident complex field distribution

$$u_0(x, y) = e^{2\pi\lambda k_x x}\theta(x - x_0) \tag{3.23}$$

clipped at the position of the edge[10]. A computational grid of size $1024 \times 256 \times 1830$ with the experimental pixel-size of $\Delta x = 0.29\mu m$ was used. We simulated both the coherent and incoherent case. For the coherent case we set $\lambda = \lambda_0 = 0.47\mu m$, for

[9] Experiment done with Kaushikaram Subramanian and Moritz Kreysing (MPI-CBG)
[10] Heavyside step-function $\theta(x)$

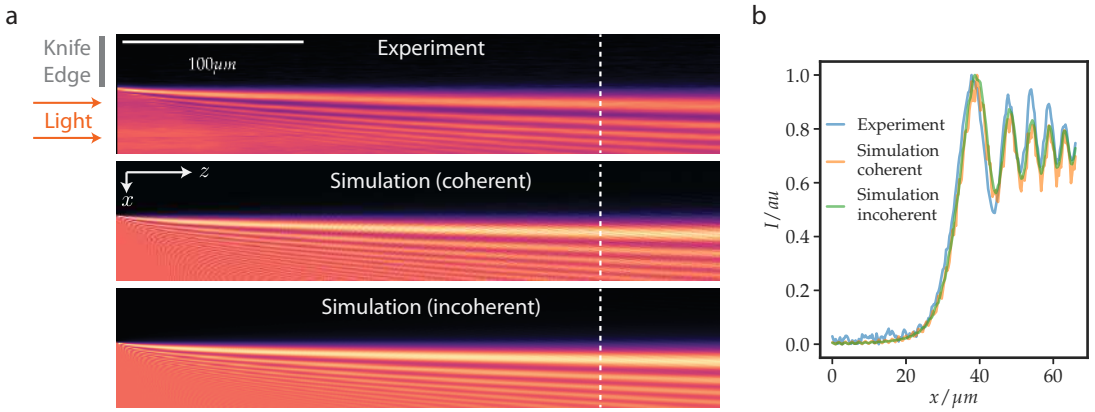


Figure 3.8: Experimental validation with diffraction pattern around a knife edge. a) Light is focused with an incoherent light source (M470L3 Thorlabs, $\lambda_0 = 470nm$) such that an almost plane wave ($NA = 0.001$) illuminated a knife edge. The diffracting light was imaged at different depths z from the edge and the intensity recorded (upper image). The simulation was done on a computational cell of size $(1024 \times 256 \times 1830)$ with voxel size $\Delta x = 0.3\mu m$. We simulated the diffraction in the case of a single plane wave (coherent, middle) and the incoherent superposition of 100 incident plane waves of uniformly sampled wavelengths $\lambda \in [460nm, 480nm]$, corresponding to the measured spectral width of $\pm 10nm$ of the light source (incoherent, bottom). b) Intensity plot at a given axial position (dashed line in a) for experiment and the two simulations. (Reproduced with permission from Weigert et al. 2018b, Creative Commons license CC BY 4.0)

the incoherent case we uniformly sampled $\lambda \sim \mathcal{U}(0.46\mu m, 0.48\mu m)$ and averaged the intensities of 100 such simulations, thus taking into account the spectral width of the illumination source.

A xz-slice of the resulting intensity distribution is depicted in Fig. 3.8a, with both coherent and incoherent simulations matching well the experimentally acquired data. We further plot the normalized intensity as a line cross section in Fig. 3.8b. Here, the additional smoothness of the incoherent result can be seen, when compared to the coherent case. Again, the spatial intensity variation of the experimental data is well reproduced, with slight differences appearing at bigger lateral (x) positions.

Image Projection through Micro-spheres

In this experiment, we use a custom micro projection setup to accurately project patterned light fields on a sample and collect the transmitted light at different focal planes, using an influx/efflux system with a 20x/0.45 NPL Fluotar and Olympus UPLSAPO 20x/0.75 objective respectively[11]. A negative USAF (R1DS1N, Thorlabs) test-chart was illuminated incoherently (M470L3 Thorlabs LED) and projected onto the backside of Borosilicate sphere ($n = 1.48$, $110\mu m$ diameter, Cospheric LLC, USA) embedded in OptiPrep ($n = 1.42$), as depicted in Fig. 3.9a. The detection objective was then focused at the same projection plane and the resulting image captured using a Andor Zyla 5.5 sCMOS camera. To simulate the experimental situation we used a computational grid of size $1024 \times 1024 \times 730$ with a pixel size of $\Delta x = 0.161\mu m$. A spherical refractive index distribution with $n = 1.48, r = 110\mu m$ was defined, using

[11] Experiment done with Kaushikaram Subramanian and Moritz Kreysing (MPI-CBG).

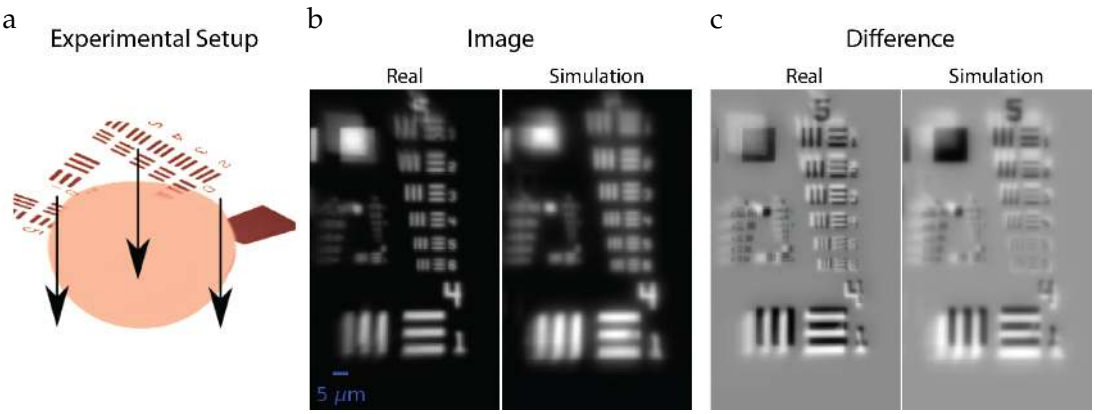


Figure 3.9: Experimental micro-projection of a test-chart through a glass sphere and comparison with simulation. a) Experimental setup: The negative USAF (R1DS1N, Thorlabs) test-chart was illuminated incoherently (M470L3 Thorlabs) and projected behind a glass sphere (Borosilicate material, $n = 1.48$, $110\mu m$ diameter, Cospheric LLC, USA). The transmitted image at the same focal plane was then recorded. b) The images from the experiment (Real) and the simulation. c) The difference images as calculated with respect to the undistorted test-chart image, showing that the real sphere-induced image distortions are qualitatively reproduced by the simulation, although with additional image blur at the image edges. Scale-bar (blue) $5\mu m$. (Reproduced with permission from Weigert et al. 2018b, Creative Commons license CC BY 4.0)

$n = 1.42$ as the refractive index of the outer medium. We then calculated a 16×16 grid of detection PSFs as described in Section 3.2, corresponding to an imaging plane directly behind the sphere. To ensure non overlapping signal of the individual PSFs, we multiplexed the PSF simulation according to a 2×2 checkerboard pattern. Finally, we performed a spatially varying convolution of the input (test-chart) image with the obtained grid of PFSs. As can be seen from Fig. 3.9b the lower order, geometric, distortions (*e.g.* image shift and spatially varying defocus blur) are qualitatively well reproduced. This can be seen more clearly in Fig. 3.9c, where we show the difference of both real and simulated images compared to the undisturbed reference image when removing the refracting sphere. Still, the simulation results seem to produce additional image blur not found in the experiment, especially at the border of the image. This might be due to rays refracting at the border of the sphere at high angles and thus failing to get reproduced by the BPM approximation.

Light-sheet Microscope Scattering

Finally, we compare experimental scattering results acquired on a full light-sheet microscopy setup with the image formation pipeline as outlined in Section 3.2. To that end, we measured the intensity of a cylindrical light sheet scattered by a micro-sphere embedded in agarose, as depicted in Fig. 3.10a[12]. We used a commercial Z.1 light-sheet microscope (Zeiss, Germany) equipped with a 20x/1.0 detection objective and a 10x/0.2 illumination objective. To obtain a stationary light sheet, we deactivated the pivot and sheet scanner. The illumination iris was adjusted to obtain a light sheet with a waist of $1.7\mu m$ and a lateral width $\sim 100\mu m$

[12] Experiment done with Sebastian Bundschuh and Kaushikaram Subramanian (Kreysing group, MPI-CBG).

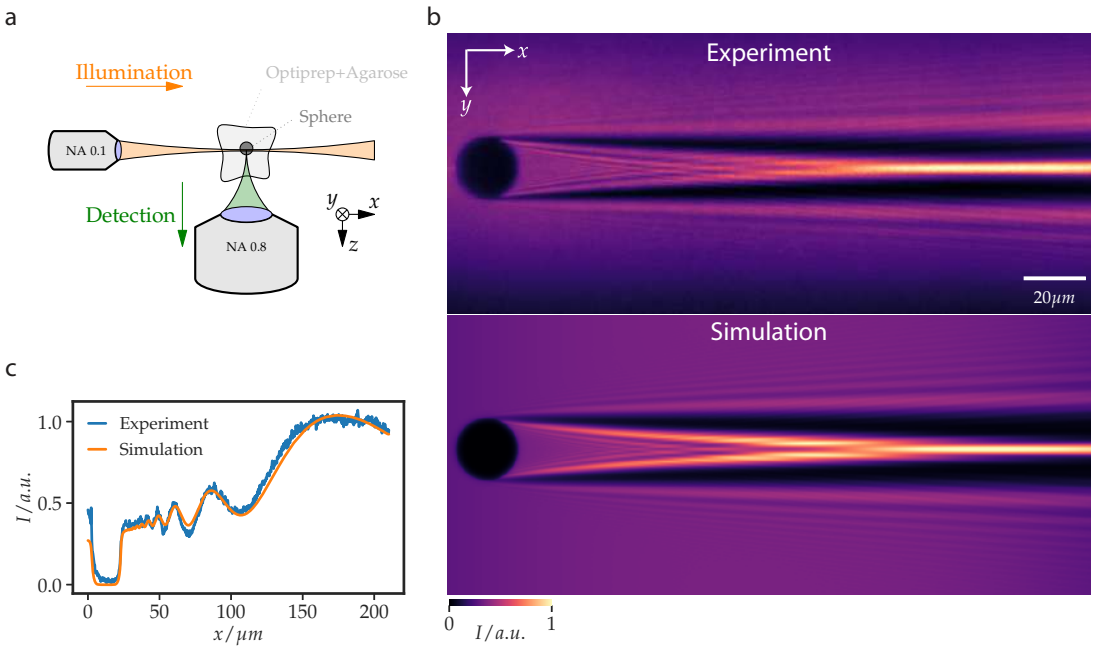


Figure 3.10: Experimental validation on a commercial light-sheet microscope. a) Polymethylmethacrylate (PMMA) sphere with a diameter of $20\mu m$ and refractive index of $n = 1.495$ were embedded in an block of OptiPrep (Progen Biotechnik GmbH) / agarose (Sigma Aldrich) with refractive index of $n \approx 1.43$) and which was labelled with Alexa Fluor 488. A stationary illuminating light sheet with a waist of $1.7\mu m$ and a lateral extension of $\approx 100\mu m$ was generated with a LZ1 (Zeiss) light-sheet microscope, incident on the agarose embedded sphere. b) Simulation results of the intensity distribution behind the sphere at a plane incident to the sphere center and comparison with the simulation. c) Line-plot of the intensities in b) along x through the center of sphere for both cases. Scalebar is $20\mu m$. (Reproduced with permission from Weigert et al. 2018b, Creative Commons license CC BY 4.0)

(FWHM). Polymethylmethacrylate (PMMA) microspheres ($n = 1.495$) with a diameter of $20\mu m$ (PolyAn GmbH, Germany) were mounted within a block of low-gelling agarose mixed with OptiPrep and labelled with Alexa Fluor 488 to result in a refractive index of $n \sim 1.429$. The final stacks were acquired with a pixel-size of $0.23\mu m \times 0.23\mu m \times 0.42\mu m$. An excitation wavelength of $488nm$ and a detection bandpass of $505 - 545nm$ was used. The image focused at the center of the scattering sphere is shown at the top of Fig. 3.10b.

To reconstitute the experimental data, we used a computational grid of size $1024 \times 512 \times 512$ with a pixel size of $\Delta x = 0.205\mu m$. Next, we defined a spherical refractive index distribution with $n = 1.495, r = 10\mu m$ at the position given by the experimental image stack. As incident illumination field, a cylindrical light sheet with $NA = 0.2$ and $\lambda = 0.49\mu m$ was assumed and calculated according to Eq. (3.7). Next a grid of (16×16) detection PSFs was generated (detection $\lambda = 0.52$) and spatially varying convolved with the illumination intensity to give the final image.

In Fig. 3.10b, we compare the intensity distributions for the experimental and the simulated case. As can be seen, the scattered field is well reproduced by the simulation and specific diffraction patterns within the focused region are reconstituted accurately. This can be seen ever more when plotting the normalized intensity cross section through the center of the sphere as done in Fig. 3.10c.

3.4 Applications

After having validated the adequacy of the described light propagation method on several experimental examples, we now show for several applications how its computational efficiency allows the *in-silico* investigation of optical phenomena in realistically large biological samples: *i)* We perform the simulation of the wave-optical image formation process in light-sheet microscopy of a refracting tissue model. This includes the accurate propagation of the cylindrical light sheet, the spatially resolved distribution of aberrated PSFs, and the reproduction of the resulting common image artifacts. *ii)* We investigate the spatial correlation of aberrations in scattering tissue motivated by its relevance for adaptive optics. To that end we first reproduce the *conventional shift-shift memory effect*, which describes the area over which an aberration correction focal spot preserves its shape when being translated (size of the *isoplanatic patch*). Next, we demonstrate via simulations that by a specific decomposition of the aberration correction this area can be increased $2-3$ fold (*extended memory effect*). Finally, *iii)* we investigate the optical properties of the outer nuclear layer (ONL) of mouse retina, which light has to pass before being detected at the photoreceptors. Here, we provide new computational insights into the hypothesized link between the inversion of nuclear DNA in diurnal animals with an improved scattering properties and subsequent visual sensitivity. Specifically, we create refractive models of large 3D retinal tissue models obtained from microscopy and demonstrate that nuclear inversion reduces side scattering substantially, as hypothesized.

3.4.1 Simulation of Light-Sheet Microscopy Image Formation

We first demonstrate the utility of our method by presenting the wave-optical simulation of the volumetric image-formation process in a light-sheet microscope. A light-sheet microscope uses two distinct light paths, one for excitation and one for the collection of fluorescent light. While *biobeam* enables to flexibly implement simulation of different imaging modalities, we will here only consider the use of a single cylindrical light sheet with the numerical aperture of illumination being $NA_{illum} = 0.1$ and detection $NA_{det} = 0.7$.

Tissue model: We first generated a synthetic refractive index distribution $n(x, y, z)$ of a tissue model mimicking a developing embryo (*cf.* Fig. 3.11a). To that end, we first started with a computational geometry of $1024 \times 2048 \times 1024$ and isotropic pixel-size $\Delta x = 0.098 \mu m$, resulting in a simulation volume of $100 \mu m \times 200 \mu m \times 100 \mu m$. We then created the refractive index map of an ellipsoidal body ($r_{min} = 40, r_{max} = 80$, $n = 1.37$, including a shell with $n = 1.38$), embedded $\sim$1400 spherical compartments mimicking cell nuclei ($r = 2.5 \mu m$, $n = 1.36$), added a larger central

[13] Absorbing material is modeled by complex refractive indices with non-zero imaginary part
[14] Refractive indices for the different cell constituents were chosen as reported by (Choi et al. 2007)

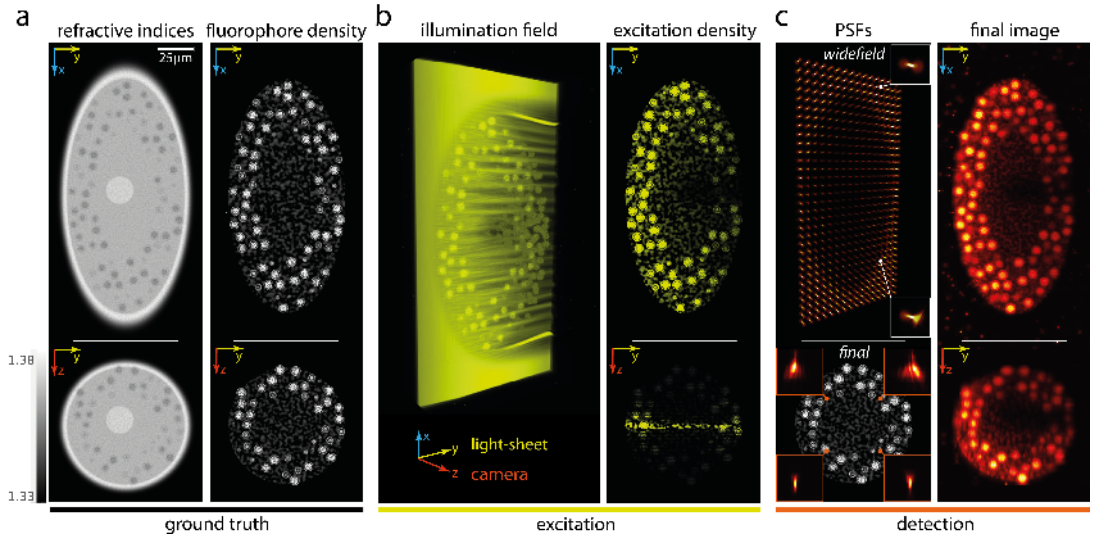


Figure 3.11: Rigorous wave-optical simulation of image formation process in light-sheet microscopy.(a) Synthetic tissue phantom of a multicellular organism ($100 \times 200 \times 100\mu m$) comprising a complex refractive index distribution (left, $n = 1.33 - 1.38$) and a fluorophore distribution of interest (right). (b) Wave optical simulation of the illuminating light sheet and resulting excitation distribution within the sample at a given z position. (c) Partially coherent simulation of the detection path by multiplexed calculation of all independent point spread functions (left) and the resulting simulated camera image combining illumination and fluorescence path of light through the scattering sample. (Reproduced with permission from Weigert et al. 2018b, Creative Commons license CC BY 4.0)

compartment ($r = 10\mu m$, $n = 1.38 + 0.01i$) mimicking a highly absorbing[13] tissue organelle, and assumed water ($n = 1.33$) as surrounding medium[14]. We finally added Perlin noise (Perlin 1985) with $\sigma_n = 0.005$ to account for high frequency refractive index variations (*cf.* Fig. 3.11a left). As fluorescence signal density $f(x, y, z)$ we used a combination of the nuclei masks, Perlin noise auto-fluorescence background, and distributed single pixel signal pixel mimicking embedded sub-diffraction beads (*cf.* Fig. 3.11a right).

Image formation simulation: To generate the image as seen by the camera at focal plane z_F, we first simulated a the interaction of a cylindrical light sheet ($NA_{ill} = 0.1$) with the refracting embryo model, propagating along y with its focus at the center of the lateral plane orthogonal to z_F as seen from the camera (*cf.* Fig. 3.11b left). From this, we obtained the fluorescence excitation field at every point in the volume as the fluorophore signal density modulated by the scattered sheet intensity $f(x, y, z) \cdot u_{ill}(x, y, z)$. Next, for the same z_F a full set of detection PSFs is obtained by propagating light from multiplexed, diffraction-limited point-sources orthogonally through an idealized, refocusing lens towards the camera as described in Section 3.2.6. A numerical aperture of the detection objective of $NA_{det} = 0.7$ was assumed. In this way, we obtained a 32×16 grid of spatially varying, volumetric PSFs at z_F with position-dependent aberrations that stem from distortions and scattering in both the illumination and the detection paths (Fig. 3.11b right). Convolving the exhaustively sampled sets of spatially-varying PSFs with the fluorescent object finally yields the wave-optical image as seen by the camera at axial

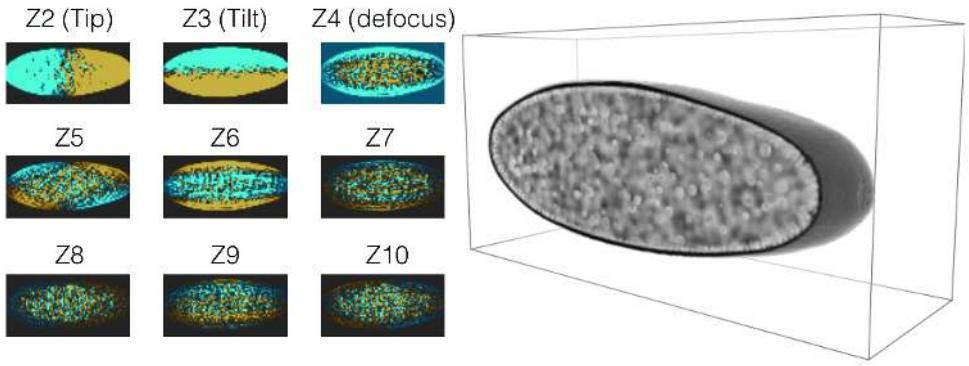


Figure 3.12: Calculating of aberrations of the detection point spread function (PSF) for a given z plane within a synthetic tissue model. The model's physical size is ($200\mu m$, $100\mu m$, $100\mu m$) and the dimensions of the computational grid are $1024 \times 512 \times 512$. The detection wavelength is $\lambda = 522nm$, the numerical aperture is $NA = 0.5$ and the aqueous immersion medium has a refractive index of $n_0 = 1.33$. The refractive index distribution of the tissue model mimics an eggshell, cell nuclei and granular random fluctuations within the biological plausible range of $n \in (1.35, 1.43)$. (Reproduced with permission from Weigert et al. 2018b, Creative Commons license CC BY 4.0)

Video 5

position z_F Fig. 3.11b right). Repeating these steps for a desired number of focal planes (we chose $n_z = 128$) results in a volumetric image stack as would be recorded by a microscope with the assumed specifications (*cf.* Supp. Video 5). The large grid size of $1024 \times 2048 \times 1024$ and sampling density requirements (typically $0.5 - 5\mu m$), resulted in a total of $\sim 10^6$ PSF that were calculated in well under 15 minutes on a single graphics card. In contrast, the non-multiplexed calculations would have required more than two months on a single CPU. As a result of this computational pipeline, *biobeam* generates faithfully calculated 3D microscopy data sets that account for both refraction and diffraction based imaging artifacts. These include image blur, contrast loss, and the spatially varying PSFs and tissue induced aberrations (Fig. 3.11b & c).

Aberration maps Often (*e.g.* for investigating adaptive optics strategies) it is desirable not only to simulate sets of PSFs distributed over a sample, but to directly extract spatial maps of aberrations that would result when imaging a tissue or embryo that is represented in a refractive index model. Given the efficiently multiplexed determination of PSF as described in Section 3.2.5, also aberration terms can efficiently be calculated for any point in a sample in the same manner. The main benefit again is that a single multiplexed simulation through the tissue is sufficient. Let $h(x - x_0, y - y_0, z_F)$ be the lateral slice of the PSF associated to a given initial position (x_0, y_0) at focal position z_F, then the corresponding pupil function $P(\theta, \phi)$ is given by the angular spectrum (Fourier transform) of $h(x - x_0, y - y_0, z_F)$. Projecting $P(\theta, \phi)$ onto a Zernike[15] basis Z_n yields the associated Zernike terms of the aberrations present at (x_0, y_0, z_F). In Fig. 3.12 we show these aberration calculations for the central focal plane of a tissue model mimicking an embryo with similar refractive index distribution as before. We observe that aberrations terms are distributed as expected, with mirror-symmetric Tip/Tilt (Z_2/Z_3) aberrations, the largest defocus

[15] With the Noll index n. See Appendix A.3 for a list of of zernike modes.

term (Z_4) located at the center, and higher order aberrations (*e.g.* astigmatism Z_5, Z_6 and coma Z_7, Z_8) exhibiting mirror/point symmetry as expected (Tao et al. 2012).

3.4.2 Simulation of Aberration Correlations in Scattering Tissue

An important aspect in adaptive optics microscopy is the size of the *isoplanatic patch* , *i.e.* the area over which a given wavefront correction for a single point within a sample stays valid when this point is translated. In the following, we will investigate this behavior of a translated focus in a random scattering medium. First, we will reconstitute *in-silico* the recently found *shift-shift memory effect* that relates the properties of a translated focal spot with the speckle pattern of a plane wave after transmission through the medium, thus validating our method for that important use case. Secondly, we propose a decoupling scheme for correcting wavefronts that we show via simulations to increase the range over which these correlations hold and thus the size of the isoplanatic patch.

Reconstitution of the conventional shift-shift memory effect It is well established and exploited in imaging that perfect imaging foci can be created behind strongly scattering screens (Vellekoop et al. 2007) and within biological samples by appropriately shaping the wavefront before entry into the scattering medium. Here, the *shift-shift memory effect* (Judkewitz et al. 2015) describes the behavior of an aberration corrected wavefront that forms a diffraction limited focal spot when laterally translated within a scattering tissue. Given a fixed position g, such focus can be created by pre-shaping the incoming wavefront with the correct wavefront aberration terms ϕ_g (Vellekoop et al. 2007). When translated laterally by a length Δx, the quality of the resulting spot at the translated position quickly deteriorates, albeit in a precise manner: The correlation coefficient as a function of Δx between the original and the laterally shifted focal spot is equal to the autocorrelation function of the speckle pattern resulting from an incident plane wave at the same depth (Judkewitz et al. 2015).

To reconstitute this effect *in-silico*, we first created synthetic refractive index distributions with different depths ($L_x / L_y = 100\mu m, L_z = 20 \dots 80\mu m$) and with refractive index variation of $n = 1.36 \pm 0.03$ by using generated Perlin noise of the given variation. Next, diffraction limited guide stars were propagated from inside the tissue model towards the surface, where the guide-star specific aberration patterns ϕ_g were recorded. Phase conjugation of these fields at the tissue surfaces leads to the precise recovery of the initial diffraction limited focal spot inside the tissue (Fig. 3.13a). The memory effect can then be investigated by translating ϕ_g by Δx and propagating the translated complex field through the tissue again (Fig. 3.13b). The focus quality for a given Δx was then determined as the normalized spatial correlation $C(\Delta x)$ of the back-shifted transmitted complex field $u(x + \Delta x, y, z_0)$ with the original focused field $u_0(x, y, z_0)$. In Fig. 3.13c we show $C(\Delta x)$ for three different tissue depth, corresponding to 1 ($20\mu m$), 2 ($40\mu m$), and 4 ($80\mu m$) mean free path lengths. For all simulations we used the average of 50 different guide stars at 150 different

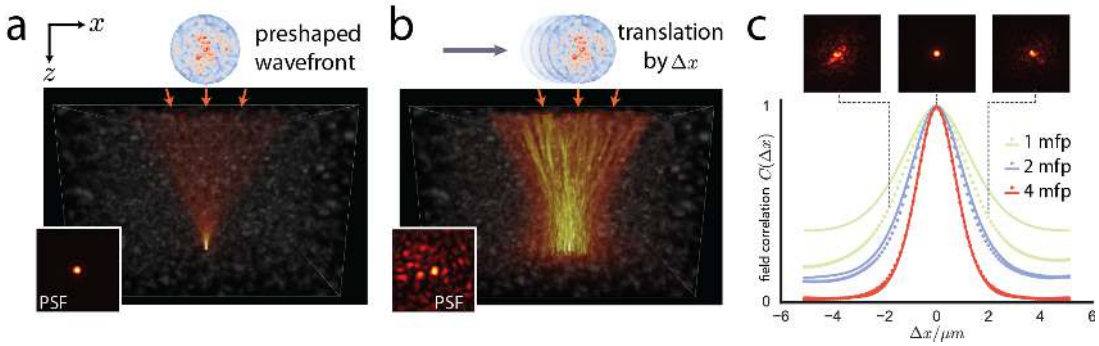


Figure 3.13: Simulation of shift-shift memory effect: a) Simulation of guide-star assisted diffraction-limited focusing in scattering tissue model with a pre-shaped wavefront that compensates the tissue induced aberrations. b) Lateral translation of this aberration compensated beam leads to gradual degradation. c) Quantification of focus degradation via its correlation function $C(\Delta x)$ wrt to the original sharp focus a function of translated distance (dotted line) for 3 different tissue models of increasing depth (1 mfp/20μm, 2 mfp/40μm, and 4 mfp/80μm, with mean free path length mfp). We also show the auto-correlation function of the speckle pattern arising from a plane wave propagating through the same tissue (solid line). The agreement of both confirms the findings of (Judkewitz et al. 2015). This agreement is most noticeable deep inside the tissue ($L_z > 1$ mfp). (Reproduced with permission from Weigert et al. 2018b, Creative Commons license CC BY 4.0)

translation position Δx, which for the simulated 3 different tissue depths comprised 22500 individual field propagation simulations. As experimentally and theoretically described in Judkewitz et al. 2015 we found $C(\Delta x)$ to be similar (for $L_z = 1$ mfp) or virtually identical ($L_z > 1$ mfp) to the absolute value of the spatial auto-correlation of a speckle field that results from an incident plane-wave. In the same way, the width of $C(\Delta x)$ and with it the size of the isoplanatic patch decreases with increasing tissue depth, as expected. This *in-silico* reconstitution shows, that our simulations can faithfully reproduce an emergent wave-optical phenomenon responsible for the significant robustness of adaptive imaging against lateral focus displacements.

Extended memory effect We now show that the decoupling of *site-specific* and *translational-invariant* aberrations results in an *extended memory effect* that might be used for adaptive optics strategies with fewer guide-stars. Compared to the conventional case in which the full aberration pattern ϕ_g is co-translated with the shifted beam (Fig. 3.14a), we hypothesized that a strategy in which a site-specific aberration pattern remains statically conjugated with the surface of the sample would yield an increased range of focus persistence (Fig. 3.14b). In detail, we decompose the wavefront $\phi_g = \phi_{local} + \phi_{invariant}$, where the translation invariant part $\phi_{invariant}$ is the average aberration correction across the lateral dimension (*e.g.* global defocus or tip/tilt terms). Moving the focus is then achieved by only translating $\phi_{invariant}$, while leaving ϕ_{local} constant. By calculating the spatial focus correlations $C(\Delta x)$ as before, we found that the length over which the focal spot sustained its compact shape was significantly increased compared to the conventional shift-shift memory effect (Fig. 3.14c). This effect is visible at various penetration depths, but becomes most pronounced beyond two mean-free-paths (MFPs). At 2.8 MFP the size of the isoplanatic patch, evaluated as the area-integrated Strehl ratio, was extended in all

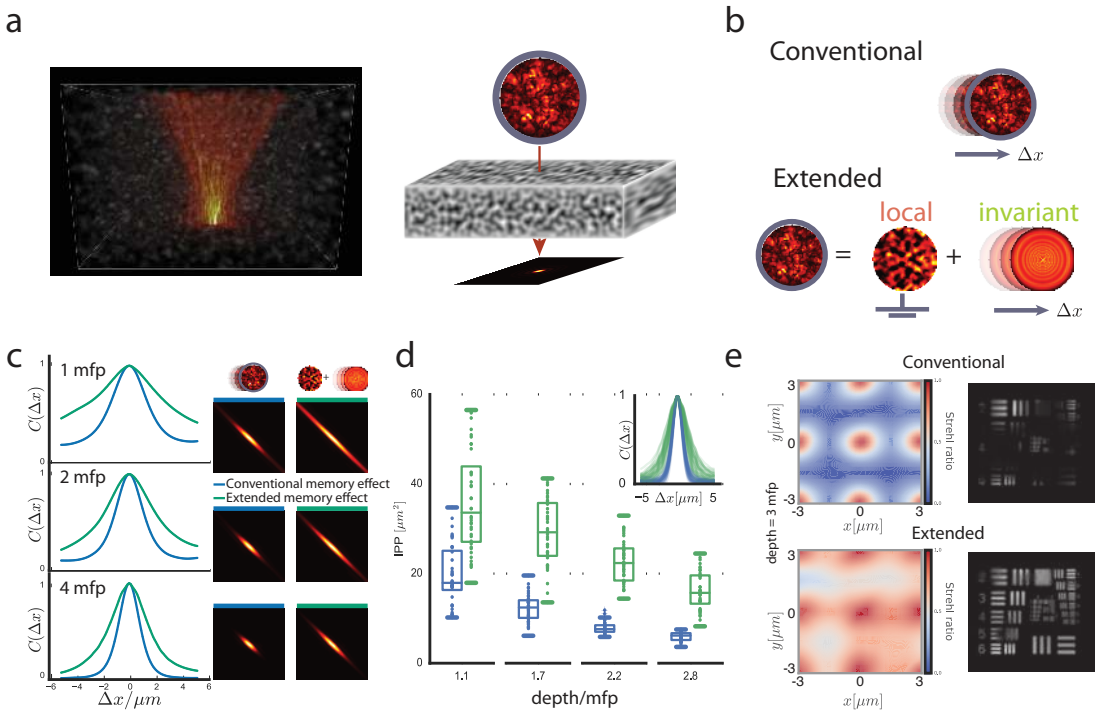


Figure 3.14: Proposed extension of conventional shift-shift memory effect. a) Guide-star assisted diffraction-limited focusing in scattering tissue model enabled by FIXEM. b) Extended memory effect: Decoupling of site-specific local and translational-invariant aberrations, which c) leads to improved focus correlations (left) and focus persistence (right) with shift distance Δx. d) The resulting area increase of isoplanatic patch (IPP) is most reliable at high penetration depths, and e) relaxes the required density of guide stars as illustrated for adaptive imaging.

simulation instances with a mean area gain of 2.7 compared to the conventional shift-shift memory effect (Fig. 3.14d). This proposed extension could be particularly relevant for adaptive deep-tissue imaging where an increase of the isoplanatic patch-size for a fixed Strehl ratio implies a reduction in the required sampling density of embedded or multi-photon induced guide stars (Fig. 3.14e).

3.4.3 Scattering Properties of Retinal Tissue

As a final application example, we demonstrate the potential of fast light simulations to elucidate the optical properties of retinal tissue. An intriguing property of the vertebrate eye is its peculiar structure: Before light is detected by the outer segments (OS) of the photoreceptor cells[16] at the back of the retina, it first has to travel through $\sim 200\mu m$ of layered neuronal tissue (*cf.* Fig. 3.15a, Solovei et al. 2009; Kreysing et al. 2010). The nuclei of photoreceptor cells (PRC) itself form the particularly densely packed *outer nuclear layer* (ONL, $50 - 60\mu m$ in mouse retina), which constitutes a potential source of severe light scattering thus threatening to impede visual sensitivity by distorting the light incident on the photoreceptors (Kreysing et al. 2010). Interestingly, Solovei et al. 2009 found that the architecture of PRC nuclei in the ONL radically differs between diurnal (*e.g.* pig) and nocturnal (*e.g.* mouse)

[16] *Rods* and color-sensitive *cones*

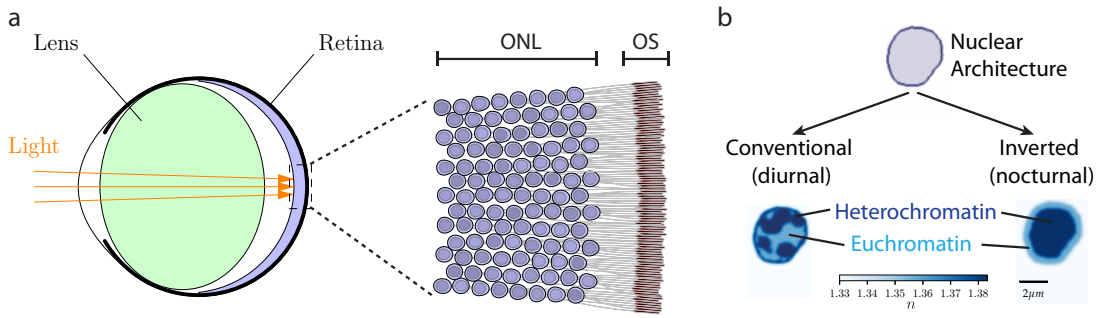


Figure 3.15: Overview of retinal tissue structure: a) Light entering the eye and passing the lens has to traverse several layers of retinal tissue before being detected by the outer segments (OS) of photoreceptor cells (PRC, rods and cones). An important source of potential light scattering is the densely packed *outer nuclear layer* (ONL) composed of PRC nuclei. b) Difference of ONL rod nuclei architecture between diurnal and nocturnal mammals: Optically dense heterochromatin forms chromocenter around a less dense euchromatin core in diurnal animals (conventional architecture), whereas a shell of euchromatin surrounds a heterochromatin core for nocturnal animals (inverted architecture).

mammals: Nuclei of diurnal animals show a *conventional* distribution of chromatin, where chromocenters of condensed and optically denser ($n \approx 1.38$) heterochromatin are distributed around a core of optically less dense ($n \approx 1.35$) euchromatin located at the nuclei center (Fig. 3.15b, Błaszczak et al. 2014). In contrast, nocturnal species exhibit an *inverted* nuclear architecture, where the heterochromatin resides at the nuclei center and is surrounded by euchromatin. This nuclear inversion is unique to rod cells of nocturnal mammals[17] and is thought to be an evolutionary adaption to nocturnal vision (Solovei et al. 2009). Specifically, it has been hypothesized that the chromatin reorganization in the inverted architecture improves the optical transmission property of the ONL by reducing the scattering of incident light. This has been corroborated by light simulations in 2D and 3D with synthetic nuclei models, which found for single nuclei models that the nuclear inversion indeed results in less scattering especially at larger scattering angles (Kreysing et al. 2010; Błaszczak et al. 2014). Due to its computational complexity, these simulations were however confined to single nuclei. An accurate simulation of light propagation and scattering at tissue scale, *e.g.* through realistically large ONL model that capture the *in-situ* geometry of real ONL nuclei arrangements, is therefore highly desirable. We will address this question by *i)* extracting a volumetric ONL model from microscopy acquisitions of mouse retina, *ii)* simulating the light propagation through *conventional* and *inverted* refractive index distributions that are derived from the biological ONL model, and *iii)* calculating the side scattering cross-sections for both cases. That way, we will show that side scattering, especially for large angles, is substantially reduced for the inverted case of chromatin organization.

Extraction of realistic ONL models: To create a 3D model of the nuclei distribution, mouse retina samples were stained (Hoechst 33258) and imaged[18] with a two-photon

[17] At low light levels vision is almost exclusively provided by the more light-sensitive rods.
[18] Sample preparation by Heike Petzold, imaging by Alfonso Garcia Ulloa.

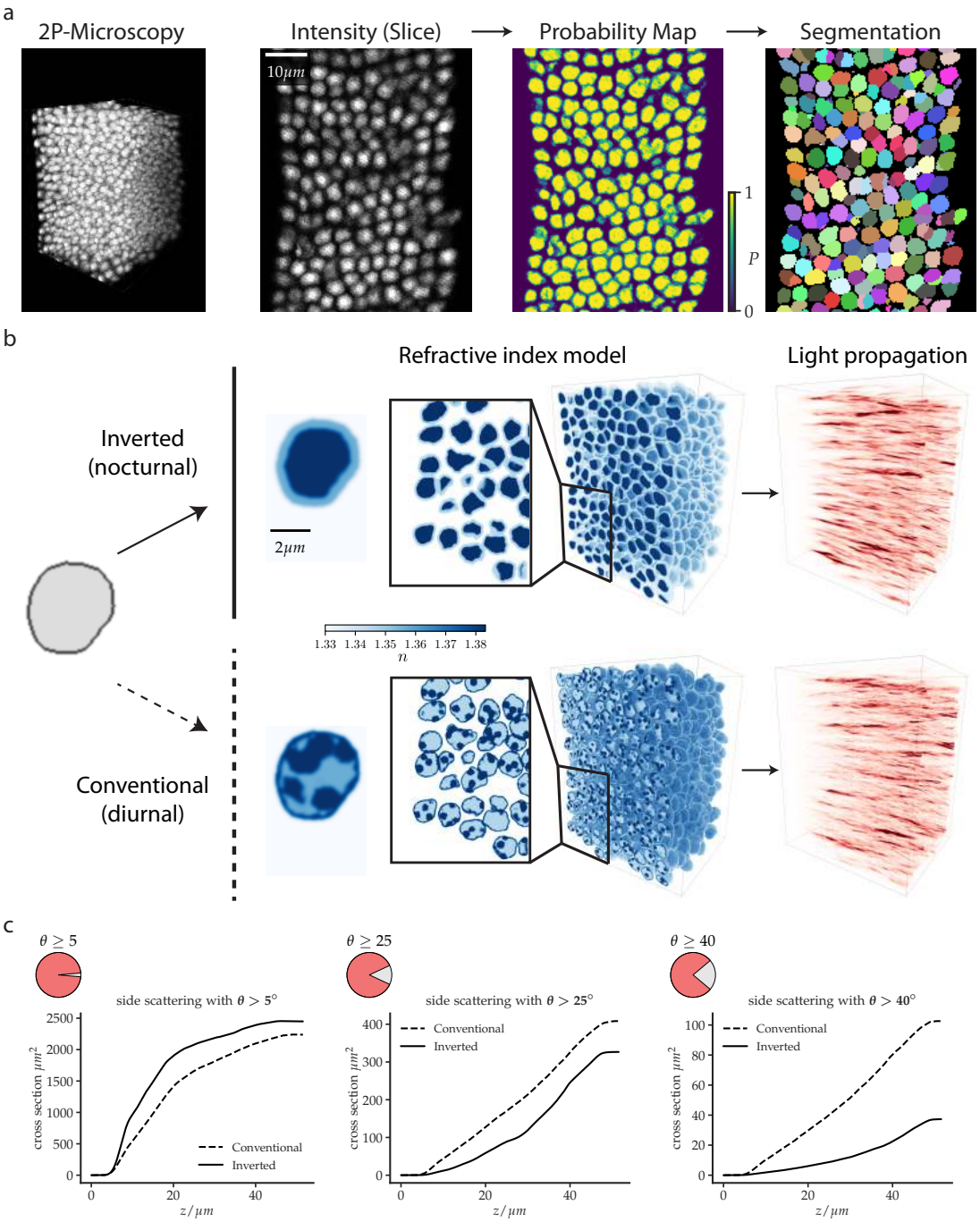


Figure 3.16: Scattering properties of retinal tissue for different nuclear architectures: a) Extraction of realistic tissue models of the outer nuclear layer (ONL) in mouse. From left to right: Two photon microscopy of stained mouse retina. A single slice through the acquired 3D image of size $83 \times 83 \times 250nm$. Probability map of nuclear regions as obtained with a Random-Forest classifier. Final segmentation result with individual nuclei regions via watershed transform. b) Simulation of light propagation. From the segmentation two refractive index models of the ONL were created by assigning each nuclei region either a conventional or inverted refractive index distribution. Afterwards, a plane wave ($\lambda = 500nm$) was propagated through the two different models and the complex field retained. c) Side scattering cross sections for three different cut-off angles $\Theta = 5°, 25°, 40°$ along the propagation axis z for the two models. Only the scattering at angles $\theta > \Theta$ contribute to the respective side scattering cross section. When considering almost all angles ($\Theta = 5°$) scattering for the inverted case is slightly larger than the conventional. In contrast, side scattering for higher angles ($\theta > 25°$ and $\theta > 40°$) is vastly reduced for the inverted case.

confocal microscope (Zeiss LSM 780, 63x/$NA = 1.3$), resulting in an acquired intensity image of $190\times190\times82\mu m$ with pixel-sizes of $83\times83\times250nm$ (Fig. 3.16a). To create a realistic refractive index map of packed nuclei within the ONL, we first segmented the intensity image into nuclei regions. To that end, we trained a random forest classifier via Fiji (Schindelin et al. 2012; Arganda-Carreras et al. 2017) to densely classify each pixel into background or foreground (nuclei), and applied a watershed segmentation (Walt et al. 2014) on the probability map with manually generated seed points, resulting in 1758 individual nuclei instances (Fig. 3.16a). Finally, the refractive index distribution inside each nuclei region was generated according to the two different models:

- *Inverted*: Consisting of two refractive phases with $n_1 = 1.357$ and $n_2 = 1.382$, corresponding to euchromatin and heterochromatin, respectively. Each nuclei mask is split into shell and core regions of equal volume (via morphological shrinking operations on each mask), which are then assigned the respective refractive indices (n_1 for shell, n_2 for core).
- *Conventional*: Here, we randomly picked $8-12$ chromocenters within the nuclei mask and assigned points close to either the nuclei border or those chromocenters to the high refractive index phase n_2 until its joint volume reached half the full nuclei volume. The other points were then assigned the less dense refractive index n_1.

For both inverted and conventional models both refractive phases (n_1 & n_2) occupied the same total volume. We furthermore created isotropic refractive index distributions (pixel-size $83nm$) via bi-linear interpolation and assumed $n_0 = 1.33$ as surrounding medium. Finally, we blurred the resulting refractive index distribution in both cases with a Gaussian kernel of small size ($\sigma = 2$px) to create a smooth distribution (Fig. 3.16b).

Light propagation simulations and side scattering calculation: To simulate the light propagation through both tissue models we used a computational simulation grid of size $1024\times1024\times645$ with isotropic pixel-size $\Delta x = 83nm$. We then propagated a plane wave with wavelength $500nm$ through the two different refractive distribution models corresponding to the conventional and inverted ONL model. We finally calculated for every position along the propagation direction z the side scattering cross sections $\sigma^{\Theta}_{side}(z)$, which represents the amount of light scattered away at an polar angle $\theta > \Theta$ at position z:

$$\sigma^{\Theta}_{side}(z) = \int_0^{2\pi}\int_{\Theta}^{\pi} d\varphi d\theta \, \sin\theta |f(\theta,\varphi,z)|^2$$

Given the complex field $u(x,y,z)$, this integral can be calculated from the angular spectrum $U(k_x,k_y,z)$ as described in Eq. (3.17).
In Fig. 3.16c we show the side scattering contributions for three different cutoff angles: Although the overall scattering ($\theta > 5°$) after full traversal of the tissue

($z = 55\mu m$) is comparable for both the conventional and inverted architecture, the side scattering at larger angles is substantially reduced for the inverted case (80% and 36% of the conventional side scattering for $\theta > 25^\circ$ and $\theta > 40^\circ$ respectively). This demonstrates, that scattering in the inverted case is predominantly forward directional and thus less problematic for loss of visual sensitivity at the retina. In contrast, the conventional architecture suffers from greatly enhanced side scattering, detrimental to the quality of the resulting image.

3.5 Summary

In this chapter, we demonstrated how a GPU accelerated light propagation method can serve as a flexible and particularly powerful platform to systematically study wave-optical phenomena in weakly scattering biological tissues of realistically large size. By introducing a multiplexing scheme of PSF calculations and providing an efficient GPU implementation (*biobeam*), we demonstrated that faithful whole-tissue wave-optical simulations of light-sheet microscopes is possible. Beyond the reproduction and identification of commonly occurring imaging artifacts, we demonstrated the optical capabilities and accuracy of our method by demonstrating that emergent wave-optical phenomena such as the shift-shift memory effect are quantitatively reproduced deep inside tissues and that the range of this effect can be extended by decoupling the shifted incident wavefront. Finally, we demonstrated via light simulation through realistically large retinal tissue models, that the inversion of nuclear architecture in mouse retina results in significant reduced side scattering, thus providing a computational evidence in support of this evolutionary beneficial adaptation in nocturnal mammals. While we chose a variant of BPM as a low-level field stepping routine, the here presented strategy of multiplexing PSF calculation is more general in nature, and may also be used in combination with other light propagation algorithms, *e.g.* in scenarios where higher accuracy at large angles or genuinely isotropic scattering is of relevance (Osnabrugge et al. 2016; S. Schmidt et al. 2016). Prospectively, we see *biobeam* helping to improve microscope design, enhancing data driven deconvolution and segmentation strategies by providing realistic imaging data-sets along with ground-truth data, and paving the way for a new generation of smart, adaptive microscopes that learn to treat the sample as a part of the optical path. Of great importance will be the availability of refractive index maps of embryos, tissues, cells, and sub-cellular compartments which may profit from recent advances in tomographic phase microscopy (Choi et al. 2007; K. Kim et al. 2017; Schürmann et al. 2017), optical diffraction tomography (ODT, Chi et al. 2018), or multi-guide star laser tomography used in astronomy (Hart 2010).

4 Content-Aware Image Restoration (CARE)

Overview

In the previous two chapters, we investigated how volumetric acquisitions degraded by noise and image blur can be improved in real time via low-level image processing and how the disruptive effect of the optical tissue inhomogeneities can be described and studied by fast, multiplexed light simulation. Now, we will reconsider the problem of microscopy image restoration, *i.e.* the problem of recovering the *latent* true image from an imperfect observation by the microscope, yet taking a different, more powerful *data-driven* approach. In fluorescence microscopy the quality at which biological processes can be faithfully recorded is not only determined by the spatial resolution of the used optical device, but also by the desired temporal resolution, the total duration of an experiment, the required imaging depth, the

achievable fluorophore density, bleaching, and photo-toxicity (Icha et al. 2017; Laissue et al. 2017). All these aspects cannot be optimized at the same time – one must make trade-offs, for example, sacrificing signal-to-noise ratio by reducing exposure time in order to gain imaging speed. Such trade-offs are often depicted by a *design-space tetrahedron* (J. B. Pawley 2006; Laissue et al. 2017) that has resolution, speed, phototoxicity, and depth at its four vertices (Fig. 4.1a) with the volume being limited by a total photon budget (Cox et al. 1986). These trade-offs can be addressed by optimizing the microscopy hardware, yet there are physical limits that cannot easily be overcome, and consequently computational procedures to improve the quality of acquired microscopy images are increasingly important. For example, one is often forced to image samples at low signal intensities, resulting in difficult to analyze, low signal-to-noise ratio (SNR) images. One way to improve SNR is to increase laser power or exposure times which, unfortunately, is usually detrimental to the sample, limiting the possible duration of the recording and introducing artifacts due to photo-damage. An alternative solution is to image at low SNR, and later computationally restore acquired images. Classical approaches, such as denoising (Buades et al. 2005), can in principle achieve this, but without leveraging available knowledge about the data at hand. Other examples of sophisticated image restoration algorithms that can push the limit of the design-space tetrahedron include structured illumination microscopy (M. G. Gustafsson 2000; Müller et al. 2016), super-resolution radial symmetry (SRRF) (N. Gustafsson et al. 2016), deconvolution (Richardson 1972; Preibisch et al. 2014), surface projection algorithms (Blasse et al. 2017; Shihavuddin et al. 2017), and denoising methods (Buades et al. 2005; Dabov et al. 2009) that allow one to recover important biological information that would be inaccessible from the raw images alone. Yet most of common image restoration problems, such as denoising or deconvolution, have multiple possible solutions, and require additional prior assumptions or *regularization* in order to select one solution as the final result. These assumptions are typically general, *e.g.* requiring certain level of smoothness of the restored image, and therefore are not dependent on the specific content of the images to be restored. Typical biological imaging experiments are often highly stereotypical, in the sense that the same organism is often studied in several experimental runs, resulting in large experimental image corpora of the same biological structure labeled in an identical way. As a result, image restoration methods that are able to leverage available prior structural knowledge about the biological data under investigation are poised to reach superior restoration results.

Machine learning based methods, such as *deep learning* (DL), provide such an avenue to perform complex tasks by learning from the inherent statistical correlations present in specific data distributions (LeCun et al. 1998; LeCun et al. 2015). In particular, DL employs large multi-layered *artificial neural networks* that can implement intricate input-output functions after being trained on annotated example

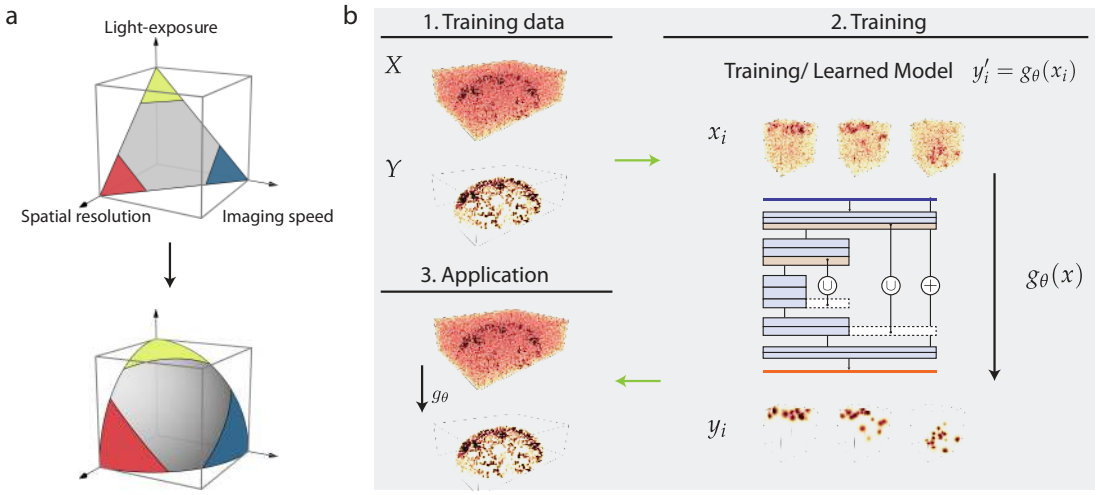


Figure 4.1: Content-aware image restoration for fluorescence microscopy. a) The design-space of fluorescence microscopes entails trade-offs between imaging speed, spatial resolution, and light exposure to best capture a given sample. Image restoration aims to enlarge the design-space by restoring image aspects that suffered due to the trade-off used during imaging. b) Principle of *content-aware image restoration* (CARE) for a low signal-to-noise ratio (SNR) restoration problem: Training data is acquired directly at the microscope, where each training data pair (x_i, y_i) consists of two registered low and high SNR images of the same biological sample. A deep convolutional neural network g_θ is then trained to restore y_i from x_i and applied to previously unseen, potentially very low SNR images yielding restored images. (Adapted with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

data, *i.e.* gold-standard, *ground-truth* data. The use of neural networks as classification and regression models has a long, undulating history. Already McCulloch et al. 1943 proposed a mathematical model of logical propositions based on networks of artificial neurons implementing logical operations. Soon after, Lettvin et al. 1959 discovered the role of local, convolutional neural connections in the retina serving as feature detectors. At the same time one of the earliest neural network architecture successfully used for classification appeared in the form of the *perceptron* (Rosenblatt 1958), a one-layer feed-forward network with linear activation functions, that was able to learn Boolean functions of real vector inputs. The insight provided by Minsky et al. 1969, that due to its linearity any function whose pre-image is not linearly separable (*e.g.* XOR) could *not* be learned was regarded as a fatal blow to the nascent field and led to a rapid decline in research thereafter. Inspired by biological findings regarding the hierarchical structure of cells in the mammalian visual cortex (Hubel et al. 1962) the 80s and 90s saw the emergence of network architectures implementing feature-cascades as in the *Neocognitron* (Fukushima et al. 1982) and convolutional neural networks (LeCun et al. 1990) that enabled basic image recognition tasks. Furthermore it was shown that neuronal networks with *nonlinear* activation functions and at least a single hidden layer are *universal function approximator*, *i.e.* they are able to approximate any continuous function with arbitrarily small error[1] (Cybenko 1989; Hornik et al. 1989), thus evading the fundamental limitation of the earlier perceptron. Application of neural networks to large classification and regression problems

[1] Given enough - *i.e.* potentially exponential many - hidden neurons.

was, however, hampered by their data inefficiency, plagued by convergence issues (*e.g.* the *vanishing gradient problem*, Hochreiter 1991), and they subsequently fell out of favor compared to methods such as support vector machines (SVM, Hearst et al. 1998), random forests (Breiman 2001), and graphical models (Nasrabadi 2007; U. Schmidt et al. 2014). With the advent of general purpose GPU programming, the availability of large sets of image training data (Martin et al. 2001; Deng et al. 2009), and the employment of non-saturating activation function[2], deep (*e.g.* many layered) neural networks regained popularity after winning several computer vision competitions in object classification and detection starting with Cireşan et al. 2012 and Krizhevsky et al. 2012. Since then, the employment of such methods has led to dramatic improvements for example in the classification of natural images into semantic categories (He et al. 2016) or playing classical strategy board-games (Silver et al. 2016). In biology, deep learning based approaches have recently been applied to the automatic extraction of connectomes from large electron microscopy data (T. Beier et al. 2017), for classification of image-based high-content screens (Caicedo et al. 2017), fluorescence signal prediction from label-free images (Christiansen et al. 2018; Ounkomol et al. 2018), resolution enhancement in histopathology (Rivenson et al. 2017), or for single molecule localization in super resolution microscopy (Nehme et al. 2018; Ouyang et al. 2018).

The direct application of DL methods to image restoration tasks in fluorescence microscopy, however, is complicated by *i)* the huge diversity of biological samples, data normalization, imaging artifacts, experimental conditions and acquisition methods, *ii)* the strong requirements to assess the restoration quality, and *iii)* the absence of sufficiently large training data sets.
In this chapter we show how to remedy these problems and demonstrate how DL based models, in particular *residual convolutional neural networks*, can be used for a wide range of common image restoration problems in fluorescence microscopy. Specifically we will show how *i)* volumetric acquisitions of extremely low signal-to-noise ratio can be nearly perfectly recovered, enabling up to 60 fold reduced exposure/laser-power, *ii)* the two dimensional surface of epithelia tissue can be extracted from noisy three dimensional volumes that would allow to image 8-13 times faster thus increasing temporal resolution by the same factor, *iii)* the typically poor axial resolution of microscopes can be restored, enabling up to 10 fold less axial acquisitions, *iv)* neural network based restorations of whole embryo developmental time-lapses can be computed in real-time by interleaving acquisition and multi-GPU inference, and *v)* super-resolved images of tubular structures can be obtained from single widefield images acquired with 100 fold faster imaging speed. In all these examples, we will demonstrate that what we call *content-aware re*storation (CARE) allows to produce results that were previously unobtainable. In the case of low SNR restoration, for example, we acquired pairs of images at low and high signal-to-noise

[2] *E.g.* rectified linear units instead of the then classical sigmoid function.

ratios, use them as input and ground-truth to train CARE networks, and apply the trained networks to remove noise in previously unseen data (*cf.* Fig. 4.1b). That way it is possible, without sacrificing image quality, to use shorter exposure times or less laser-power, in turn enabling acquisitions at higher temporal and/or spatial resolution. We further will propose practical strategies to generate training data in several task depending contexts, specifically when such training image pairs can be *i)* acquired *physically* by changing the acquisition parameters of the microscope, *ii)* created *semi-synthetically* by simulating the degradation process (*e.g.* axial blur), or *iii)* created *fully synthetically* by using an *in-silico* model of both the biological structures and the image formation process. We thus argue, that the application of CARE to biological images allows to transcend the limitations of the design-space tetrahedron (Figure 4.1a), pushing the limits of the possible in fluorescence microscopy through machine learned image computation.

This chapter is structured as follows: In Section 4.1 we will first review some of the related work. Next, we will describe the basis of our approach in Section 4.2 and comprehensively investigate on a simple denoising example of cell images the different design choices that can influence the restoration quality in Section 4.3. We then apply our method on tasks that allow the *physical acquisition* of training data. In particular, we show the restoration of low SNR volumes that commonly appear in live cell imaging exemplified by flatworm and flour beetle acquisitions. In this way, we provide - to the best of our knowledge - the first practical framework to enable live-cell imaging of light sensitive flatworms. We next show in Section 4.4.2 that in the same way low SNR live-cell imagery of fly wing discs acquired with few focal planes can be jointly denoised and almost isotropically upsampled. Along these lines, we demonstrate in Section 4.4.3 the utility of our approach on the more complex, composite task of jointly denoising and projecting the two dimensional epithelia surface of the developing wing of fruit flies and show that biological relevant information can be obtained while reducing light dosage up to 10-fold. In Section 4.4.2 we will give an example of improving axial resolution of low SNR volumes of developing wing-discs of fruit fly larvae from anisotropic acquisitions using 4 fold fewer acquired focal planes and reduced photo-damage. We next describe in Section 4.5 how the inherent resolution anisotropy of common fluorescence microscopy techniques can be improved even without access to isotropic training data by simulating the degrading axial blur and *semi-synthetically* generating training data from the lateral slices of already acquired volumes (IsoNet). We demonstrate the utility of that approach on several examples, ranging from liver tissue acquisitions, to *Drosophila* developmental time-lapses, acquisitions of the developing zebrafish retina, and show that satisfying image quality can be achieved with up to 10 fold less axial acquisitions. Moreover, by interleaving inference and acquisition we give the first example of a neural network based restoration pipeline that runs in real-time on a light-sheet microscope in Section 4.5.4. Finally, we will investigate in Section 4.6 the utility of

fully synthetically created training data via modeling both biological structures and the image formation and demonstrate how this approach can be used to restore super-resolved images of microtubules and secretory granules from single wide-field images acquired with up to 100 fold faster imaging speed. We close with a discussion of the limitation of data-driven learned approaches to image restoration in Section 4.7.

Contributions

1. Introducing a deep learning based image restoration method for fluorescence microscopy via *content-aware image restoration* networks (CARE).
2. Proposing different strategies to create training data *physically*, *semi-synthetically* , or *fully synthetically*.
3. Recovery of extremely low signal-to-noise ratio with up to 60 fold reduced exposure/laser-power, enabling live-cell imaging of the light sensitive species *Schmidtea mediterranea*.
4. Joint denoising and surface projection of *Drosophila* epithelia tissue that allows to image 6-10 times faster.
5. Proposing a novel recovery strategy (IsoNet) for the typically poor axial resolution of microscopes enabling up to 10 fold less axial acquisitions.
6. Demonstrating the real-time ability of a neural network based whole tissue restoration pipeline running on a custom light-sheet microscope.
7. Resolving diffraction-limited structures such as microtubules and granules from wide-field images at 20-times higher frame-rates compared to state-of-the-art methods.

Significant parts of this chapter are published in Weigert et al. 2017, Weigert et al. 2018a, and (Sui et al. 2018).

4.1 Related Work

One of the earliest uses of neural networks for image restoration tasks were given by Zhou et al. 1988 and Paik et al. 1992 that used *Hopfield networks* to recover images corrupted by noise and motion blur. After the resurgence of *deep learning* that was made possible by commoditization of parallel computing via GPUs and the increased abundance of natural image corpora (Martin et al. 2001; Deng et al. 2009), many new neural network based methods have been proposed to perform image restoration tasks on natural images. One of the first applications demonstrating superior performance of trained neural networks compared to non-learned methods have been given by Jain et al. 2008 and Burger et al. 2012, where *convolutional neural networks* (CNN) and *multi-layer perceptrons* (MLP) respectively were used to denoise images corrupted by artificial Gaussian noise. Since then, advanced network architectures *e.g.* using *symmetric skip connections* (Long et al. 2015; Ronneberger et al. 2015) constitute the basis of state-of-the art methods for denoising and upsampling/super-resolution of natural images (Mao et al. 2016; K. Zhang et al. 2017a). This includes

denoising strategies that include only neural network based priors (K. Zhang et al. 2017b; Ulyanov et al. 2018) or that rely on corrupted images only as training data, such as *noise2noise* (Lehtinen et al. 2018). Additionally *conditional generative adversarial networks* (GANs, Goodfellow et al. 2014) have been employed as general image-to-image translation methods (*pix2pix*, Isola et al. 2017), and have found applications *e.g.* for deconvolution in astronomy (Schawinski et al. 2017) or in super-resolution microscopy (Ouyang et al. 2018). In the life sciences, neural network based methods have recently been adopted to a variety of different image restoration problems such as resolution enhancement in histopathology (Rivenson et al. 2017), for single molecule localization in super resolution microscopy (Boyd et al. 2018; Nehme et al. 2018), or for the restoration of medical images from MRI (Oktay et al. 2016; S. Wang et al. 2016), computer tomography (Pelt et al. 2013; Jin et al. 2017), and electron microscopy (Heinrich et al. 2017). Deep learning based methods for general image restoration problems in fluorescence microscopy have been, however, so far missing from the literature.

4.2 Image Restoration with Convolutional Neural Networks

Image restoration, in general, is the problem of recovering the true (latent) image given a distorted (corrupted) observation of it. In the context of fluorescent microscopy, these degradations include camera and photon noise, the image blur incurred the optical point-spread-function (PSF), axial resolution loss due to undersampling of focal planes, or the optical aberrations induced by the heterogeneity of the tissue. The distorted, observed image x can be regarded of being the result of a *forward model* $x = f(y)$ applied to the true (latent) image y. The problem then is to recover y, given x. Typical image restoration tasks with their corresponding forward models are:

- Denoising: $f(y_i) = \mathcal{P}(x_i) + \eta$ (with Poisson photon shot-noise $\mathcal{P}$ and Gaussian camera noise η)
- Deconvolution: $f(y) = h \otimes x$ (with the microscope PSF h)
- Upsampling: $f(y) = \mathcal{S}_\sigma(x)$ (with a subsampling factor σ)

Although f can often be efficiently evaluated, the inverse $y = f^{-1}(x)$ is typically hard to compute, and is commonly not even uniquely defined (*i.e.* there are a multitude of potential latent images). A classical approach to solve this problem is to find the minimizer of an energy $\mathcal{E}$

$$\tilde{y} = \arg\min_y \mathcal{E}(y|x), \quad \text{where} \quad \mathcal{E}(y|x) = \mathcal{D}_f(x, y) + \Phi(y)$$

that is composed of a *data term* D that encourages the restored image to comply with the forward model[3], and a *regularization term* Φ that favors solutions based on

[3] *E.g.* a L_2 loss, *i.e.* $\mathcal{D}_f(x, y) = \|x - f(y)\|^2$

Task	$\mathcal{X}$	$\mathcal{Y}$	typical loss	$\mathcal{L}(y, \tilde{y})$
Classification	$\mathbb{R}^n$	$\{1, \dots, K\}$	categorial cross entropy	$\sum_i y_i \log \tilde{y}_i$
Segmentation	$\mathbb{R}^n$	$\{0, 1\}^n$	binary cross entropy	$y \log \tilde{y} + (1 - y) \log (1 - \tilde{y})$
Regression	$\mathbb{R}^n$	$\mathbb{R}^n$	mean squared error	$\frac{1}{n} \sum_i (y_i - \tilde{y}_i)^2$

Table 4.1: Overview of image based supervised learning: Nomenclature of different tasks, the input/output spaces and the typical loss function used.

prior assumptions about the latent image[4]. Although this approach can work well in many cases, it has major problems. First, the forward model (and thus the data term) is sometimes not fully accurate by neglecting to model some sources of image corruption which are either unknown or hard to describe mathematically. Second, and more important, the regularization term typically encodes rather *weak* assumptions about the latent image, often simply favoring smooth images by penalizing strong differences of neighboring pixels. Third, minimizing the cost function $\mathcal{E}$ is often computationally demanding, resulting in long runtimes or the need for approximations to find solutions more quickly. In fact, this issue is connected to the first two problems since data and regularization terms are often modified to render the minimization of $\mathcal{E}$ easier. In the following, we will take an alternative approach and directly *learn* a function g that computes the restored image $\tilde{y} = g(x) \approx \bar{f}^{-1}(x)$ for input image x, thereby effectively approximating the inverse of the true forward model $\bar{f}$ (which may be unknown). We will do that via *supervised learning* on the basis of a sufficiently large set of pairs (x, y) of corrupted images x and their corresponding true images y, and by implementing g_θ as an *artificial neural network* with learned parameters θ.

Supervised learning

Supervised learning is one of the paradigmatic problems within the field of machine learning (Goodfellow et al. 2016). Here, one is given example pairs $(x_i, y_i)_{i=1}^N$ of N inputs $x_i \in \mathcal{X}$ and outputs $y_i \in \mathcal{Y}$, with the task to find a mapping $g : \mathcal{X} \mapsto \mathcal{Y}$ that approximates the given data points as best as possible according to a defined error metric. In the case of image analysis, the input space will typically be the space of all real valued images $\mathcal{X} = \mathbb{R}^n$ with the n being the number of image pixels. The output space $\mathcal{Y}$ is then task dependent, for instance is chosen as $\mathcal{Y} = \{1, \dots, K\}$ for *image classification* into K classes, $\mathcal{Y} = \{0, 1\}^n$ for *dense segmentation* of each image pixel into two (*e.g.* foreground and background) classes, or an image of the same shape $\mathcal{Y} = \mathbb{R}^n$ as in *regression* (*cf.* Table 4.1). Afterwards, the sought after functional relationship will be approximated by a function g_θ with parameters θ that are chosen to minimize the *expected empirical risk* $\mathcal{R}(\theta)$

$$\mathcal{R}(\theta) = \frac{1}{N} \sum_{i=1}^{N} L(y_i, g_\theta(x_i)) \tag{4.1}$$

[4] *E.g.* Total-Variation prior $\Phi(y) = |\nabla y|$ enforcing sparse gradients

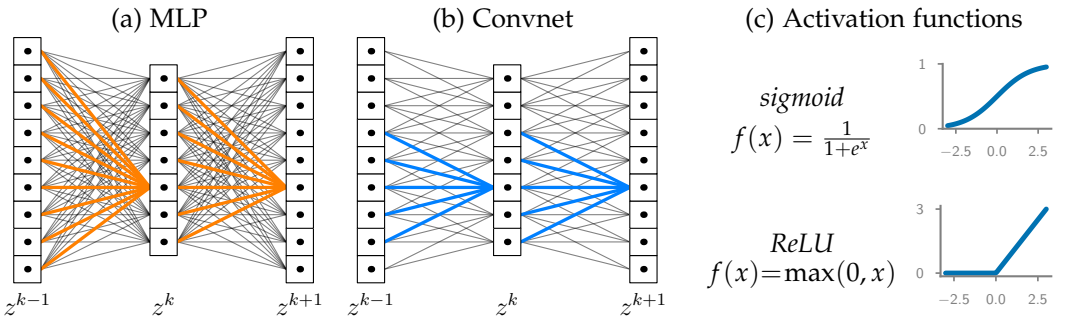


Figure 4.2: Different connectivity pattern of neural networks: a) Multi-Layer-Perceptrons (MLPs) with densely connected layers. b) Convolutional Neural Networks (CNN) with implement convolutions with local support. c) Different common non-linear activation functions.

with a suitable *loss function* $L(y, \tilde{y})$, *e.g. categorial cross entropy* loss for classification or the *mean squared error* loss for regression (*cf.* Table 4.1).

Artifical and Convolutional Neural Networks

An artificial neural network (ANN) is a specific way to implement g_θ, consisting of stacked[5] layers $L^1, \ldots, L^D$ that compute the *activations* of intermediate neurons $z^1, \ldots, z^D$ in a successive way from the first input layer $x := z^1$ towards the output layer $\tilde{y} := z^D$ (Goodfellow et al. 2016). For each given layer L_k the new activations $z^k \in \mathbb{R}^m$ are calculated from the activations z^{k-1} at the layer before, via a linear transformation according to a *weight matrix* W^k, an addition of *bias terms* b^k followed by the application of a (nonlinear) *activation function* f^k:

$$z_i^k = f^k(\sum_j W_{ij}^k z_j^{k-1} + b_i^k) \tag{4.2}$$

The structure of the non-zero entries of W^k define the dependency of each neuron from the neurons of the preceding layer. If W^k is a dense matrix, it is called a *dense layer* and the resulting stacked model is called a *multi-layer-perceptron* (MLP, Haykin 1994). If W^k is a circulant, sparse matrix the resulting network is called *convolutional neural network* (CNN, LeCun et al. 1998, *cf.* Fig. 4.2). In the case of a CNN each linear transformation mediated by W^k are local, translation invariant *convolutions* for which typically small *filters* are used, *e.g.* $3 \times 3 \times c$ pixels for a 2D image with c channels. The internal convolutional layers of the CNN are also called *feature layers*, because they often represent certain structural image features (*e.g.* edges or corners). Additionally, an *encoder-decoder* architecture is often employed, which has been found to increase the modeling power of CNNs (Vincent et al. 2008). To that end, internal feature layers are first compressed in the encoder part by reducing their spatial resolution with *pooling* layers (typically by taking the maximum in a local neighborhood), hoping that this will lead to better (more high-level) features of the image. The decoder part of the network uses *upsampling* to successively enlarge

[5] We only describe the most important case of a *feed-forward neural network* here.

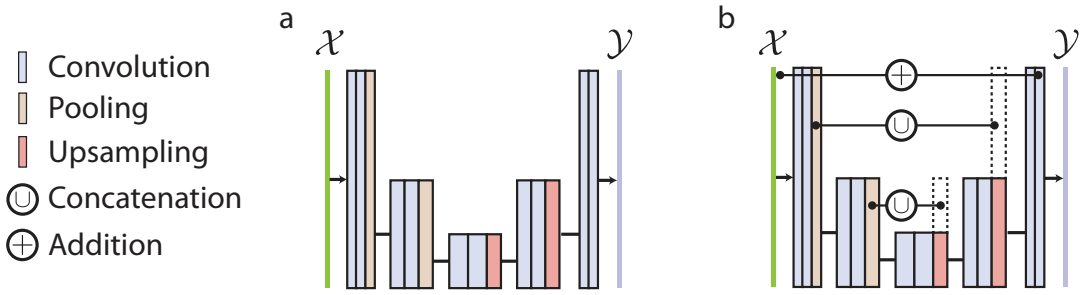


Figure 4.3: Different architectures of convolutional neural networks. a) Encoder-Decoder network, where interleaved convolutional and pooling layers first extract high-level features to a bottleneck layer, which then is subsequently enlarged via upsampling layers to yield the final image. b) Skip connections via concatenation layers help to restore high resolution details as pioneered by (Eigen et al. 2015), and *U-Net* (Ronneberger et al. 2015). Adding a a residual/addition layer facilitates convergence during training (He et al. 2016).

feature layers to eventually yield an output that is typically of the same resolution as the input. However, using upsampling layers can lead to blurred CNN outputs that cannot delineate object boundaries well anymore. Using high-resolution feature layers in the decoder part can remedy this issue. One approach to do this, that we also adopt in this work, is to use *skip connections* that concatenate earlier feature layers from the encoder part with the upsampled later layers in the decoder part (Eigen et al. 2015 and *U-Net* by Ronneberger et al. 2015). Finally, it has often been shown advantageous to not directly learn the output of interest (such as the restored image), but instead the *residual* $\tilde{g}$ to the input of the network (He et al. 2016), *i.e.* $g(x) = x + \tilde{g}(x)$, which can be implemented by a simple addition operation at the end of the network architecture. It is this topology of a residual convolutional network with skip connections, which we will use extensively.

Optimization and neural network training

As we described before a CNN implements a function $g_\theta : \mathbb{R}^N \mapsto \mathbb{R}^N$ defined via the parameters θ of all its layers, *i.e.* the union of parameters of all convolution filters W^k and its bias terms b^k. Given training data comprised of N input-output image pairs $\{(x_i, y_i)\}_{i=1}^N$, the parameters are typically chosen by minimizing the empirical risk or *training loss*

$$\theta^* = \arg\min_\theta \mathcal{R}(\theta) = \arg\min_\theta \frac{1}{N} \sum_{i=1}^{N} L(y_i, g_\theta(x_i)) \tag{4.3}$$

that assigns a cost to each prediction $g_\theta(x_i)$ via the loss function L based on the knowledge that y_i is the correct output. This minimization problem is in general non-convex, thus the resulting loss landscape can be extremely complex admitting potentially many critical points, and a general optimization theory is still missing from the literature (Choromanska et al. 2015). Nevertheless, it is currently accepted as best practice to use a variant of *stochastic gradient descent* (SGD, Goodfellow et al. 2016) as optimization method, where the parameters are repeatedly updated based

on the gradient of the loss evaluated only on random subsets of the training data, called *mini-batches* of typical small size N_b (*e.g.* $N_b = 16$ images):

$$\theta_{n+1} = \theta_n - \alpha \frac{1}{N_b} \sum_{i=1}^{N_b} \nabla_\theta L(y_i, g_\theta(x_i)) \qquad \text{(SGD)} \tag{4.4}$$

Crucially, the topology of the directed computational graph of feed-forward neural networks allows the efficient calculation of the gradient $\nabla_\theta L$ via the recursive application of the chain rule to each layer. The resulting method is called *reverse automatic differentiation* (Linnainmaa 1970) or *back-propagation* (Rumelhart et al. 1986) and constitutes the numerical backbone of efficient gradient based methods for neural network optimization[6]. Furthermore, an *epoch* denotes the aggregate of such optimization steps that has processed all of the available training data once, and *learning rate* designates the step-size α of the gradient-based parameter updates in Eq. (4.4). Apart from plain SGD, several improved variants have been proposed that use adaptive learning-rates and employ momentum, as done for instance with AdaDelta (Zeiler 2012) or Adam (Kingma et al. 2015). Finally, to evaluate the model performance in an unbiased way, it is standard practice to split the initial body of data $\{(x_i, y_i)\}_{i=1}^N$ into three distinct sub sets: *training set*, *validation set*, and *test set*. Here, only the training set is used during optimization, whereas the validation set is used to monitor the generalization ability during training, to detect over-fitting, and to select the final model. The final reported error is then calculated on the held out test set.

Loss functions and probabilistic loss

As we exclusively consider image restoration via regression, the loss function L will be typically the mean squared error (MSE) or the mean absolute error (MAE)

$$L(y, \tilde{y}) = \|y - \tilde{y}\|^2 \quad \text{(MSE)} \qquad\qquad L(y, \tilde{y}) = |y - \tilde{y}| \quad \text{(MAE)}. \tag{4.5}$$

Later we additionally will investigate to predict a pixel-wise *distribution*, consisting of the *location* μ and *scale* σ of a Laplace probability density function

$$p(z; \mu, \sigma) = \exp(-\tfrac{1}{\sigma}|z - \mu|) / (2\sigma) \tag{4.6}$$

and will use as loss function the negative log-likelihood

$$L(y, \mu, \sigma) = -\log p(y; \mu, \sigma) = \tfrac{1}{\sigma}|z - \mu| + \log 2\sigma \tag{4.7}$$

which we call *probabilistic loss*.

[6] Efficient in the sense, that the function evaluation of $L(\theta)$ and the gradient evaluation $\nabla_\theta L$ have the *same* computational complexity $\sim \mathcal{O}(P)$ with $P = |\theta|$ being the number of parameters.

Image normalization and restoration quality metrics

For both training and prediction it is important to normalize the input images to a common range, which is typically done by scaling the input to a given minimum and maximum value. Microscopy images, however, often exhibit isolated pixels with extremely dim and bright values that do not represent the actual image content (*dead* and *hot* pixels of the camera). Hence, we use a simple *percentile*-based normalization instead, which we define for an image $y \in \mathbb{R}^N$ as

$$\mathrm{N}(y; p_{\mathrm{low}}, p_{\mathrm{high}}) = \frac{y - \mathrm{perc}(y, p_{\mathrm{low}})}{\mathrm{perc}(y, p_{\mathrm{high}}) - \mathrm{perc}(y, p_{\mathrm{low}})}, \tag{4.8}$$

where $\mathrm{perc}(y, p)$ is the p-th percentile of all pixel values of y. Typically values are $p_{\mathrm{low}} \in (1, 3)$ and $p_{\mathrm{high}} \in (99.5, 99.9)$. Since the corrupted input image x, the restored image $\tilde{y}$, and the corresponding ground-truth image y typically differ considerably in their dynamic range of pixel values, we cannot simply use standard image quality metrics (*e.g.* PSNR, SSIM) without normalizing the images first. To that end, we first normalize the ground-truth image y by using Eq. (4.8) with $p_{\mathrm{low}} = 0.1$ and $p_{\mathrm{high}} = 99.9$. Secondly, for any image u we wish to compare against y, we apply an affine transformation $\chi(u) = \alpha u + \beta$ that scales and translates every pixel of u based on parameters

$$\alpha, \beta = \underset{\alpha', \beta'}{\arg\min} \left\| (\alpha' u + \beta') - \mathrm{N}(y, 0.1, 99.9) \right\|^2 \tag{4.9}$$

In other words, α, β are chosen such that the *mean squared error (MSE)* between $\chi(u) = \alpha u + \beta$ and $\mathrm{N}(y, 0.1, 99.9)$ is minimal. Note that α, β can be easily computed in closed form (*cf.* Appendix A.5). Based on above definitions, we define the *normalized root-mean-square error (NRMSE)* between the ground-truth y and the restored image $\tilde{y}$ as follows:

$$\mathrm{NRMSE}(y, \tilde{y}) = \sqrt{\mathrm{MSE}\big(\chi(\tilde{y}), \mathrm{N}(y, 0.1, 99.9)\big)}. \tag{4.10}$$

Furthermore, we will apply the same normalization scheme to y and $\tilde{y}$ before we evaluate other metrics *e.g. structural similarity index (SSIM)* (Z. Wang et al. 2004), which measures the perceived similarity between two images ($\mathrm{SSIM} \in (0, 1)$, with $\mathrm{SSIM} = 1$ designating perfect similarity), or *peak signal to noise ratio* PSNR

$$\mathrm{PSNR}(y, \tilde{y}) = 20 \log_{10} I_0 - 20 \log_{10} \mathrm{NRMSE}(y, \tilde{y}) \tag{4.11}$$

with I_0 being the dynamic range of the images (which after normalization is set to $I_0 = 1$).

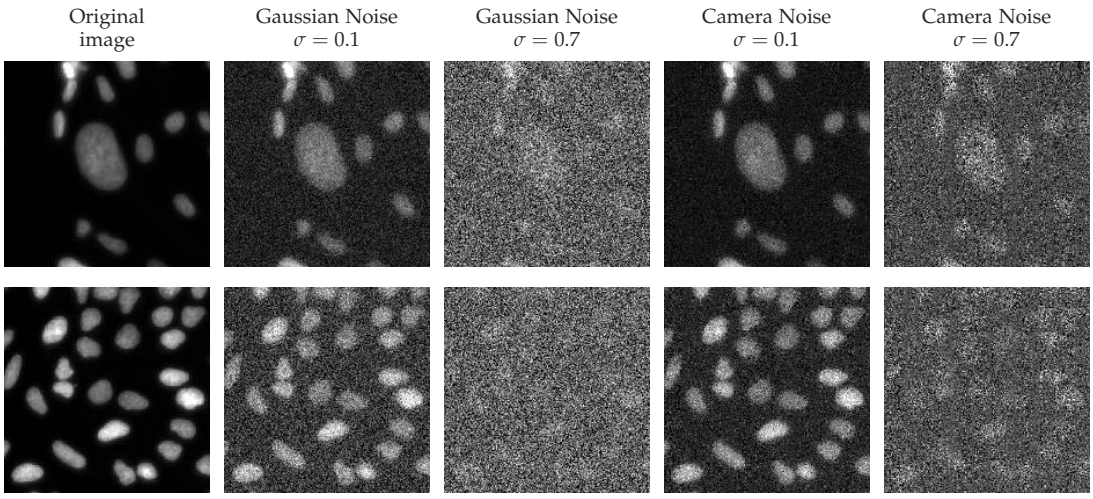


Figure 4.4: Images used for 2D denoising experiments: We show two examples of clean 2D cell images of size 256×256 (left column), and 4 different variants with applied Gaussian noise and Camera noise and different noise strength σ.

4.3 An Educational Example: Denoising of 2D Cell Images

Before we turn our attention to the restoration of complex volumetric acquisitions of biological samples, we first demonstrate on a simple example the utility of the method for fluorescence microscopy. In particular, we will consider the case of *denoising of two dimensional cell images*. This way, we will be able to comprehensively study the implications of network design and hyperparameters choices on the restoration quality, allowing informed design decision later on, where the computational complexity often is prohibitive for an extensive search of the model design space (as would be the case for large volumetric reconstruction tasks). In the following, we will first compare results obtained via neural network based reconstruction with different popular denoising methods. Next, we will investigate how the restoration quality varies with the size of available training data and varying noise levels. Finally, we will compare a multitude of different neural network hyperparameters by comparing $\approx 3,400$ individually trained, disparate models, from which we deduce practical heuristics for neural network hyperparameter selection.

4.3.1 Restoration Quality and Comparison with other Methods

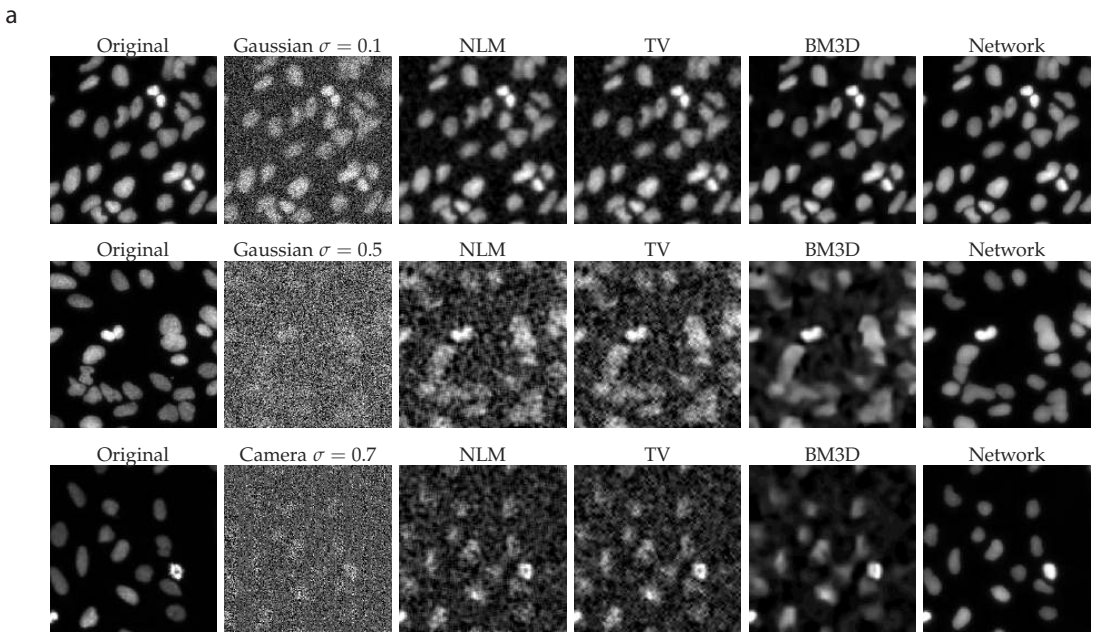
Data set For our experiments, we used the image set BBBC006v1 from the Broad Bioimage Benchmark Collection (Ljosa et al. 2012). This data set contains focused and defocused 2D images of U2OS cells with nuclei staining (Hoechst), from which we selected the 768 in-focus images as our corpus of cell images. From these we extracted 2900 non-overlapping crops of size 256×256 from which we separated 100 images as hold out test set to be used later for model evaluation (Fig. 4.4). We then first normalized the images according to Eq. (4.8) with $p_{\text{low}} = 0.1$ and $p_{\text{high}} = 99.8$ to ensure a common dynamic range. Finally we create different corrupted versions

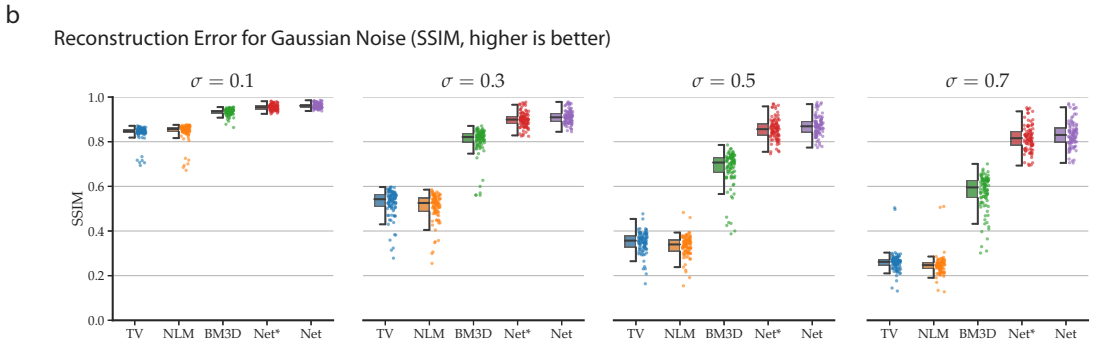
by applying two different noise types: *i) Gaussian noise* of given standard deviation σ, and *ii) Camera Noise* where we created for each image a stripe-like pixel depending gain masks g and applied an intensity dependent noise (with variance $\sim gy\sigma^2$, with y the pixel intensity) mimicking the noise characteristic of a sCMOS camera. For both noise types we created 4 different variants of the image corpus with $\sigma = (0.1, 0.3, 0.5, 0.7)$. Two example images and its different noise corrupted versions are depicted in Fig. 4.4.

Network training As the restoration network we used a residual U-Net as described before, with a depth (*i.e.* number of maxpooling/upsampling layers) of 2, convolutional kernel sizes of 3×3, and an initial number of 32 filters (which get doubled after each pooling operation). This results in a model with a total of 332641 parameters, which is of relatively modest size compared to typical classification networks[7]. To investigate the effect of noise variability, we trained *i) noise specific networks* using training images from only the specific noise level σ and noise type (Gaussian/camera), and *ii)* a *common model* using training images selected form all noise levels and types. Each model was trained for 100 epochs with a mini-batchsize of 16, using MAE loss and the Adam optimizer (Kingma et al. 2015) with a learning rate of $r = 5 \cdot 10^{-4}$. The whole training took 13 min on a single GPU[8].

Restoration quality and comparison with other methods For evaluation, we applied each trained model to the held-out test set of 100 images for all different noise levels and noise types, and compared against several popular and potent classical denoising methods: Non-local-means (NLM, Buades et al. 2005), Total-Variation denoising (TV, Rudin et al. 1992), and BM3D (Dabov et al. 2009). To obtain the best possible competing denoising results, we tuned the parameters (*e.g.* denoising strength) of each classical method via a line-search to yield minimal (MSE) error compared to the ground-truth independently for every single test image, which would be impossible in a more realistic setting. In Fig. 4.5a we depict three example images, its noise corrupted version, and the restoration results from all compared methods. For small noise levels (Gaussian $\sigma = 0.1$) all restoration methods yield results that are visually close to the ground-truth, with BM3D and the noise specific network resulting in considerable less artefacts at background regions. In the case for moderate and high noise levels (Gaussian/camera with $\sigma = 0.5/0.7$), the network reconstruction is of substantially higher quality compared to the other methods. This observation is corroborated quantitatively by computing the restoration error compared to ground-truth for all test images (Fig. 4.5b & c). Both the NRMSE and SSIM are significantly improved in case of the network reconstruction, specifically for higher noise levels: For example in case of Gaussian noise with $\sigma = 0.7$ the noise specific network results exhibited a median SSIM of 0.84, a substantial improvement compared to BM3D (0.60), NLM (0.25), and TV (0.26).

[7] *E.g.* the original U-Net of Ronneberger et al. 2015 used $\sim$ 22 Mio. parameters, while AlexNet of Krizhevsky et al. 2012 used $\sim$60 Mio. parameters.
[8] NVIDIA GTX 1080





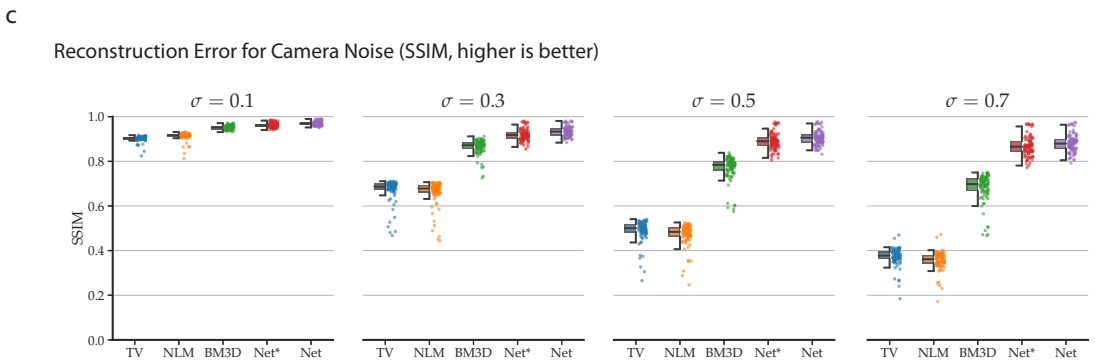


Figure 4.5: Denoising of 2D cell images: a) Three clean example images, their corrupted versions with different noise level and noise type and the restoration results of classical methods Non-loal-means(NLM, Buades et al. 2005), Total-Variation denoising (TV, Rudin et al. 1992), BM3D (Dabov et al. 2009), and of the noise specific neural network models (Network). b) Reconstruction quality for Gaussian noise as quantified by SSIM (Z. Wang et al. 2004, higher is better, SSIM $= 1$ signifies perfect agreement). Both noise specific model (Net) and the single model trained on all noise levels (Net*) are shown. c) The same for camera noise type. Boxes show interquartile range (IQR), line signify median, and whiskers extend to 1.5 IQR.

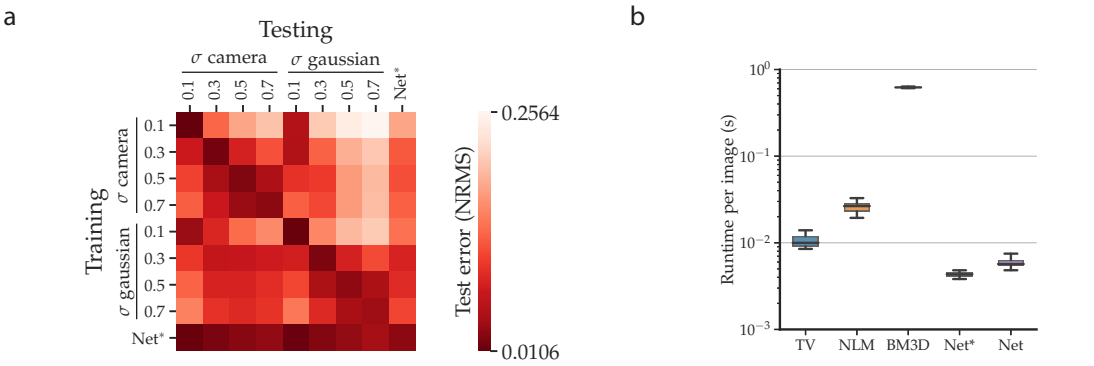


Figure 4.6: Test error for models trained for a specific noise level an types applied to different noise levels and types. The common model (Net*) trained for all noise levels simultaneously performs well across all noise levels and types, whereas the specific models suffer in performance when applied to noise they were no trained for (off diagonal). b) Runtime per image (256×256) for all compared methods. Boxes show interquartile range (IQR), line signify median, and whiskers extend to 1.5 IQR.

Different noise levels and runtime Interestingly, the restoration quality of a single common network trained for all noise levels at once (median SSIM $= 0.82$, Net* in Fig. 4.5b & c) almost equaled that of the noise specific one. Additionally, we show in Fig. 4.6a the test error (NRMSE) of different models on test images with different noise level: Whereas networks trained for a specific noise level perform poorly on images with different noise, the common model is able to attain low reconstruction error for all noise levels, demonstrating the denoising capability of neural networks across different noise scales. Another important advantage of neural network based restoration is that inference is fast, resulting *e.g.* in an order of magnitude smaller runtime per image than BM3D (Fig. 4.6b).

Influence of learning rate To investigate how the choice of the learning rate r influences the restoration result, we trained different models each with a different learning rate r and depict the final test error after 100 epochs in Fig. 4.7a. To counter the effect of random initial conditions, we computed for each r the average and standard deviation of 10 runs with different random seeds. As can be seen in Fig. 4.7a, the final test error is equally small over a wide range of $r < 3 \cdot 10^{-3}$, and quickly grows when choosing a learning rate $r > 3 \cdot 10^{-3}$, saturating at the error of a model predicting empty images. Due to the relative robustness of training convergence we conclude of learning rate range of $r \approx 1 \cdot 10^{-4} - 1 \cdot 10^{-3}$ to be a good default choice.

Effect of training data size Finally, we study the effect of training data size by computing the final test error when training models with only a random subset of n images of the total 2800 original training images. As can be seen in Fig. 4.7b using only a extremely small subset of $n = 10$ images results in an only approximately 21% larger test error than the using the full training data set (NRMSE $= 3.75$vs3.1), and the training error plateaus for $n > 1000$. This indicates that the dependency on training data size is relatively noncritical and that using a relatively small corpus (*e.g.* $n = 1000$) results in already satisfactory restoration results.

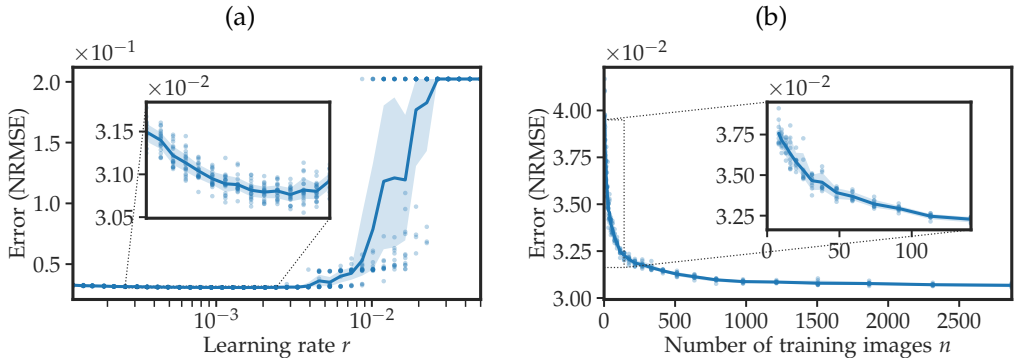


Figure 4.7: Effect of learning rate and number of training images on test error for the denoising experiment on images with Gaussian noise of $\sigma = 0.5$. a) Test error for different learning rates r. The solid blue line is the average of 10 training runs per condition, the shaded area is the respective standard deviation. Each single run is indicated by a blue dot. Note that for high learning rates ($r > 10^{-2}$) the variability of different runs greatly increases (some runs converge to a reasonable solution, some converge to predicting the empty image) and then saturates at the error of the empty image prediction.

4.3.2 Evaluating Network Architectures

To conclude this section, we will study the effect of several hyperparameters relating to the network architecture. This includes the topology of the network (*i.e.* the way in which layers are connected), the size of convolutional filters, the used activation function (*i.e.* non-linearity), and others. To that end, we perform an extensive search through that design space by constructing 680 different model architectures, train them as described before using the same number of epochs and learning rate, and finally report for each hyperparameter combination the average test error of 5 independent runs. To facilitate this task, we used a GPU cluster with 46 usable GPUs (NVIDIA GTX 1080). The results are shown in Fig. 4.8, and we will in the following describe the effect of all different architectural parameters one by one:

Architecture topology (Fig. 4.8b) We compare the following 4 different network topologies:

unet: the U-Net architecture from Ronneberger et al. 2015 (*cf.* Fig. 4.3)

resunet: a *residual* U-Net that predicts the difference between input and output (*cf.* Fig. 4.3)

convnet: a simple convolutional network, *i.e.* a simple stack convolutional layers without pooling operations

stagenet: an architecture consisting of several consecutive convolutional blocks, whose intermediate outputs are summed to yield the final result (inspired by Pelt et al. 2017). Again, no pooling operation is used.

As can be seen from Fig. 4.8b, both the *unet* and *resunet* yield lower test errors in a larger number of cases, than the simple *convnet* and the *stagenet*. This indicates, that pooling operations and subsequent the existence of a bottleneck layer seems to be an important building block for the denoising task at hand. We further conclude the residual U-Net (*resunet*) to be the topology of choice for all further considerations.

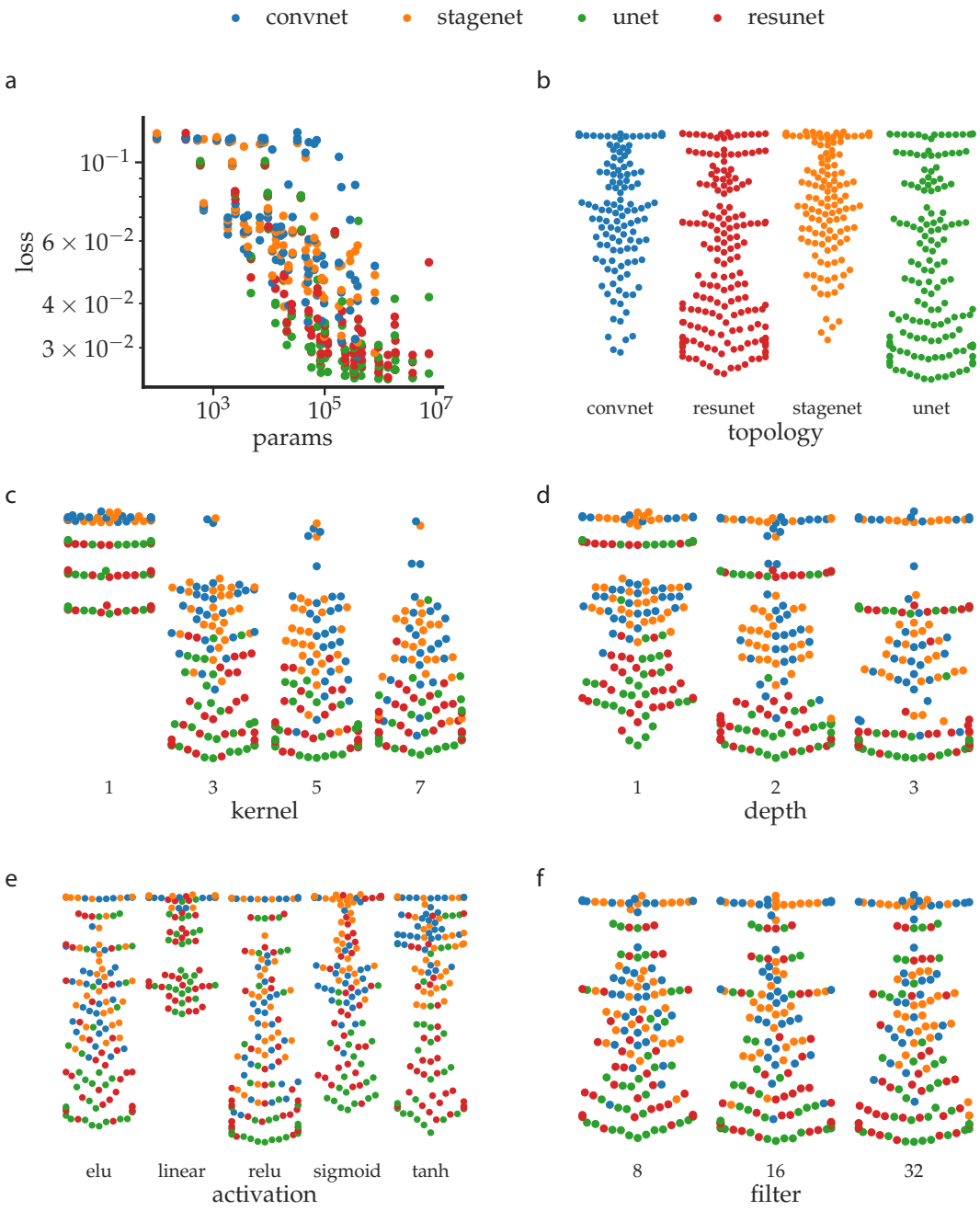


Figure 4.8: Influence of several hyperparameters relating to model architecture on the final test error. Depicted is the test loss (MSE, lower is better) for a total of 680 different model architectures that differ in used kernel size, activation function, number of filters, depth, and topology. For each model architecture, we used the average final test loss of 5 different experimental runs. The y axis always depicts the test loss with the same scaling as in a). Model architecture topology is color coded as indicated at the top. Shown is test error depending on a) number of parameters, b) model topology, c) convolutional kernel size ($1 \times 1, 3 \times 3, \ldots$), c) model depth (number of pooling steps), d) nonlinear activation function, and f) initial number of filters (which get doubled after each pooling step).

Convolutional kernel sizes (Fig. 4.8c) As can be seen, choosing convolutional filters of size 1×1 results in poor performance. This is expected, as in this case every

pixel value is predicted independently of every other input pixel, effectively reducing the model to a per-pixel MLP that ignores spatial correlations. Low test error can be achieved with filters of size 3×3, 5×5, and 7×7, with the latter showing a slight error increase, due to the increased number of parameters and the resulting overfitting. A convolution size of 3×3 can thus be regarded a sensible default choice.

Network depth (Fig. 4.8d) Increasing network depth, *i.e.* number of consecutive convolutional blocks/pooling operations, results generally in improved restoration results. The most pronounced decrease in test error was obtained when increasing the depth from 1 to 2, with only minor improvements for the larger depth 3. As increasing depth comes at the expense of increased parameter number and therefore increased training/inference runtime, we regard a network depth of 2 as a good default setting.

Activation functions (Fig. 4.8e) Several non-linear activation functions have been proposed for classification, starting with the *sigmoid* function (Cybenko 1989), and similar functions like tanh (LeCun et al. 1989)), the *rectified linear unit* (ReLU, Nair et al. 2010), or the *exponential linear units* (ELU, Clevert et al. 2015). As demonstrated in Fig. 4.8 e) the test error is lowest for ReLU, ELU, and tanh (with only small variations between those), while the sigmoid activation results in slightly larger loss. For completeness, we additionally show the results for a *linear* activation function, which are substantially worse. This is expected, as the concatenation of layers with linear activation is equivalent to a single linear transformation that reduces to whole model to a linear transform. Based on our earlier observation, we will regard the ReLU activation as a good default choice.

Number of filter per layer (Fig. 4.8f) Finally, we consider the base (initial) number of features per layer (which gets doubled after each pooling operation). Interestingly, the restoration loss is less sensitive to this design parameter. We conclude that an initial feature number of $16 - 32$ is a sensible choice in most cases.

4.4 Restoration with Physically Acquired Training Data

We now will apply the outlined restoration strategy to several biological imaging examples, where training data can be acquired *physically*, *i.e.* by changing the acquisition parameters of the microscope. In particular, we will consider the important case of restoring (denoising) low signal-to-noise (SNR) volumes by using corresponding images of the same specimen acquired with different camera exposure and laser power. We will demonstrate on the examples of volumetric live-cell imaging of light-sensitive samples like flatworm (Planaria) and flour beetle (Tribolium), that the recovery of extremely low signal-to-noise ratio images with up to 60 fold reduced exposure/laser-power is possible. Following the same strategy, we will show how the common task of surface projection of developing *Drosophila* epithelia tissue can

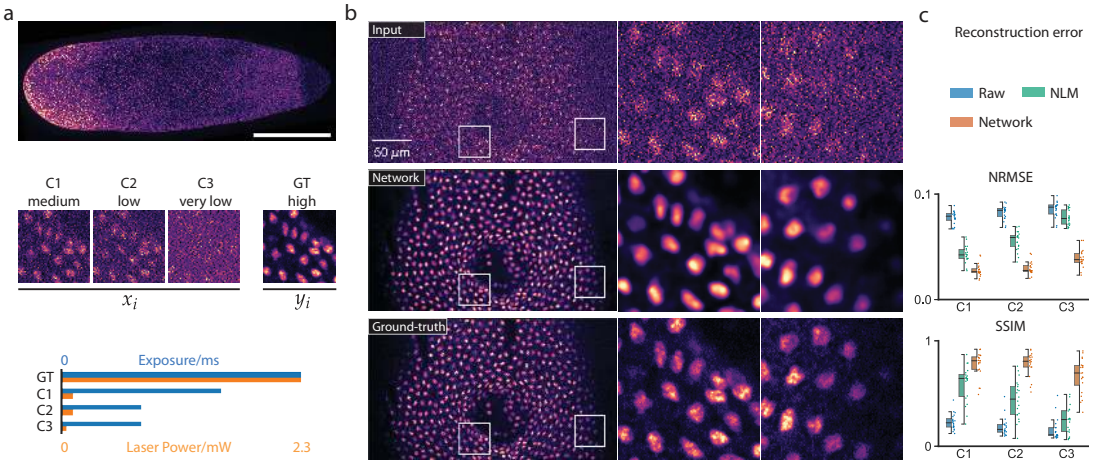


Figure 4.9: Restoration of noisy (low SNR) volumes of the flatworm *Schmidtea mediterranea*. a) Overview of different imaging conditions of *fixed* nucleus-stained (RedDot1) *Schmidtea mediterranea* training data. We use high laser power and camera exposure for ground-truth (GT) and reduced exposure and laser power (C1-C3) compatible with live-cell imaging. We imaged 96 *fixed* specimen at all different imaging conditions, resulting in a corpus of $\approx$ 17000 aligned low (C1-C3) and high (GT) SNR training pairs (x_i, y_i) of size $64 \times 64 \times 16$. Scalebar $500\mu m$. b) The trained network applied to previously unseen images (condition C2). Shown are a single image-plane of a raw input stack (top row), the network prediction (middle row), and the high SNR gold-standard/ground-truth (bottom row). c) Quantification of prediction error on hold-out test volumes for different laser intensities and exposure times (C1 to C3, $n = 20$). Box plots show normalized root-mean-squared error (NRMSE) and structural similarity (SSIM, higher is better) for the input, for a denoising baseline (NLM (Buades et al. 2005)), and for network restorations. Colored boxes indicate interquartile range (IQR), lines therein depict the median, and whiskers extend up to 1.5 IQR. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

be achieved on low SNR volumes using up to 6 times less light dosage. Finally, we will consider the joint task of denoising and axially upsampling of *Drosophila* wing-disc images. In this case, photo-toxicity imposes limitations on the maximally available exposure time and number of focal planes leading to noisy and axially blurred images, that prohibit the proper quantification of tissue folding. Here, we will show that these limitations are amendable, by *physically* acquiring high SNR and axially isotropic volumes on the same microscope as training data.

4.4.1 Denoising of Low SNR Microscopy Images

Planaria Data

We first apply the outlined approach to the imaging of the planaria (flatworm) *Schmidtea mediterranea*, a model organism for studying tissue regeneration. This organism is exceptionally sensitive to even moderate amounts of laser light, suffering muscle flinching at desirable illumination levels even when anesthetized (Shettigar et al. 2017) (Supp. Video 10). Using a laser power that reduces flinching to an acceptable level results in images with such low SNR that they are impossible to interpret directly (Fig. 4.9). Consequently, live imaging of *S. mediterranea* has thus far been intractable.

Video 10

Condition	Exposure	Laser power	NRMSE			SSIM		
	(ms)	(mW)	Input	NLM	Network	Input	NLM	Network
GT	30	2.31	-	-	-	-	-	-
C1	20	0.12	0.0838	0.0513	**0.0376**	0.2117	0.5425	**0.7077**
C2	10	0.12	0.0882	0.0628	**0.0394**	0.1707	0.4022	**0.7080**
C3	10	0.05	0.0910	0.0829	**0.0478**	0.1441	0.2219	**0.6097**

Table 4.2: Experimental conditions and restoration error for denoising of *Schmidtea mediterranea* for the 3 different conditions C1–C3. Shown are NRMSE (lower is better) and SSIM (higher is better) for the input, the network restoration and the Non-local-means (NLM) denoising baseline.

To generate training data, *fixed* samples at different developmental stages of nucleus-labeled (RedDot1) *Schmidtea mediterranea* were imaged at high SNR conditions using high laser power and camera exposure time[9]. To that end, fixed animals were mounted in refractive index matched 1.5 % agarose (50% w/v Iodixanol) and imaged with a spinning disc confocal microscopy setup using a 30x NA 1.05 silicon oil immersion objective as described in Boothe et al. 2017. As it is common nomenclature in many fields, we refer to the high quality, high SNR acquisitions as *ground-truth* (GT) data. This is justified since they do indeed represent the desired quality of our results. Next the same samples were acquired at 3 different levels of reduced laser power and/or reduced exposure times (C1–C3, *cf.* Table 4.2): Crucial to the process is that corresponding image stacks are *well aligned* for all imaging conditions. The different imaging conditions were therefore interleaved during the plane-by-plane acquisition, thus ensuring properly registered input and ground-truth stacks in all cases. In total, 96 stacks of average size $1024 \times 1024 \times 400$ were acquired, resulting in $\sim$ 80 GB of imaged training data. From the so produced well-aligned pairs of high/low SNR volumes, we next extracted on the order of $\sim$ 17000 randomly positioned sub-volume pairs (x_i, y_i) of size $64 \times 64 \times 16$ voxels, which were subsequently used for training and validation of a CARE network. As network architecture a residual version of a 3D U-Net type topology (resunet) was used as described before, with depth 2, convolutional filters of size $3 \times 3 \times 3$, and ReLU activation, resulting in a networks with $\approx 10^6$ parameters (the whole topology is plotted in Fig. A.1). As training loss, the probabilistic loss was used Eq. (4.6)[10]. For training, we used a batch-size of 16, the Adam optimizer with learning rate $4 \cdot 10^{-4}$, and 100 epochs (all training parameters can be found in Table A.2).

After training the network was applied to 20 previously unseen volumes of *Schmidtea mediterranea* images for each of the 3 different conditions. In Fig. 4.9c, Fig. 4.10 we show restoration results for all different conditions. It can be consistently observed that the reconstructed image data are of very high quality, even if the SNR of the input data to be reconstructed was very low, *e.g.* being acquired with a 60-fold reduced light-dosage (Supp. Video 11).

Video 11

[9] Experiments done together with Tobias Boothe (Rink group MPI-CBG).
[10] Using MSE and MAE loss Eq. (4.6) results in similar performance.

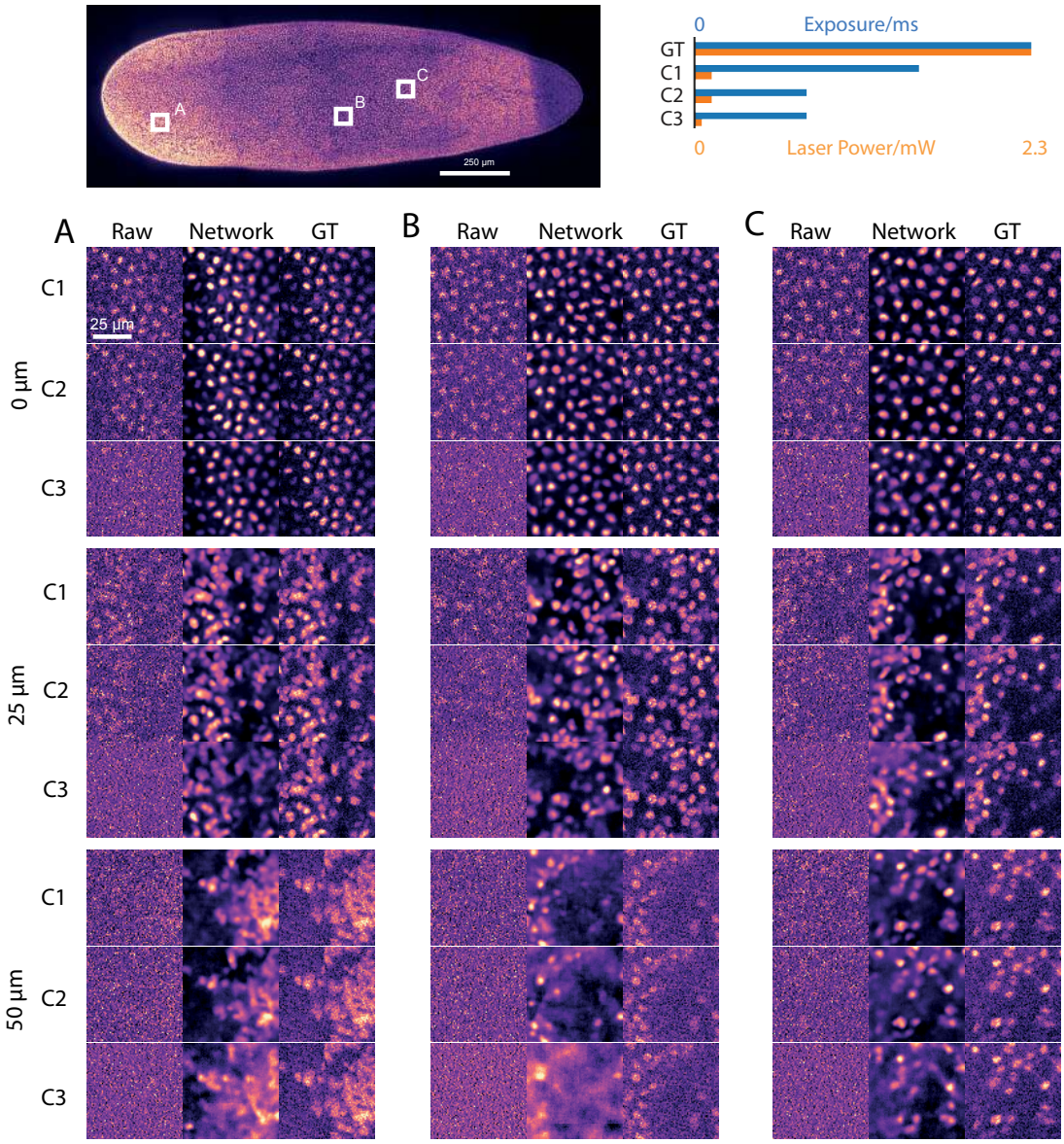


Figure 4.10: Restoration of low-SNR images of (RedDot1 stained) *Schmidtea mediterranea* (Planaria) acquired at 3 different low-SNR conditions (C1-C3) and a high-SNR condition (ground-truth, GT). Depicted are insets of the input, the network reconstruction and the high-SNR ground-truth at different positions and tissue depths ($0 - 50\mu m$). (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

Method	GPU	Typical runtime	Language	Implementation/Source
Lowpass filter	Yes	$\approx 1.5s$	`Python`	https://github.com/maweigert/gputools
Median filter	No	$\approx 4.5min$	`Python`	https://www.scipy.org
Bilateral filter	Yes	$\approx 2.8s$	`Python`	https://github.com/maweigert/gputools
NLM (Buades et al. 2005)	Yes	$\approx 17s$	`Python`	https://github.com/maweigert/gputools
TV (Chambolle 2004)	No	$\approx 1.8min$	`Python`	http://scikit-image.org
BM3D (Dabov et al. 2009)	No	$\approx 47min$	`C++`	https://github.com/gfacciol/bm3d
BM4D (Maggioni et al. 2013)	No	$\approx 2h$	`MATLAB`	https://www.cs.tut.fi/~foi/GCF-BM3D
Network (ours)	Yes	$\approx 40s$	`Python`	

Table 4.3: Overview of compared 3D denoising methods and their implementations. Runtime is given for a typical volume of size $1024 \times 1024 \times 256$.

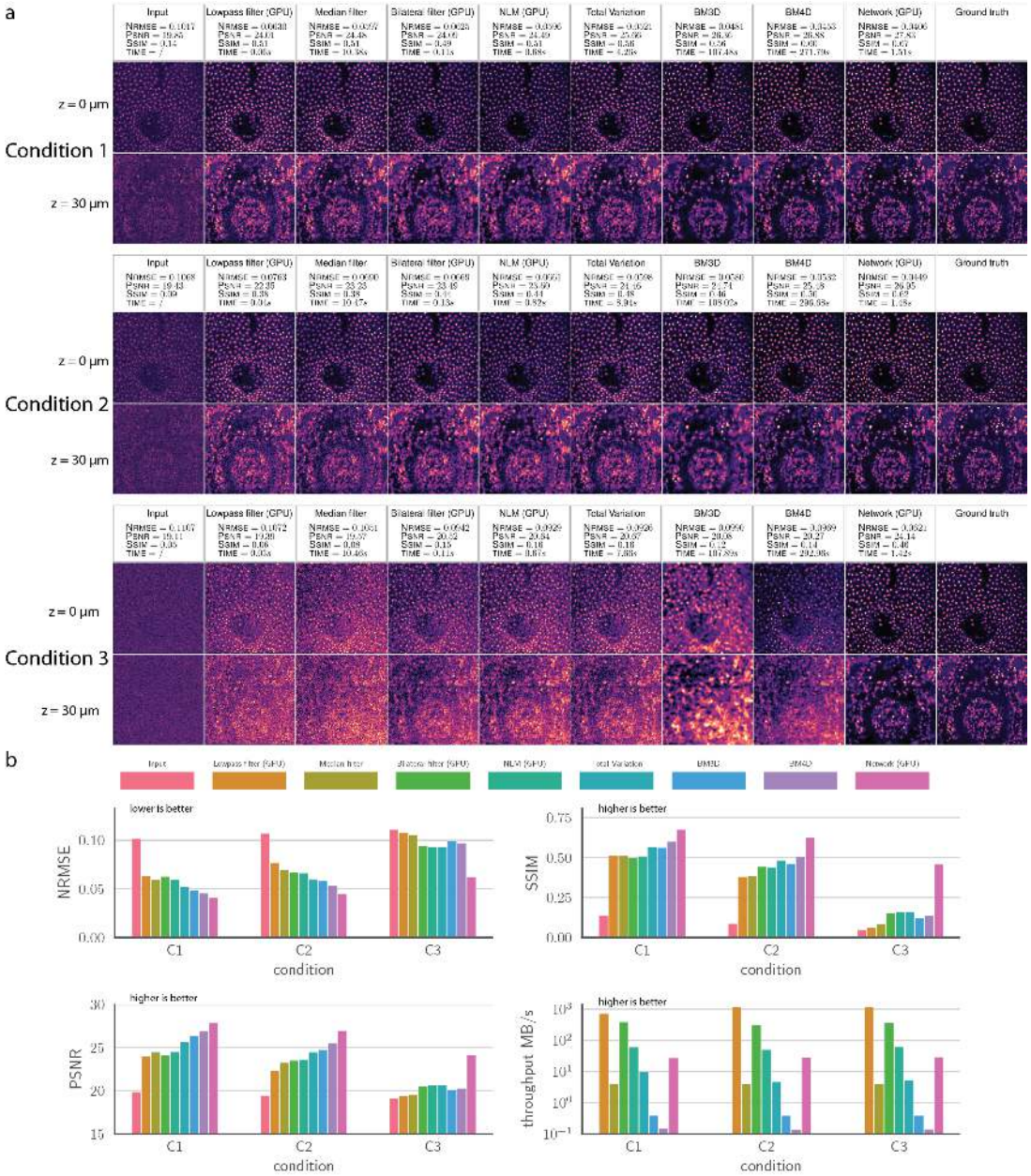


Figure 4.11: Comparison of different denoising methods (Table 4.3 for details). Note that we tuned the hyper-parameters of all methods (*except our Network*) on the test image to yield their best possible results. a) For each of the 3 conditions we show two slices ($z = 0\mu m$ and $z = 30\mu m$) through the same volume of size $(500 \times 500 \times 41)$. All methods are 3D methods, apart from BM3D, which was applied plane-by-plane. b) Error metrics and throughput for each method. We show NRMSE (lower is better), SSIM (higher is better), PSNR (higher is better) and throughput (input size/time, higher is better). Note that the slow runtime of BM3D and BM4D render both methods impractical for typical biological data sets ($\gg$ 1GB). (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

As can be seen in Fig. 4.10, the restored signal is close to the ground-truth even deeper inside the tissue, where the signal contrast additionally suffers due to light scattering and where raw images appear to be impossible to interpret. For additional comparisons and line plots see Fig. A.2. To quantify the reconstruction error between prediction and previously unseen ground-truth we compute the NRMSE

and SSIM defined before and compare it to results obtained by a strong baseline denoising method (Non-local-means, see Fig. 4.9d, Table 4.2). To ensure the strongest baseline results possible, the parameters of Non-local-means (σ, and search-size and -windows) were tuned on the test set, thus providing an upper bound of its performance. In reality, this would of course not be possible (one would have to guess the right parameters), thus the improvement of CARE over that baseline are likely to be higher in practice. As can be seen in Fig. 4.9d and Table 4.2, the results obtained via the CARE network performs substantially better than NLM denoising (*e.g.* SSIM $= 0.71$ *vs*. SSIM $= 0.40$ at condition C2). The improvement in restoration is especially pronounced at the extremely low SNR conditions C3, where the CARE network attains a SSIM $= 0.61$ compared to SSIM $= 0.23$ for the NLM baseline. Moreover, while training CARE networks can take several hours, the restoration time for a volume of size $1024\times1024\times100$ is less than 20 seconds on a single GPU (Nvidia Titan X). In this case, the presented method allows to take low SNR input data unusable for biological investigations and recover high-quality time-lapse data, providing the first practical framework for live-cell imaging of *S. mediterranea*.

To substantiate the superiority of image restoration of low SNR planaria images via CARE, we compare to results obtained by other alternative potent classical denoising methods (*cf.* Table 4.3). These includes low-pass, median and bilateral filter, as well as Total-Variation denoising (Chambolle 2004) and the collaborative filter methods BM3D/BM4D (Dabov et al. 2009; Maggioni et al. 2013). Specifically BM3D and BM4D are considered to be among the strongest non-learned denoising methods and are commonly used as potent comparison methods to state-of the art methods in computer vision (U. Schmidt et al. 2014; K. Zhang et al. 2017a). As before, we tuned the parameters of all competing methods on the test data, thus providing them with an advantage over CARE. For all three conditions, we computed NRMSE, SSIM, and PSNR of the restoration. Due to the slow runtime and large memory requirements of BM4D we used a single test volumes of size $500\times500\times41$. As can be observed from Fig. 4.11 the restoration quality of the network reconstruction is again substantially improved compared to all other methods, with BM4D being the best baseline method. Again this is most pronounced at the lower SNR condition C3, where the BM4D achieves a SSIM of 0.14, compared to 0.46 for the CARE network. Additionally, the large runtime renders BM4D virtually impractical for typical biological data sets(Fig. 4.11b): whereas the network is able to restore a volume of typical size $1024\times1024\times256$ in well under a minute, BM4D would take several hours (Table 4.3). To investigate the dependency of restoration quality on the amount of available training data, we trained several networks with only a subset of the original training stacks. Specifically, we computed for each condition the final test error when using $n \in (17, 66, 266, 1063, 4251, 17005)$ training patches of size $64 \times 64 \times 16$. We observed that already a small number of training images (*e.g.* $n = 100$) leads to an acceptable image restoration quality, specifically for conditions C1 and C2 (Fig. 4.12).

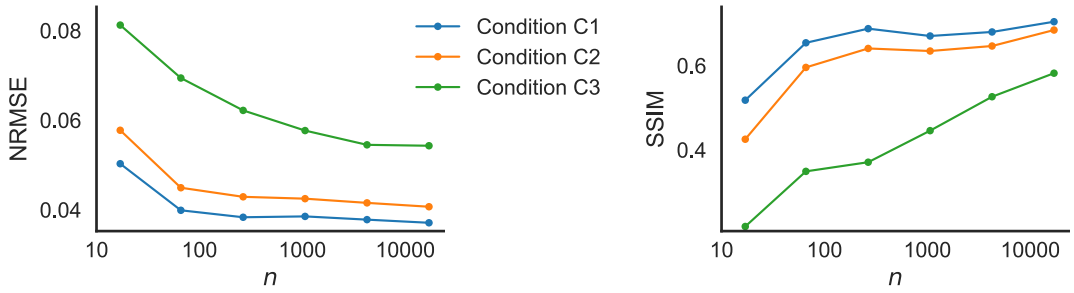


Figure 4.12: Dependency of restoration quality on amount of training data for denoising of Planaria (*Schmidtea mediterranea*) volumes. a) Exposure and laser power used as imaging conditions C1-C3 and a small sample images that was acquired using these conditions. b) Test error (NRSME, SSIM, PSNR) of networks trained with $n \in (17, 66, 266, 1063, 4251, 17005)$ small crops of size $64 \times 64 \times 16$.

Condition	Laserpower	NRMSE			SSIM		
		Input	NLM	Network	Input	NLM	Network
GT	20 %	-	-	-	-	-	-
C1	0.1 %	0.0955	0.0518	**0.0330**	0.1704	0.4635	**0.8827**
C2	0.2 %	0.0891	0.0384	**0.0275**	0.229	0.7157	**0.9066**
C3	0.5 %	0.0762	0.0325	**0.0235**	0.3661	0.8437	**0.9221**

Table 4.4: Experimental conditions and restoration error for denoising of *Tribolium castaneum* for the 3 different conditions C1–C3. Shown are NRMSE (lower is better) and SSIM (higher is better) for the input, the network restoration and the Non-local-means denoising baseline.

Tribolium Data

As a second denoising example, we considered live-cell imaging of developing *Tribolium castaneum* (red flour beetle) embryos. Here, the photo-damage accumulated during long time-lapses limits the maximally attainable SNR for every time point. We used a nuclei-labeled EFA::nGFP transgenic line of *Tribolium castaneum* to image the embryonic development (Sarrazin et al. 2012). Samples were prepared according to standard protocols (Brown et al. 2009), and a Zeiss 710 confocal microscope with a 25× multi-immersion objective was used for imaging[11]. As before, we imaged at 3 different low-SNR conditions and a high-SNR ground-truth condition by varying the laser power (Table 4.4). Again all laser conditions were interleaved during acquisition to ensure registered image pairs. In total, we collected 26 training stacks for each condition, acquired from different samples at different developmental stages. The stack size was $\sim 700 \times 700 \times 50$ for each condition, totaling in 14 GB of training data. As neural network we used the same architecture as for the planaria data, *i.e.* a resunet of depth 2, with convolutional kernel size of $3 \times 3 \times 3$ and ReLU activations. The resulting CARE model was again trained for 100 epochs using probabilistic loss.

[11] Experiments done together with Akanksha Jain (Tomancak group, MPI-CBG).

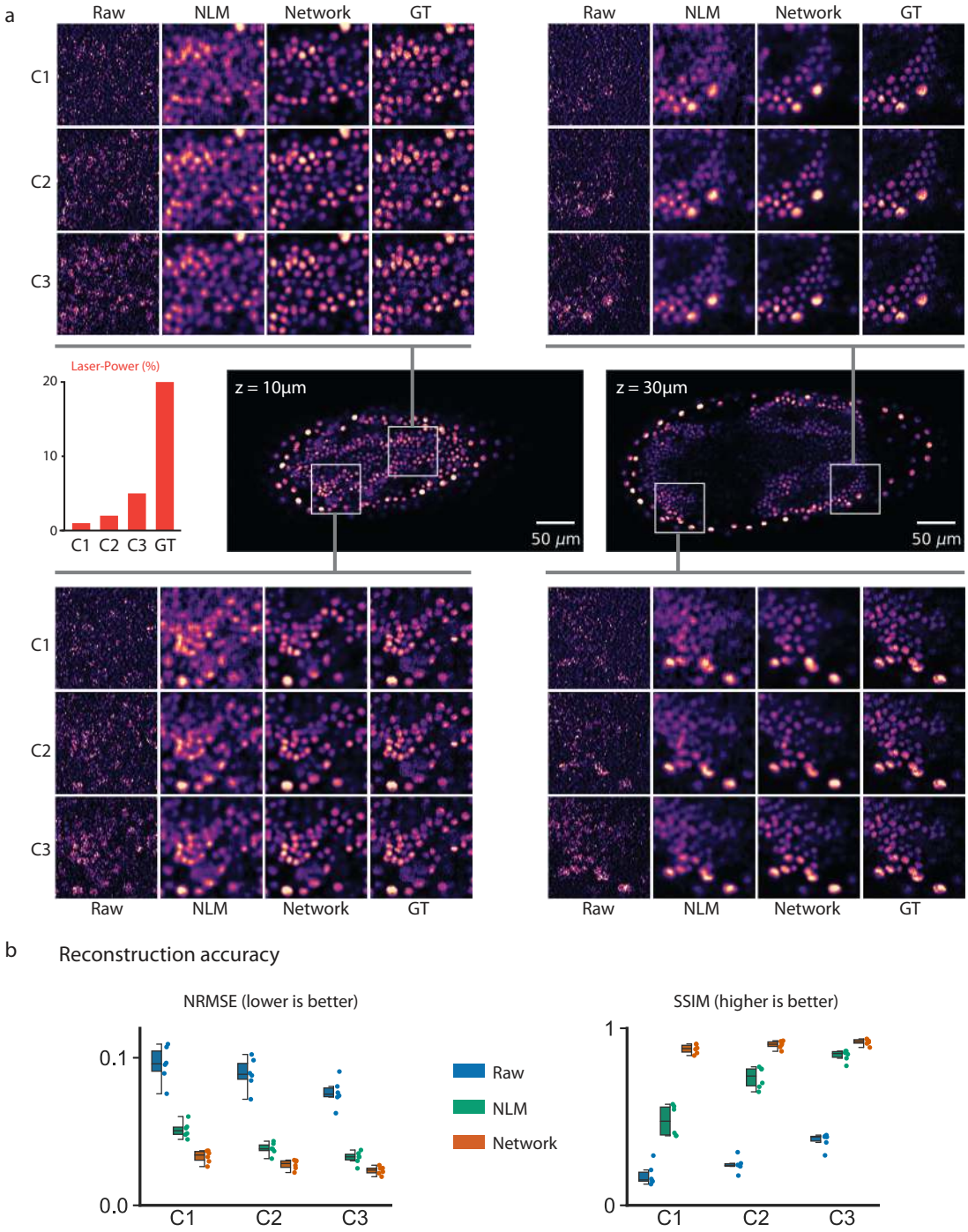


Figure 4.13: Restoration of low-SNR images of *Tribolium castaneum*. Spinning disk confocal acquisitions of a developing embryo (EFA::nGFP labeling) at 3 different low-SNR conditions (C1-C3) and a high-SNR condition (GT). a) Depicted are insets of the input, the best Non-Local-means (NLM) denoising result, the network reconstruction and the high-SNR ground-truth at different positions and tissue depth. b) Reconstruction accuracy measured on held out test set (comprising 6 volumetric stacks for each condition). Shown are normalized root mean squared error (NRMSE, lower is better) and structural similarity (SSIM, higher is better). Colored boxes indicate interquartile range (IQR), lines therein depict the median, and whiskers extend up to 1.5 IQR. Note, that the parameters of the NLM baseline were optimized on the test-set, which would normally not be possible. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

To test the restoration performance, we used a held-out set of 6 different unseen volumes per condition, again acquired at different developmental stages (*cf.* Supp. Video 12). To compare against a baseline denoising method, we again used Non-local-means denoising whose parameters were tuned on the test set. In Fig. 4.13a we depict exemplary slices of a test volume for all conditions, and show the input, the baseline NLM denoising result, the results obtained via the CARE network and the ground-truth condition. Again, the network based image restoration performs well even on extremely noisy input data with up to $50-200$-fold reduced light-dosage when compared to typical imaging protocols (GT). To quantify the reconstruction error, we again computed the NRMSE and SSIM on the test set (*cf.* Fig. 4.13b and Table 4.4). The biggest improvement in restoration quality compared to the baseline was found again for the lowest SNR conditions (C1/C2), where the network results achieved a SSIM of 0.88 (C1) and 0.91 (C2) compared to NLM denoising with 0.46 (C1) and 0.72 (C2).

Video 12

4.4.2 Upsampling of Anisotropic Volumes

A different application that can be addressed via the physical acquisition of training data, is the common case of anisotropic resolved volumes due to axially undersampled acquisitions. Here, the result are noisy and axially blurred volumes, owing to the limited laser power and number of focal planes that are compatible with specimen health or the temporal resolution. This is in contrast to acquisitions of fixed specimen, where such constraints are often not present, and where the signal-noise-ratio (SNR) and axial sampling can be chosen almost freely. A simple strategy to restore these noisy and axially blurred acquisitions is therefore to acquire samples at high SNR condition using a high isotropic focal plane sampling, and then image the same specimen at low SNR condition with less focal planes similar to what is used in a live-cell imaging setting. These corresponding stacks with different noise and axial resolution can then be used as training data for a neural network that is able to learn this joint denoising and upsampling process. Note, that similar approaches for the case of upsampling have been applied for 2D natural images by (Dong et al. 2016; J. Kim et al. 2016; Mao et al. 2016), and for electron microscopy volumes by (Heinrich et al. 2017).

We will demonstrate this approach on the example of live-cell imaging of epithelia tissue, specifically of the imaginal wing disc of *Drosophila*[12]. Here the folding of epithelia layers during development is of considerable biological interest to understand the formation of complex tissue such as the developing fly wing. Due to the light sensitivity of the tissue any volumetric time-lapse, however, needed to be acquired with sufficiently reduced laser intensity and limited number of focal planes, resulting in raw images with considerable amount of noise and deteriorated axial resolution (Fig. 4.14a). To improve signal contrast and axial quality, we acquired

[12] Data acquired by LiYuan Sui (Christian Dahmann group,TU-Dresden).

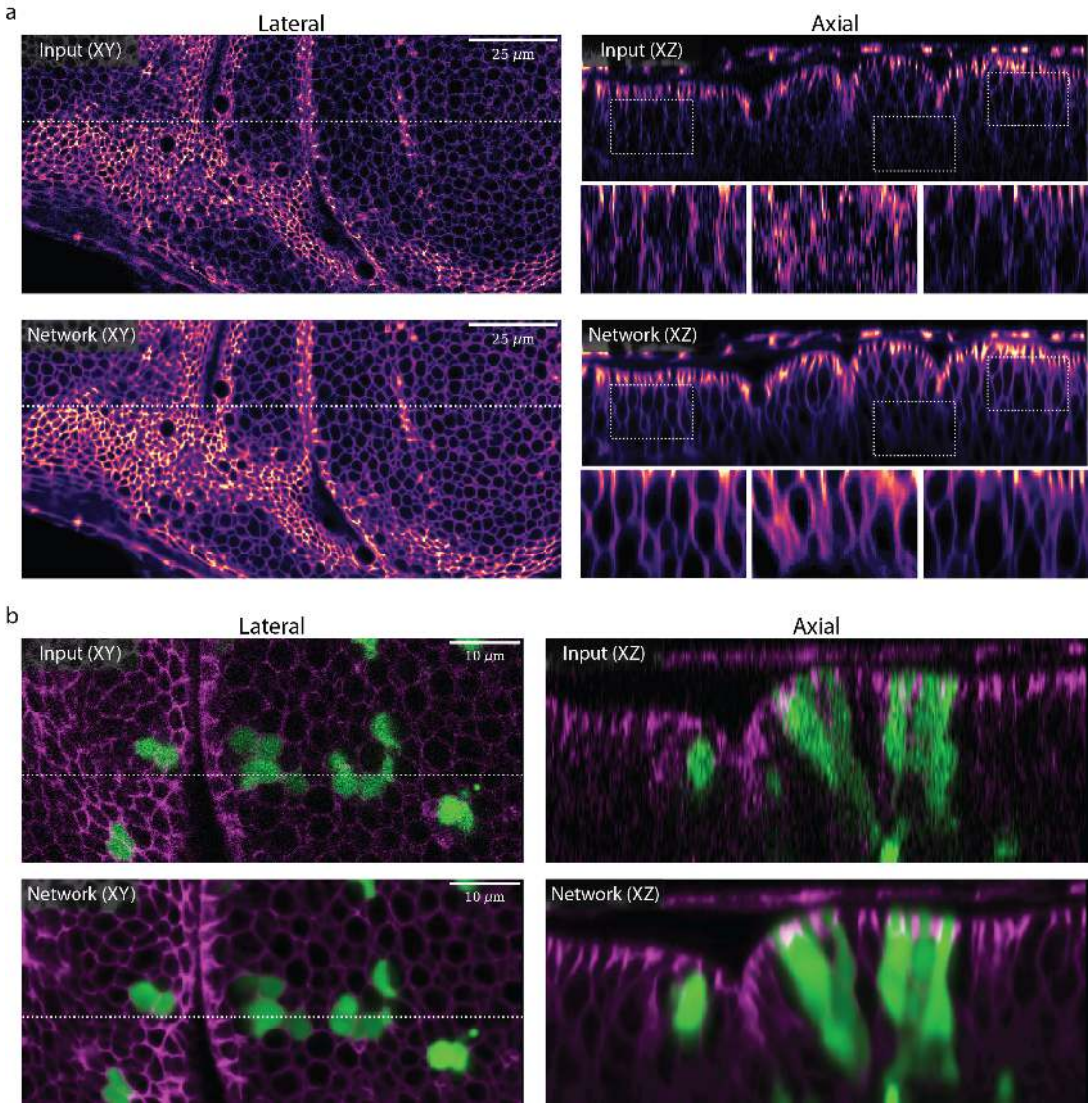


Figure 4.14: Isotropic upsampling with physically acquired training data. a) Anisotropically acquired volumetric stacks of *Drosophila* wingdisc (membrane label Indy-GFP) with pixel-size $\Delta x = 0.24\mu m, \Delta z = 1.30\mu m$. The raw stacks (input, top row) show a low axial (xz) resolution due to the lower axial sampling density. We used stacks acquired with 4-fold higher axial sampling $\Delta z = 0.325\mu m$ and high camera exposure as ground-truth to train a restoration network that jointly denoises and upsamples such volumes. Shown are the results of applying the trained network on an unseen volume (Network, bottom row). Images shown with $\gamma = 0.9$ to enhance visibility of small intensity features. b) The same with a multi-color volume (Indy-GFP and RFP-marked clones).

training volumes of several imaginal discs at different developmental stages with two different microscope settings: the first corresponding to the original imaging conditions, i.e. with reduced laser and large z stepping (low-quality), the second with increased laser power and a 4-fold increased number of focal planes (high-quality, almost isotropic). A Leica SP5 MP confocal microscope with a 40x/1.25 oil immersion objective was used for imaging, and different conditions again interleaved to ensure well-registered volumes. The low quality stacks were acquired with a z spacing of $1.30\mu m$ and low camera exposure, compared to a z spacing of $0.325\mu m$ and high exposure time for the high quality stacks. As neural network model we used again a residual U-Net with depth 2, convolutional kernel size of

3×3×3 and ReLU activations, which was trained for 100 epochs using mean absolute error (MAE) loss. As the specimen was labeled with two different tags (Indy-GFP as membrane marker, and RFP marked clones) we trained a different network per marker channel.

Finally, we applied the resulting network to the raw time-lapse data, producing images with considerable less noise and improved axial resolution (Fig. 4.14a). Notably, the cell membranes of neighboring cells within the pseudo-stratified epithelia are almost invisible in the raw axial slices, whereas they are distinctively present in the network reconstruction (Fig. 4.14a). The same can be observed for multi-color stacks, where additionally RFP-marked clones are present. Here, we applied a differently trained network to each color channel independently. We observe that not only the GFP signal of the cell membranes is restored as before, but the recovered RFP signal of clones integrates between cell membranes as expected (Fig. 4.14b). In both cases, the imagery recovered via CARE provided a substantially improved outline of the folded wing disc geometry and thus directly contributed to the important biological investigation of tissue folding in *Drosophila*.

4.4.3 Surface Projection of Epithelia Tissue

In biological experiments involving the study of sheet-like tissue development, it is often useful to image 3D volumes and *project* them to 2D surfaces for analysis, for example when studying cell behavior in developing epithelia of the fruit fly *Drosophila melanogaster* (Aigouy et al. 2010; Etournay et al. 2015; Guirao et al. 2015). An ideal projection would only collect voxel intensities that correspond to the smooth manifold containing the structure of interest and ignore spurious image intensities that are otherwise contained in a given raw image volume. A commonly used maximum projection, however, would be problematic in this case, as the relevant surface-like signal is buried by the non-informative noise outside of that surface, and more sophisticated methods are usually applied (Blasse et al. 2017; Shihavuddin et al. 2017). In a typical imaging pipeline to study the development of the wing of *Drosophila* melanogaster pupae, a mosaic of overlapping volumes that cover the full wing is recorded at many successive time-points. Importantly, the temporal resolution of such time-lapses is, again, limited by the photon-budget. While one would prefer to use low laser-power to avoid phototoxicity, and low exposure times to achieve higher frame rates, the desired SNR for downstream analysis can so far only be achieved with at least moderate laser intensities. Increasing laser-power for a given exposure time results in abnormal development and eventually nonviable wings; increasing the exposure leads to insufficient temporal resolution to capture rapid tissue reorganization. Simultaneously denoising and projecting volumetric time-lapse data would directly benefit biological and biophysical studies of developing fly wings and other projects with similar imaging needs, and presents the opportunity to test if CARE networks can deal with such composite tasks.

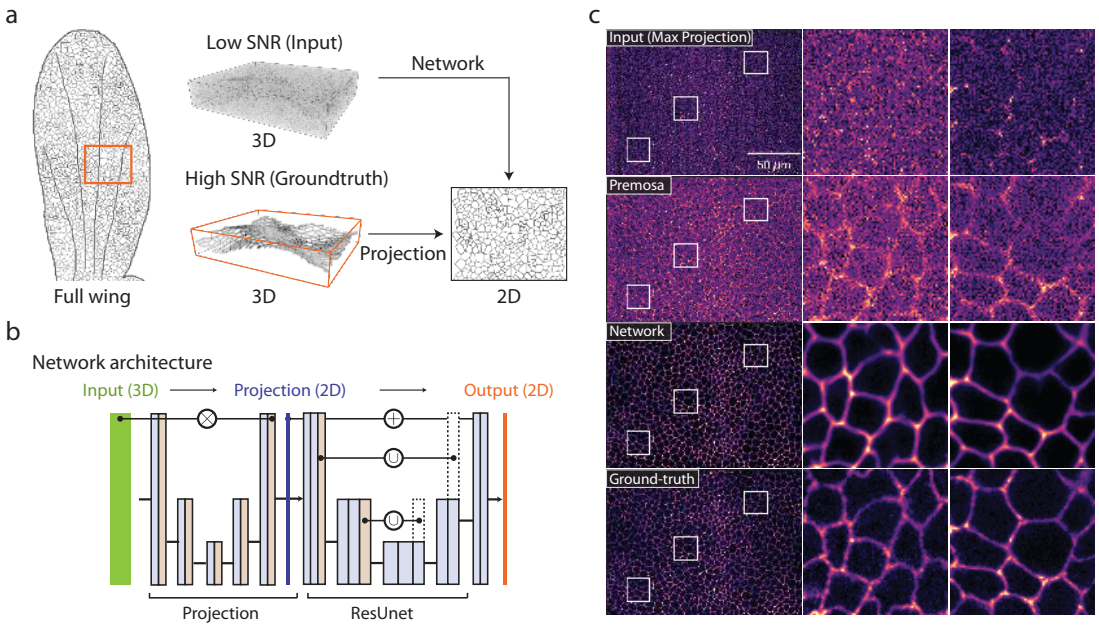


Figure 4.15: Joint surface projection and denoising of epithelia tissue. a) A developing *Drosophila* wing is a single layer of cells, embedded in a 3D volume. Imaging of long time-lapses requires to image at low SNR to avoid photo-toxicity and bleaching. With CARE, the cell layer of interest can be projected onto a 2D image, while also being denoised by the same, composite network. b) The architecture of the proposed CARE network consists of two consecutive sub-networks: a first subnetwork performs the projection of voxel intensities (top half), and a second subnetwork denoises this projection. c) Results on E-cadherin labeled fly wing data. Shown is a max-projection of the raw input data (top row), result obtained by applying the state-of-the-art projection method PreMosa (Blasse et al. 2017) (second row), the solution computed by our trained network (third row), and the desired (ground-truth) projection obtained by applying PreMosa on a very high laser-power (high SNR) acquisition of the same sample. The ground-truth data was obtained at a laser intensity that cannot be used for live-cell imaging without causing damage to the sample. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

Condition	Exposure	Laser	NRMSE			SSIM		
			Input	PreMosa+	Network	Input	PreMosa+	Network
GT	240*ms*	20%	-	-	-	-	-	-
C0	120*ms*	2%	0.1416	0.1168	**0.0798**	0.0925	0.2561	**0.5695**
C1	120*ms*	3%	0.1079	0.0744	**0.0690**	0.3562	0.5668	**0.6232**
C2	120*ms*	5%	0.1235	0.0882	**0.0732**	0.2358	0.4620	**0.6130**

Table 4.5: Acquisition conditions and restoration accuracy for surface projection of developing fly wing of *Drosophila melanogaster*. A spinning disk confocal microscope was used to acquire volumes at 3 different low-SNR conditions (C1-C3) and a high-SNR condition (GT). Shown are NRMSE (lower is better) and SSIM (higher is better) for the input, the network restoration and the best competing baseline (PreMosa+ = PreMosa from Blasse et al. 2017 with subsequent Non-Local-Means denoising).

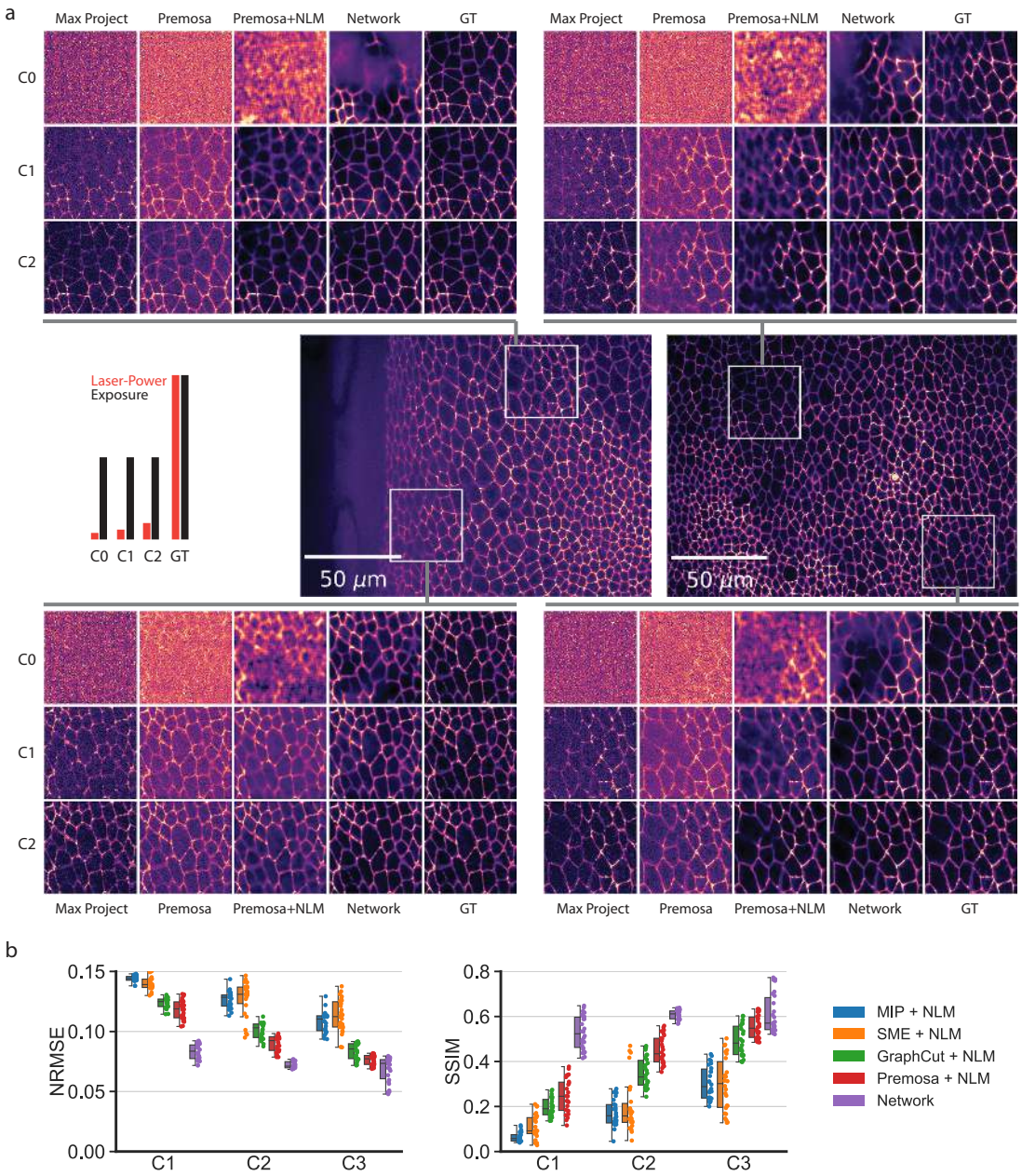


Figure 4.16: Joint surface projection and denoising of developing *Drosophila melanogaster* wing epithelia and comparison with baseline methods. a) Depicted are insets of the input, the network reconstruction and the high-SNR ground-truth at different locations. b) Reconstruction accuracy measured on held out test set (comprising 26 volumetric stacks for each condition) for data acquired at different laser intensities and exposure times (conditions C1–C3). Shown are normalized root mean squared error (NRMSE, lower is better) and structural similarity (SSIM, higher is better). We compare the result of our Network with MIP (Maxmimum projection), SME (smooth 2D manifold extraction Shihavuddin et al. 2017), GraphCut (Minimum cost surface projection Wu et al. 2002; K. Li et al. 2006) and PreMosa (2D surface projection Blasse et al. 2017), where we additionally applied NLM (Non-Local-Means Buades et al. 2005) on the output of all methods (*except ours*). Colored boxes indicate interquartile range (IQR), lines therein depict the median, and whiskers extend up to 1.5 IQR. Note that we tuned the NLM hyper-parameter for best results on the validation data, giving all baseline results an additional advantage. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

To train a CARE model in this case, we acquired pairs of high (GT) and low SNR image stacks for several conditions (C0-C2) (see Table 4.5). Again, all conditions were interleaved as channels to ensure well aligned images. For each condition, we acquired 180 different 3D stacks (of size $\sim 700 \times 700 \times 50$), totaling 35 GB of training data. As prediction target we used the surface projected 2D ground-truth signal obtained via *PreMosa* (Blasse et al. 2017), thus pairing noisy 3D volumes with the noise free 2D projection as training pairs (Fig. 4.15a). We extracted $\sim$ 17000 random 3D patches of size $64 \times 64 \times 50$ from input stacks and the corresponding patches of size 64×64 from the ground-truth 2D projection. The composite task of joint projection and denoising cannot be modeled by a simple residual network as before, due to the projection step involved. We therefore conceived a dedicated network architecture that consists of two parts: *(i)* an image restoration part, similar to the architecture used before, and *(ii)* an additional projection subnetwork that specializes on content-aware data projection (Fig. 4.15b, the full topology is depicted in Fig. A.3). In total, the proposed architecture featured $\approx$ 1.3 Mio trainable parameters. Again, we trained a model for each condition using a batch-size of 16, the Adam optimizer with learning rate $4 \cdot 10^{-4}$, and 100 epochs (*cf.* Table A.2 for more details).

After training, we applied the resulting CARE networks on a held out test set comprising 26 volumes for each condition (C0-C2). As can be seen in Fig. 4.15c and Fig. 4.16a, the projection results obtained via CARE are visually comparable to ground-truth for up to a 8-fold (C2) or 13-fold (C3) reduced light dosage, whereas the baseline projection methods (Max projection, and PreMosa with additional Non-local-means denoising tuned on the test set) often struggles to discern fine membrane structures from the embedding noise. Interestingly, for the extremely noisy condition C0 we see the network to sometimes miss structures and instead predicting blurred signal, and sometimes to predict the correct structures, whereas the baselines completely fail to restore any meaningful signal (Fig. 4.16a). To quantify the error we used the same error metrics as before (NRMSE and SSIM) and computed them on the test set (Fig. 4.16b & Supp. Video 13). We compare with several baseline methods: maximum projection (MIP), smooth 2D manifold extraction (SME, Shihavuddin et al. 2017), GraphCut (Minimum cost surface projection Wu et al. 2002; K. Li et al. 2006) and PreMosa (2D surface projection Blasse et al. 2017). Additionally, we applied NLM (Non-Local-Means Buades et al. 2005) to the results of all competing method (except CARE), again tuning the parameter to the test set, given them an advantage over CARE. As evident from Fig. 4.16b, the network again achieves a substantially higher reconstruction quality especially for the lowest SNR condition (C0), where the CARE attained a SIM of 0.56 compared to the next best method (PreMosa + NLM) with an SSIM of 0.25. Still, PreMosa+NLM turns out to be the best competing method, performing considerable better than MIP, SME, and GraphCut.

Video 13

We conclude that a CARE network dedicated to the task is able to to jointly

project and denoise epithelia volumes, enabling to restore projections comparable to ground-truth quality from conditions that use $8-13$ less light dosage (C1,C2). Thus, the gained photon-budget can be used to move beyond the design-space tetrahedron, for example by increasing temporal resolution, and consequently improving the precision of tracking of cell behaviors during wing morphogenesis (Etournay et al. 2016).

4.5 Restoration of Axial Resolution with Semi-synthetic Training Data

So far, the application of CARE has relied on the availability of matching pairs of high and low quality images, both *physically* acquired at a microscope. We now investigate the case where this kind of data is not available, and whether image pairs useful for training can also be obtained by computationally modifying existing microscopy images. Specifically, we will study the fundamental problem in fluorescence microscopy, that the axial resolution of volumetric acquisitions is significantly lower than the lateral resolution. The main reasons for this *anisotropy* are *i)* the inherent axial elongation of the blur-inducing optical point spread function (PSF) of most microscopes, and *ii)* the often reduced axial sampling (by an additional factor $\sigma > 1$ compared to lateral sampling) due to considerations regarding image acquisition time and photo-toxicity. In combination, these two effects result in volumes with a lower axial resolution typically by a factor of 2 to 8 compared to the lateral resolution, compromising the ability to accurately measure properties such as the shapes or volumes of cells. As this anisotropy is intrinsic to acquisitions on most microscopy setups[13], the strategy used before to acquire training data physically at the microscope cannot be readily applied as isotropic ground-truth is not available.

To circumvent this problem, multiple techniques are known and used: Classical deconvolution methods (Richardson 1972; Lucy 1974) are arguably the most common of these. They can be applied on already acquired data, however, their performance is typically inferior to other more complex techniques. Some confocal systems, e.g. when using two-photon excitation with high numerical aperture objectives and an isotropic axial sampling, can acquire *almost* isotropic volumes (Morales-Navarrete et al. 2015; Economo et al. 2016). Downsides are low acquisition speed, high photo toxicity/bleaching, and large file sizes. Light-sheet microscopes, instead, can improve axial resolution by imaging the sample from multiple sides (views). These views can then be registered and jointly deconvolved (Preibisch et al. 2014). The disadvantage is the reduced effective acquisition speed and the need for a complex optical setup. In the following we show how to recover isotropic resolution from a single, anisotropic acquired microscopic 3D volume via a suitable training data strategy.

The overall mathematical problem can be understood as a combination of a

[13] Modalities that would allow (close to) isotropic acquisitions are rare, *e.g.* multi-view light-sheet microscopy (Preibisch et al. 2014; Chhetri et al. 2015), or 4π-confocal microscopy (S. W. Hell et al. 1994)

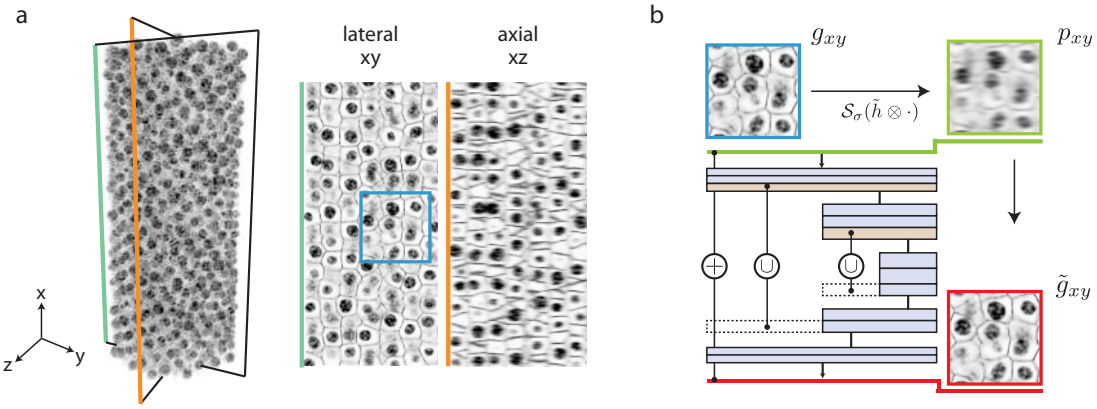


Figure 4.17: a) 3D images acquired on light microscopes are notoriously anisotropic due to axial undersampling and optical point spread function (PSF) anisotropy. b) The IsoNet-2 architecture has a U-net (Ronneberger et al. 2015) like topology and is trained to restore anisotropically blurred/downsampled lateral patches. After training it is applied to the axial views. (Reprinted by permission from Springer Nature: Weigert et al. 2017, Copyright 2017)

super-resolution problem on subsampled data, and a deconvolution problem to counteract the microscope induced optical PSF, and can be formalized as follows: Given the true fluorophore distribution $f(x, y, z)$ the acquired volumetric image g of a microscope can be approximated by the following process

$$g = \mathcal{P}\left[\mathcal{S}_\sigma^z(h \otimes f)\right] + \eta \tag{4.12}$$

where $h = h(x, y, z)$ is the point spread function (PSF) of the microscope, $\otimes$ is the 3D convolution operation, S_σ^z is the downsampling/slicing operator along z by a factor σ, $\mathcal{P}$ is the signal dependent noise operator (e.g. Poisson noise) and η is the detector noise. As the PSF is typically elongated along z and $\sigma > 1$, the lateral slices g_{xy} of the resulting volumetric images show a significant higher resolution and structural contrast compared to the axial slices g_{xz} and g_{yz} (*cf.* Fig. 4.17a). As in all examples before, we want to directly learn the mapping between blurred and downsampled images and its true underlying signal. As no ground-truth for the true signal is available, we make use of the resolution anisotropy between lateral and axial slices and aim to restore lateral resolution along the axial direction by generating *semi-synthetic* training data. Specifically, we will make the assumption, that cellular and sub-cellular structures are present in diverse orientations within the volumes of interest, and that the mentioned physical process leading to image anisotropy is well understood and can be modeled *synthetically*. This assumption is applicable for most biological imaging situations, *e.g.* for the imaging of whole embryos, where the specimen is typically randomly positioned, or for acquisitions of large tissues, where the biological substructures are not all aligned equally. It might be violated, however, for highly stratified or polarized tissues, such as cortical columns in the brain. Positing this assumption, we then proceeded as follows: For a given anisotropic volume g, the PSF of the microscope h and the axial up-sampling factor σ, we first compute for each volume g a corresponding volume p that is *i)* blurred *laterally* (along x) by an

appropriately rotated version of the point spread function h, and *ii)* down-sampled *laterally* (along x) by σ. In other words, we apply a rotated version of the image formation model in Eq. (4.12)

$$p = \mathcal{S}_\sigma^x(\tilde{h} \otimes g) \tag{4.13}$$

with a suitable chosen *rotated PSF* $\tilde{h}$. By assembling 2D lateral patches (p_{xy}, g_{xy}) from p and g, we thus obtain corresponding high resolution (g_{xy}) and computationally distorted (p_{xy}) patches, suitable for training a restoration model. This strategy to create training data from existing images is similar to the concept of *self super-resolution* used in MRI reconstructions methods (Jog et al. 2016). For $\tilde{h}$ we will consider two choices *i) full*: $\tilde{h} = h_{rot}$ where h_{rot} is a rotated version of the original PSF that is aligned with the lateral planes, and *i) split*: $\tilde{h} = h_{split}$ which is the solution to the deconvolution problem $h_{rot} = h_{iso} \otimes h_{split}$ and h_{iso} is the isotropic average of h. The later choice is motivated by the observation that convolving lateral slices with h_{split} leads to images with a resolution comparable to the axially ones. We proceed to train a network restoration model on sufficiently many pairs $(g_{xy}^n, p_{xy}^n)_{n \in \mathbb{N}}$, after which we apply the network to all unseen, anisotropically blurred, linearly upsampled *axial* slices (xz/yz) of the original volume to yield the final estimation output (*cf.* Fig. 4.17b).

4.5.1 Validation with Synthetic Data

We first validate our approach on 3 different types of synthetic data sets that resemble typical biological structures: *i)* volumes containing about 1500 simulated nuclei, *ii)* membrane structures as they are frequently seen in tightly packed cell epithelia, and *iii)* simulated cell nuclei with surrounding labeled membranes (*cf.* Fig. 4.18a). All volumes were created in-silico by combining plausible structure distributions, Perlin noise based textures and realistic camera noise. We then next simulated anisotropic acquisitions by first convolving these volumes with different PSFs, and then subsampling the volume by a certain downsampling factor σ. To test a variety of different conditions we used *i)* an anisotropic Gaussian PSF (FWHM 4px and 16px along x/y and z respectively) and subsampling factor $\sigma = 8$, *ii)* a confocal PSF ($NA = 1.1$, assuming pixel size $0.1\mu m$) and subsampling factor $\sigma = 4$, and *iii)* a light-sheet PSF ($NA_{ill} = 0.1$, $NA_{det} = 0.8$, assuming pixel size $0.16\mu m$) with subsampling factor $\sigma = 6$. For restoration we will compare two network architectures, which we will call *IsoNet-1* and *IsoNet-2*. The first, *IsoNet-1*, is a simple convolutional network identical to the SRCNN architecture, first proposed by Dong et al. 2014, where it was used for the task of upsampling natural images. It consists of 4 successive convolutional layers of decreasing kernel size, and without pooling layers, totaling $57,281$ parameters. The *IsoNet-2* is a residual U-Net as used before, with depth 2, and convolutional kernel size of 7×7, with a total of $453,503$ parameters.

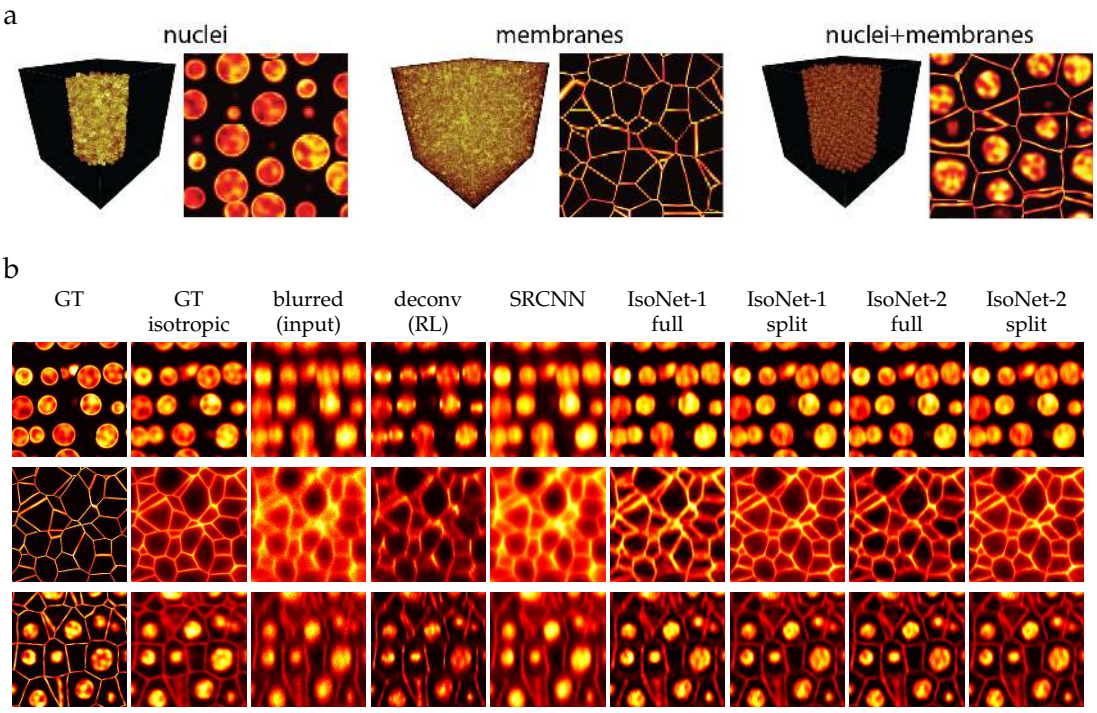


Figure 4.18: Validation on synthetic data and comparison of different methods. a) Synthetic ground-truth data sets resembling typical biological structures. Every volume is of size $512 \times 512 \times 512$. b) Results of restoration. Rows show crops of axial slices of 3D nuclei data, membrane data, and a combined data set, respectively. The columns are: *i*) ground-truth axial slice, *ii*) the same ground-truth slice convolved with an isotropic PSF, *iii*) anisotropically blurred isotropic GT image (the input images to all remaining columns, *iv*) deconvolved images using Richardson-Lucy (Richardson 1972; Lucy 1974), *v*) SRCNN (Dong et al. 2014), *vi*) IsoNet-1 with one (full) PSF, *vii*) IsoNet-1 making use of the split PSFs, *viii*/*ix*) IsoNet-2 with full PSF and split PSFs, respectively. (Panel b reprinted by permission from Springer Nature: Weigert et al. 2017, Copyright 2017)

volume (PSF/scale)	blurred (input)	deconv (RL)	SRCNN	IsoNet-1 full	IsoNet-1 split	IsoNet-2 full	IsoNet-2 split
nuclei (Gaussian/8)	25.84(0.17)	27.48(0.16)	25.89(0.19)	32.18(0.18)	32.47(0.18)	35.11(0.18)	**35.61**(**0.18**)
	24.09(0.10)	25.88(0.11)	24.15(0.10)	27.74(0.10)	27.53(0.10)	**29.51**(**0.10**)	28.84(0.10)
membranes (confocal/4)	21.83(0.13)	18.52(0.14)	21.69(0.13)	19.55(0.12)	26.19(0.13)	19.14(0.12)	**27.33**(**0.14**)
	15.95(0.01)	16.48(0.01)	15.94(0.01)	16.84(0.01)	16.49(0.01)	**17.09**(**0.01**)	16.62(0.01)
nuclei+memb. (light-sheet/6)	28.13(0.37)	25.00(0.37)	28.69(0.39)	25.59(0.38)	30.23(0.37)	25.40(0.39)	**30.95**(**0.36**)
	24.61(0.52)	26.57(0.51)	24.59(0.52)	26.86(0.51)	26.07(0.51)	**27.85**(**0.51**)	26.66(0.51)

Table 4.6: Computed PSNR values against isotropic GT (upper rows), and against GT (lower rows). PSF types are: Gaussian ($\sigma_{xy}/\sigma_z = 2/8$); confocal with numerical aperture $NA = 1.1$; light-sheet with $NA_{detect} = 0.8$ and $NA_{illum} = 0.1$. Bold values indicate best. The standard deviation (for $n = 10$ random volumes) is shown in parentheses.

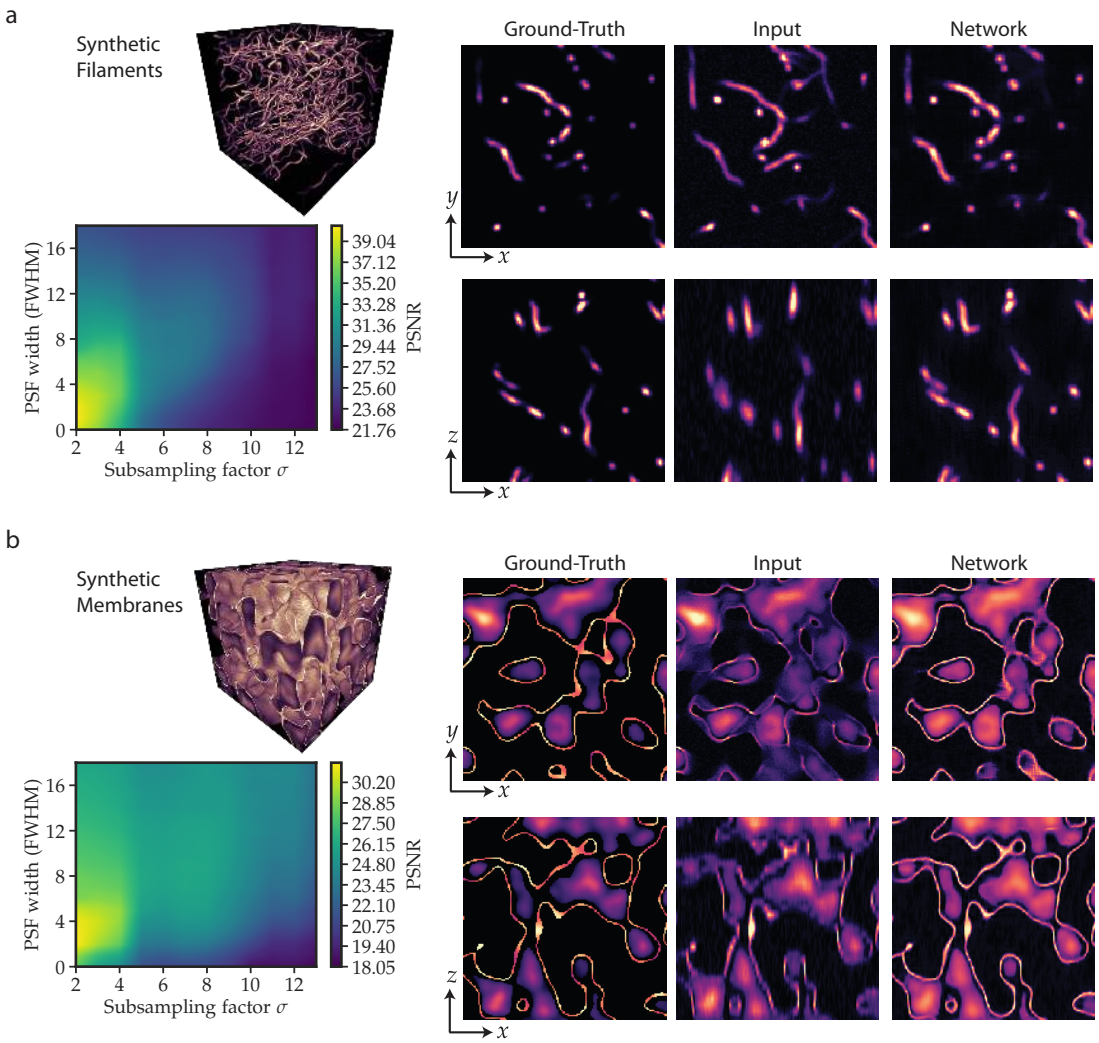


Figure 4.19: Joint effects of subsampling factor σ and PSF width (FWHM) on restoration quality. We used two data sets of a) synthetic filaments , and b) synthetic membranes. Each data set comprises 5 volumes from which we synthetically generated blurred and down-sampled input volumes for values of $2 \leq \sigma \leq 14$ and $0px \leq \text{FWHM} \leq 18px$, which were then restored as described. Shown is the average $\text{PSNR}(\sigma, FWHM)$ of the restorations. On the right, we show representative lateral (xy) and axial (xz) slices through the ground-truth, input and network result for a) $\sigma = 4, \text{FWHM} = 8px$, and b) $\sigma = 8, \text{FWHM} = 8px$.

As the problem involves a combination of a up-sampling and deconvolution, we furthermore compare against a purely up-sampling method (SRCNN, Dong et al. 2014), and performing Richardson-Lucy deconvolution only (Richardson 1972; Lucy 1974). The restoration results are presented Fig. 4.18b, where the first column shows the ground-truth images that were used to generate the isotropic ground-truth (second column), by convolving with the isotropic PSF, and the blurred images that were subsampled and convolved with realistic PSFs in order to resemble microscopic data. This third column (blurred) is then used as the input to all our and other tested methods. The subsequent columns show the results of

- Richardson-Lucy deconvolution (Lucy 1974), *i.e.* disregarding upsampling ,
- pure SRCNN (Dong et al. 2014), *i.e.* disregarding the PSF,
- the IsoNet-1 using the full PSF,

- the IsoNet-1 using the anisotropic component of the PSF h_{split},
- the IsoNet-2 using the full PSF, and
- the IsoNet-2 using the anisotropic component of the PSF h_{split}.

In Table 4.6 we compare the PSNR of the restored volumes with the two ground-truth versions, averaged over 10 different randomly created stacks per data set type. As can be seen, our method (IsoNet-2) performs significantly ($p < 0.01$) best in all cases. Note that failing to incorporate the PSF (as with pure SRCNN) or the upsampling process σ (as with pure deconvolution) results in a vastly inferior reconstruction.

Evidently, the larger the subsampling factor σ and the axial width of the PSF (FWHM), the more information is lost in the given input compared to the ground-truth, and the greater the reconstruction error will be. To systematically investigate this, we created synthetic volumes of filaments and membranes, from which we synthetically generated blurred and downsampled input volumes for a range of subsampling factors and PSF widths (Fig. 4.19). For each combination, we trained a IsoNet-2 model on the lateral slices, applied it to the axial slices as described, and computed the reconstruction error compared to ground-truth (PSNR). As can be seen from Fig. 4.19, the restoration achieves almost constantly high PSNR values for $\sigma \leq 4$ and FWHM $\leq 8px$ in both cases, after which it slowly degrades for higher values. Interestingly, for a given σ the error is not minimal for a vanishing PSF width, but becomes smallest at the presence of a small, yet noticeable, PSF width. A possible explanation could be, that a finite axial blur allows information in one plane to leak into neighboring planes and thus be useful for the restoration, rather than be cut out completely.

4.5.2 Application to Fluorescence Microscopy Acquisitions

After having validated our approach on synthetically generated volumes, we will in the following consider several real biological data sets imaged with different fluorescence imaging techniques. Specifically, will apply our strategy to increase axial resolution on *i)* mouse liver stacks acquired by confocal microscopy, where we will additionally study the effect of subsampling on the restoration error, *ii)* images of *Drosophila* embryos and *C. elegans* samples acquired with light-sheet microscopes, where we additionally quantify the resolution gain in the Fourier domain, and *iii)* zebrafish retina acquired via spinning disk confocal microscopy, where we will demonstrate that CARE networks can leverage multi-channel information.

Confocal acquisitions of mouse liver tissue

Here we used volumetric images of mouse liver tissue (Fig. 4.20). On a tissue level, the liver consists mainly of protein-producing hepatocytes that exchange molecules with complex three dimensional tubular networks[14] which are responsible for transport of blood into, and bile out of the organ (Morales-Navarrete et al. 2015). Imaging

[14] These tubular networks are the sinusoidal network and the bile canaliculus.

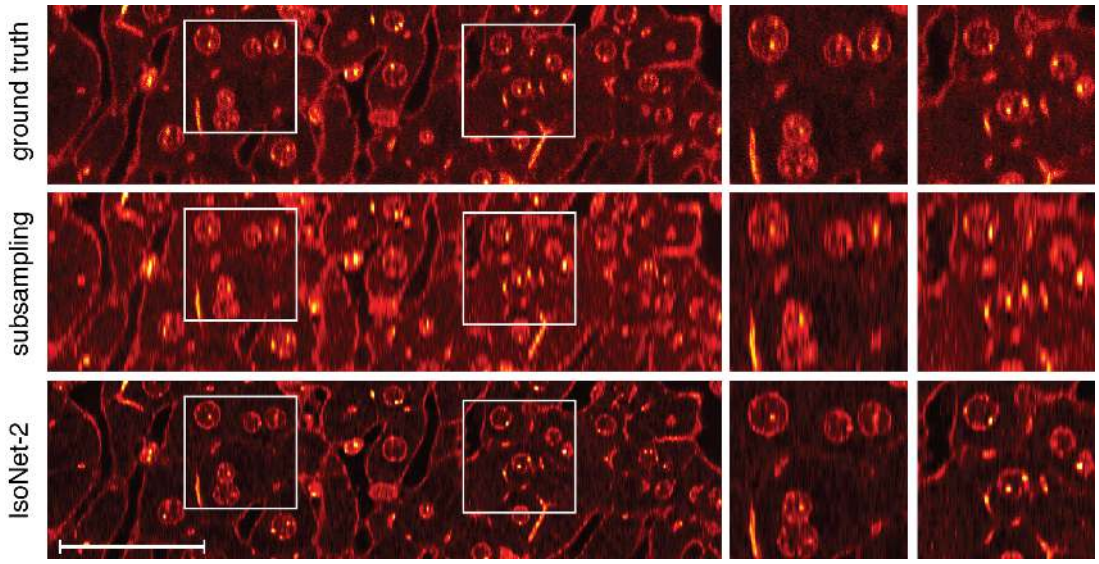


Figure 4.20: Results on fluorescence microscopy images of liver tissue (data from Morales-Navarrete et al. 2015). Nuclei (DAPI) and membrane (Phalloidin) staining of hepatocytes, imaged with a two-photon confocal microscope (excitation wavelength $780nm$, $NA = 1.3$, oil immersion, $n = 1.49$). We start from an almost isotropic acquisition (ground-truth), simulate an anisotropic acquisition (by downsampling $\sigma = 8$ fold axially), and compare the isotropic image to the IsoNet-2 recovered image. Shown are axial slices through the full volume. Scalebar is $50\mu m$. (Reprinted by permission from Springer Nature: Weigert et al. 2017, Copyright 2017)

and analyzing this organ and its complex structure, however, poses two main challenges: *i)* The sheer size of the tissue ($\sim 1mm^3$) requires very fast, tiled imaging, and *ii)* building size- and shape-preserving models of structures contained in the tissue require isotropic image resolution. This is particularly true for tubular structures bearing small radii, *e.g.* the bile canaliculi.

Mouse liver stacks containing hepatocytes tissue was acquired with a confocal microscope using a multi-channel labeling of nuclei (DAPI) and membrane marker (Phalloidin)[15]. To generate almost isotropic ground-truth volumes, acquisitions were carried out with isotropic pixel-size of $\Delta x = \Delta y = \Delta z = 0.3\mu m$. The signal from both channels was fused, resulting in a total of 8 stacks of size $750 \times 750 \times 400$ containing both nuclei and membrane structures (Fig. 4.20). We downsampled these stacks by a axial factor $\sigma = 8$ by retaining only every 8-th focal slice. This way, we obtained isotropic ground-truth stacks (to be compared against later), and corresponding anisotropic stacks (to be used as input for our method). The result were 8 anisotropic stacks of size $750 \times 750 \times 50$ from which we extracted ≈ 15000 patches of size 128×128 from the total body of data. We next trained a IsoNet-2/CARE model, while using a split PSF h_{split} calculated via a theoretical confocal PSF model ($NA = 1.3$, $n_0 = 1.49$), for 100 epochs as described before. The results show that the CARE model improved the axial resolution considerable, especially at image regions, where circular nuclei structures were severely blurred by the subsampling

[15] Data from Fabian Segovia Miranda (Zerial group, MPI-CBG).

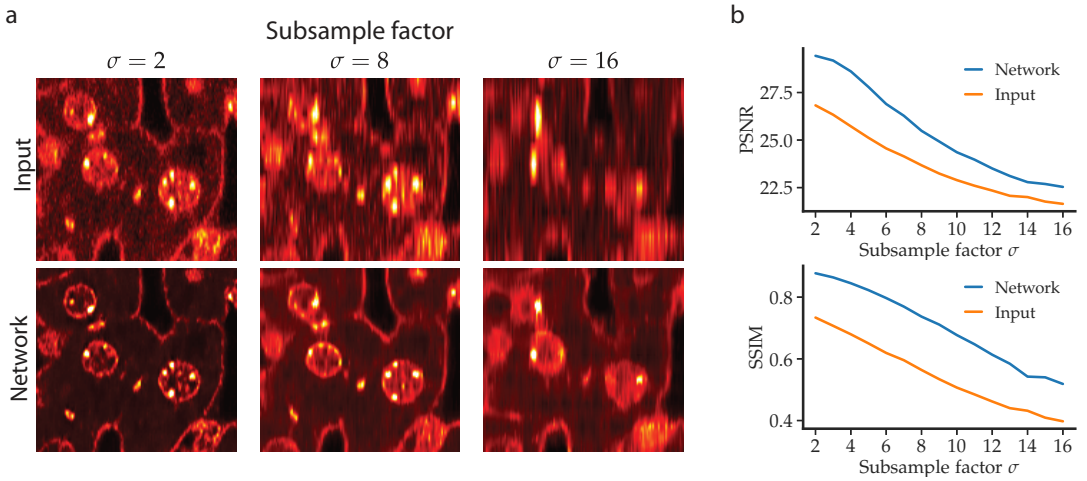


Figure 4.21: Effect of subsampling factor σ on restoration quality. a) Axial crops for both input and network reconstruction for different subsampling factors $\sigma = (2, 8, 16)$. Evidently, for larger σ information is lost in the input, which is impossible to restore for the network, *e.g.* the distribution of point like centers within the hepatocyte nuclei. b) Restoration error for $\sigma \in (2, \ldots, 16)$ compared to the isotropic ground-truth. Note that SSIM ≥ 0.7 for up to $\sigma \leq 10$.

process (Fig. 4.20). It has to be noted, that even the isotropically acquired ground-truth is subject to the anisotropic PSF of the confocal microscope. Being part of the synthetic generation process of the training data, this was partly corrected for by the network reconstruction, as suggested by the restoration of point like emitters within the hepatocyte nuclei in Fig. 4.20. As for this example almost isotropic ground-truth is available, we next studied the effect of the subsampling factor σ on the restoration quality. For that, we created artificially subsampled stacks with different $\sigma \in (2, \ldots, 16)$ and trained a restoration network for every σ as described before. We then quantified the error by computing the normalized PSNR and SSIM of both the input and the CARE reconstruction compared to the isotropic ground-truth. As can be seen in Fig. 4.21, the reconstruction quality is acceptable both visually and quantitatively (SSIM ≥ 0.7) for up to $\sigma \leq 10$. For larger σ information is already lost in the input, which is then impossible to restore for the network, *e.g.* the distribution of point like centers within the hepatocyte nuclei in the case of $\sigma = 16$.

Light-sheet acquisitions of *Drosophila* and *C. elegans*

As a next example, we consider volumes of *Drosophila melanogaster* embryos and *C. elegans* worms, acquired with a light-sheet microscopes. As the available data was already acquired with anisotropic pixel-sizes (*i.e.* $\Delta z > \Delta x/y$), we cannot compare against isotropically acquired ground-truth in this case. Rather, we will assess the perceptual isotropy of the recovered stacks, and investigate the effective resolution gain by inspecting the Fourier spectrum of the restoration results. In Fig. 4.22 we show stacks from two different sample *Drosophila* and *C.elegans* recordings where we trained *IsoNet-2* to restore the raw XZ slices as described before. *Drosophila*

[16] Data from Pavel Tomancak (MPI-CBG).

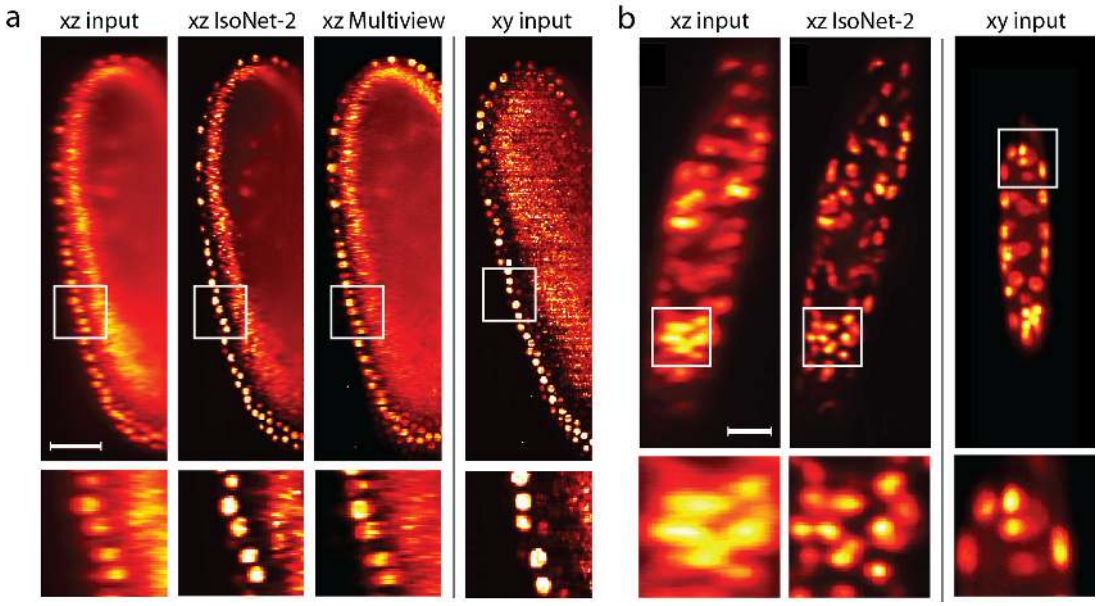


Figure 4.22: Isotropic restoration pipeline applied to a) *Drosophila* and b) *C. elegans* volumes. The image quality of the recovered axial (xz) slices is significantly improved and shows similar isotropic resolution when compared to the lateral (xy) slices. In b) we additionally compare to the result of multiview deconvolution (Preibisch et al. 2014). Scalebar a) 50 μm, b) 10μm. (Reprinted by permission from Springer Nature: Weigert et al. 2017, Copyright 2017)

images[16] were acquired with a Zeiss LZ-1 light-sheet microscope using a $20\times/1.0$ Zeiss Plan Apochromat objective and lateral/axial pixel size of $\Delta x = 0.286\mu m$ and $\Delta z = 1.00\mu m$, resulting in an subsampling factor of $\sigma = 3.5$. The *C. elegans* data was recorded with the same instrument[17], and with lateral and axial pixel size of $\Delta x = 0.091\mu m$ and $\Delta z = 0.577\mu m$, *i.e.* a subsampling factor of $\sigma = 6.34$. As depicted in Fig. 4.22, the final network reconstructions exhibit perceptual sharpness along the axial dimension (XZ) close to that of the higher quality raw lateral (XY) slices, demonstrating the ability to restore isotropic resolution from a single volume translates to different experimental settings. Additionally, we compare in Fig. 4.22b the network result with multi-view deconvolution results (Preibisch et al. 2014) obtained from 6 rotated views of the same specimen. In this case, we observe a similar good axial quality compared to the multi-view reconstruction, yet with using only a single (network) anisotropic stack instead as input.

Furthermore we consider live imaging of developing histone-labeled *Drosophila melanogaster* embryos imaged using an adaptive light-sheet microscope (Royer et al. 2016). Note that this data set was already processed (Royer et al. 2016), but still exhibited an anisotropic PSF ($4\times$) and axial sub-sampling ($5\times$) that translated into a combined 4–6 fold decrease in axial resolution (Fig. 4.23). We followed the training data strategy explained before, using a theoretical PSF for the given microscope parameters and the given sub-sampling rate $\sigma = 5$. Training was performed on patches of size 128×128 pixels, sampled from 15 equally spaced time-points during

[17] Data from Stephanie Merrett (Sarov goup, MPI-CBG).

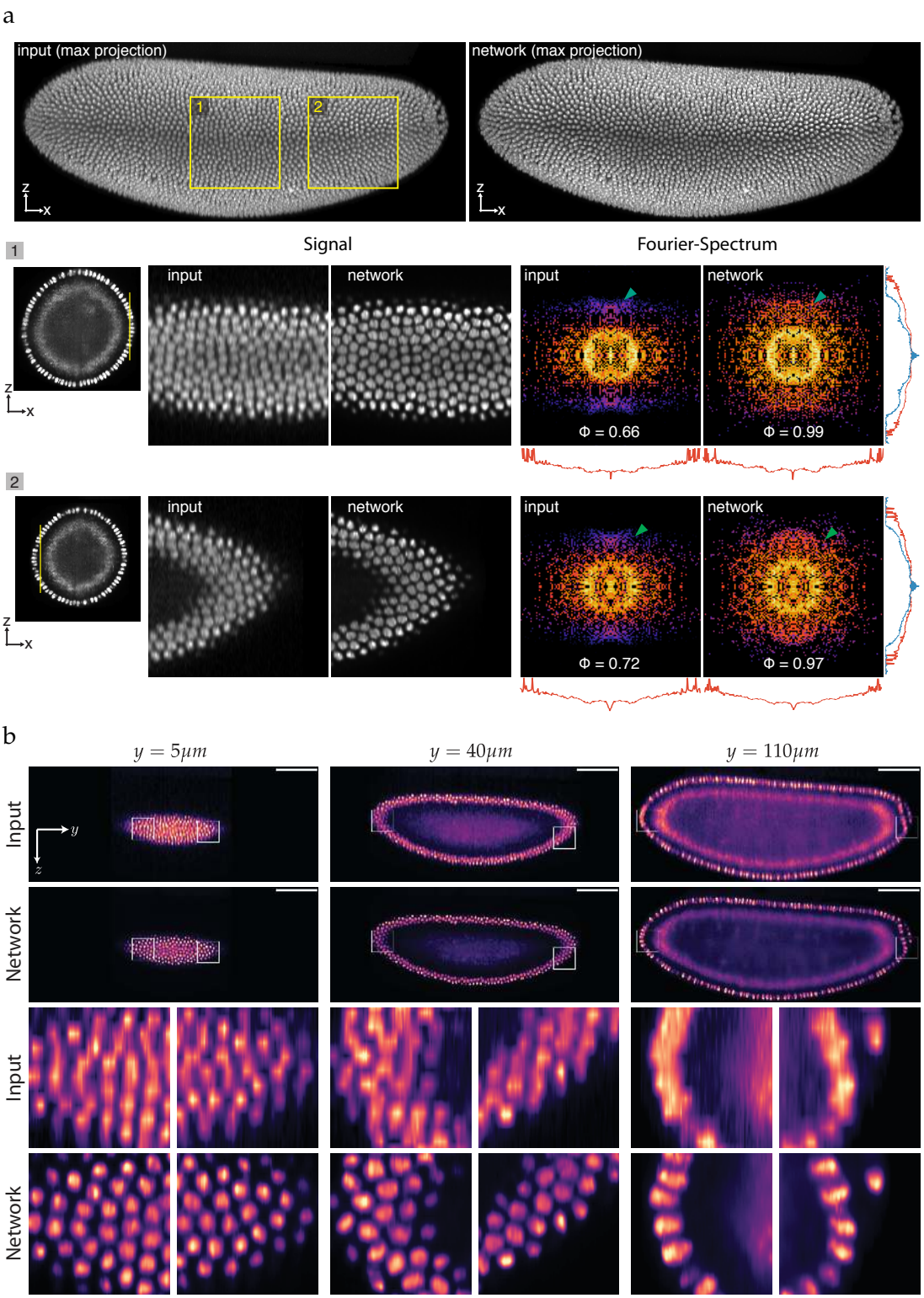


Figure 4.23: Isotropic reconstruction of time-lapse acquisitions of *Drosophila melanogaster* from (Royer et al. 2016). a) Axial resolution restoration of the original data that exhibited $\sigma = 5$-fold under-sampling in z. Restoration quality as quantified by analyzing the ratio Φ of axial *vs.* lateral spectral energy in Fourier space: $\Phi = 0.66/0.72$ (Input), $\Phi = 0.99/0.97$ (Network). Note that the missing axial information along the z-axis in Fourier space is filled up by the restoration. b) Restoration of the same volume where we additionally removed every 2nd axial slice, resulting in vastly anisotropic stacks ($\sigma = 10$). We show axial slices at different lateral (y) positions of the input and the network reconstruction, after training a new model. Scalebar $100\mu m$. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

development (between embryo cellularization and germband retraction). The final number of use patches was $\sim$ 10000. In Fig. 4.23a we show that the network prediction appear isotropic and the distinctive pattern of crowded nuclei can be clearly seen (*cf.* Supp. Video 14). To quantify the quality of this image restoration, we computed the spectral energy of the signal in the Fourier domain (Fig. 4.23a) along the axial and lateral dimension and found that the network restored previously absent frequencies along the axial dimension without altering the spectral components along the lateral dimensions. Specifically, we calculated the *spectral isotropy ratio* Φ of the axial vs lateral spectral energy and found that the restored volume achieved higher spectral isotropy ratio $\Phi = 0.99$ than the raw volumes $\Phi = 0.66$. Finally, we created from the same time-lapse an additional data set by removing every 2nd axial slice form each stack, resulting in vastly anisotropic volumes ($\sigma = 10$). After following the outline training data synthesis and training, we observed that even in this severe case of axial undersampling, the characteristic nuclear packing structure can be recovered throughout the whole embryos (Fig. 4.23b).

Video 14

4.5.3 Multi-Channel Acquisitions of Zebrafish Retina

As a last example of this section, we apply the IsoNet-2/CARE method to multi-channel acquisitions of the developing eye of *Danio rerio* (zebrafish) embryos[18]. This system is an important model of vertebrate organ formation, displaying diverse phenomena of cell migration and coordinated morphological tissue changes (Kwan et al. 2012; Sidhaye et al. 2017). The analysis of cell specific migration patterns relies on the correct tracking and segmentation of individual cells, both directly affected by the attainable volumetric resolution (Matejčić et al. 2018). The study of spatial neuron reorganization during optical cup formation, for example, is currently only possible for sparsely labeled subsets of cells (Icha et al. 2016). Since the anisotropy is a prime cause for automated segmentation methods to fail, our method is an important step towards resolving such issues and will help to analyze ubiquitously labeled tissues. While isotropy facilitates segmentation and subsequent quantification of shapes and volumes of cells, vessels, or other biological objects of interest, higher imaging speed enables imaging of larger volumes and their tracking over time.

For the restoration experiments in this case, we used a transgenic zebrafish (*Danio rerio*) line where both nuclei (DRAQ5, far-red label) and nuclear envelope (GFP-tagged LAB2P) were labeled. This allows to further investigate the ability of CARE to simultaneously restore *multi-channel* volumes. Imaging of agarose mounted embryos was performed on a spinning disk confocal (Andor Revolution WD) with a $60\times$ /1.3 objective, using excitation wavelengths of $\lambda = 638nm$ (DRAQ5 nuclei channel) and $\lambda = 488nm$ (nuclear membrane GFP channel). The lateral and axial pixel-sizes were $\Delta x = 0.195\mu m$ and $\Delta z = 2\mu m$, resulting in an axially under-sampling factor of $\sigma = 10.2$. For generating training data we acquired 5 multi-channel volumes of

[18] Experimental data acquired by Mauricio Rocha Martins, transgenic line created by Marija Matejcic (Norden group, MPI-CBG).

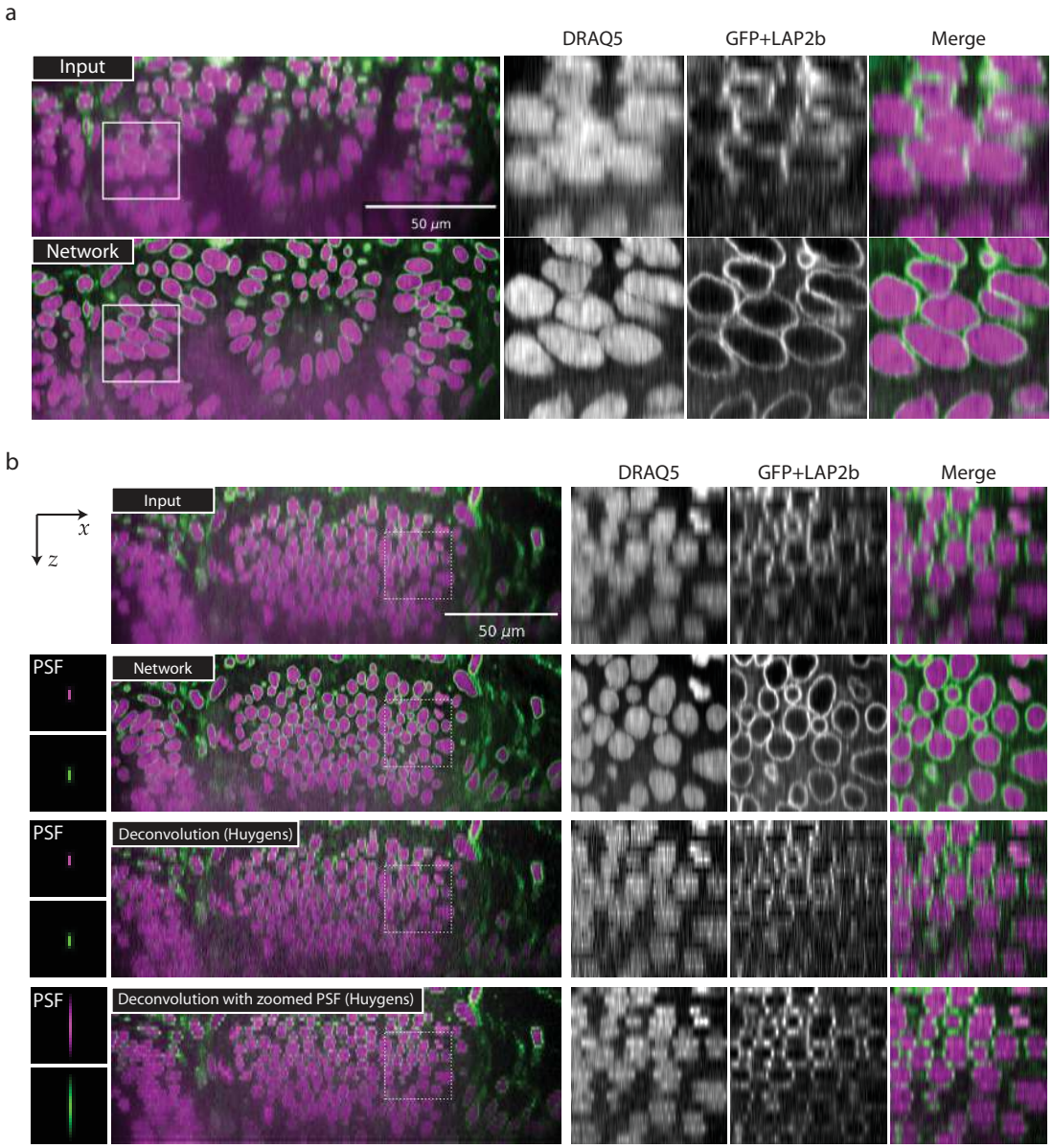


Figure 4.24: Application to multi-channel stacks of zebrafish retina. a) An axial slice through a zebrafish retina in the anisotropic raw data (top row) and the isotropic restoration with CARE. Nuclei are stained with DRAQ5 (magenta) and the nuclear envelope is labeled by GFP+LAP2b (green). b) Comparison of network restoration result with commercial deconvolution software. We compare the network output (second row) with deconvolution results obtained with `Huygens` (Scientific Volume Imaging, http://svi.nl) once with the actual PSF (third row) and once with an upsampled PSF (last row). We used the following parameters from `Huygens`: method = MLE, number iteration = 70, SNR parameter = 15, quality threshold = 0.05. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

the developing eye. We extracted lateral patches as before and applied the corresponding PSF and subsampling model, yet always keeping the information of *both* image channels. In total we used $\sim$ 25000 patches of size $128 \times 128 \times 2$ (where the last dimension equals the number of channels). After training, we again applied the network to the lateral slices of the two-channel image (Fig. 4.24a). Indeed, respective CARE networks deliver the desired axial resolution from $\sigma = 10.2$-fold fewer axial

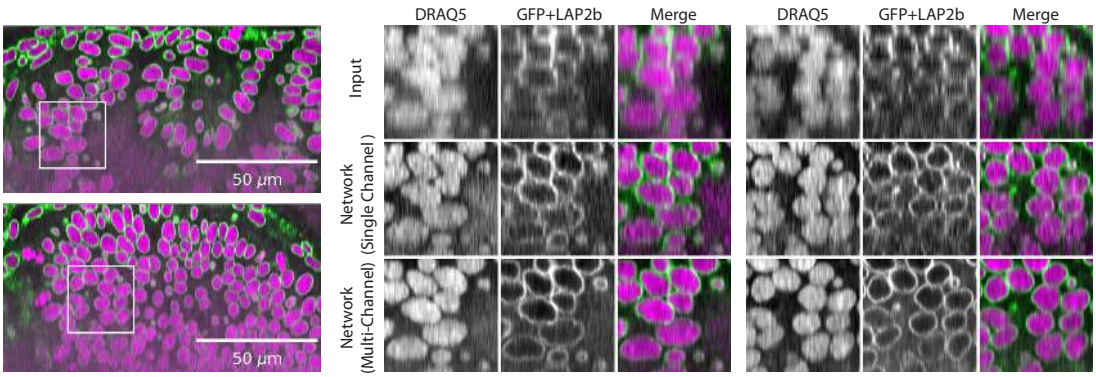


Figure 4.25: Comparison of multichannel *vs*. single channel restoration. Nucleus (DRAQ5, magenta) and nuclear envelope (GFP+LAP2b, green) labeled volumes of a developing *Danio rerio* retina. The volumetric stacks exhibit $\sigma = 10.2$ anisotropic z sampling. Different (xz) slices for different stacks/positions are shown, for the input (top row in each sub-panel), the network reconstruction when training a separate network for each channel (middle) and when training a network that restores both channels simultaneously (bottom row). The reconstruction quality of the multi-channel network is visibly superior compared to the single-channel case, showing that cross-channel information is beneficial for restoration. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

slices (Fig. 4.24a & Fig. 4.25), and were able to restore faint membrane signal from highly distorted inputs (*cf*. Supp. Video 15). We additionally compare against classical deconvolution performed with `Huygens`[19]. Here, we volumetrically deconvolved the input using *i)* the full theoretical confocal PSF corresponding to illumination and excitation wavelength for each channel, and *ii)* additionally elongated versions of the PSFs in an attempt to incorporate the additional blur induced by the anisotropy (by first axially down- and upsampling each PSF by the subsampling factor σ). We used the following parameters from `Huygens`: method = MLE, number iteration = 70, SNR parameter = 15, quality threshold = 0.05. The result in Fig. 4.24b clearly show, that pure deconvolution fails to recover any additional signal contrast and resolution. This is expected, as the used deconvolution methods disregard distortive effects of subsampling in their respective image formation models. An interesting observation is, that areas in which one channel showed a higher level of distortion, both channels where recovered similarly well. This means that the network leveraged on learned correlations between the channels and uses this cross-channel prior during image reconstruction (see Fig. 4.25).

Video 15

Taken together, increasing isotropic resolution through CARE networks, trained on semi-synthetic pairs of images, benefits both imaging speed and image quality in many biological applications. Moreover, since training data can computationally be derived from the data to be restored, this method can be applied to any already acquired data set.

[19] Scientific Volume Imaging, http://svi.nl

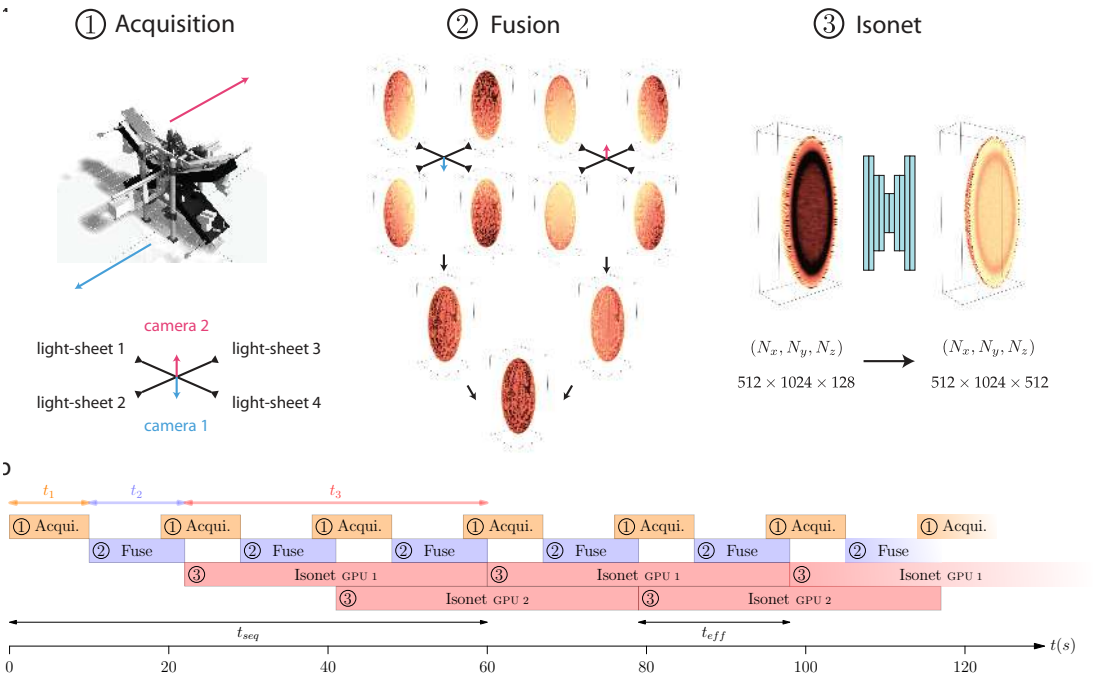


Figure 4.26: Demonstration of real-time application. a) Custom light-sheet microscope with 2 detection and 4 illumination arms for recordings of *Drosophila* embryos. The image acquisition and processing pipeline entails recording of 8 views that subsequently get fused to $\sigma = 4$ fold axially anisotropic stacks of size $\approx 512 \times 1024 \times 128$ and finally isotropically reconstructed via CARE. b) Interleaving network prediction on two GPUs renders the entire pipeline real-time compatible.

4.5.4 Real-time Application on a Custom Light-Sheet Microscope

We finally observe that the described isotropic restoration method can be evaluated fast enough even on a single workstation to be compatible with the typical acquisitions speeds of a light-sheet microscope acquiring time-series of developing embryos. To demonstrate this, we incorporated the isotropic pipeline alongside a custom light-sheet microscope[20] designed for the observation of developing *Drosophila* embryos (Fig. 4.26a). Here for every time point a volumetric image is obtained by *i)* acquisition of 8 images from 2 cameras and 4 light sheets, which are then *ii)* computationally fused to yield a anisotropic stack of typical size $512 \times 1024 \times 128$ and anisotropy factor $\sigma = 4$, *i.e.* 128 focal planes with a pixel-sizes of $\Delta x = 0.38$ laterally and $\Delta z = 1.52$ axially. Experimental considerations and the time necessary to acquire all different views lead to a typical acquisition rate of $\Delta T \approx 20s$ per stack.

Importantly, the evaluation of the full isotropic restoration pipeline yielding a final isotropic stack of size $512 \times 1024 \times 512$ takes $\approx 33s$ on a single GPU (NVIDIA GTX 1080). This includes the upsampling by $\sigma = 4$ from the original stack size and the subsequent two independent passes of the neural network[21] over the two possible axial orientations (xz/yz). We note that this is less than twice the time needed for recording of the stack, these computations are independent of the acquisition, and thus can be trivially parallelized over the individual stacks of a time-series.

[20] Developed by Nicola Maghelli, Robert Haase, Uwe Schmid (MPI-CBG), and Loic Royer (CZ Biohub).
[21] Standard residual U-Net with depth 2, kernel size 5×5, and initial filter size 16.

We confirmed this by setting up a dedicated workstation with two NVIDIA GTX 1080 GPUs that was able to perform the isotropic restoration pipeline in real-time on stacks acquired by the network connected microscope machine, by interleaving acquisition and distributing computation on the two available GPUs (Fig. 4.26b).

4.6 Restoration of Resolution with Synthetic Training Data

Having seen the potential of using *physically* acquired and *semi-synthetically* created training data for CARE, we next investigate whether reasonable restorations can be achieved from *purely synthetic* generated image data alone, *i.e.* without involving any real microscopy data during training. This situation is relevant for experimental scenarios, where neither real biological ground-truth nor the real image degradation process is accessible via pairs of experimental acquisitions, but a synthetic generative model of both the structures under investigation and the image degradation process is available. For example, when using conventional widefield microscopy to image stereotypical biological structures such as microtubules, which are smaller than the diffraction limit yet whose physical properties are well known. Although many biological applications require resolving sub-diffraction structures in the context of live-cell imaging, most super-resolution imaging modalities (*e.g.* STED, STORM, PALM) that would achieve the necessary resolution suffer from low acquisition rates, high photo-toxicity, and excessive photo-bleaching. On the other hand, widefield imaging offers the necessary speed, but lacks the required resolution. We thus will in this section ask, whether CARE restoration networks can be used to computationally resolve sub-diffraction structures from widefield images alone, given those structures and the image model can be suitably modeled *in-silico*. The main idea is to *synthetically* create ground-truth images that are realistic representations of the structures to be examined, and generate *synthetically* degraded microscopy images via simulation of the imaging process. Since the creation of realistic ground-truth structures is a complex task in its own right (Rajaram et al. 2012; Venkataramani et al. 2016; Svoboda et al. 2017), this training regime works best in cases when a simulation pipeline is already available or when the structures to be synthesized are highly stereotypical (*e.g.* tubular or point like objects). Note that this is a fundamentally different approach compared to recently proposed methods for single molecule localization microscopy that reconstruct a single super-resolved image from multiple diffraction limited input frames using deep-learning (Nehme et al. 2018; Ouyang et al. 2018). In the following, we will focus in generative models of tubular and point-like structures that are commonly studied in biology. In order to obtain synthetic image pairs, suitable for training CARE networks, we used these generated structures as ground-truth, and computationally modified them to resemble actual microscopy data (Fig. 4.27). Specifically, we created synthetic ground-truth images of tubular meshes resembling *microtubules*, and point-like structures of various sizes mimicking *granules*. Then we computed synthetic input images by simulating the image degradation process by applying a PSF, camera noise, and background auto-fluorescence (Fig. 4.27).

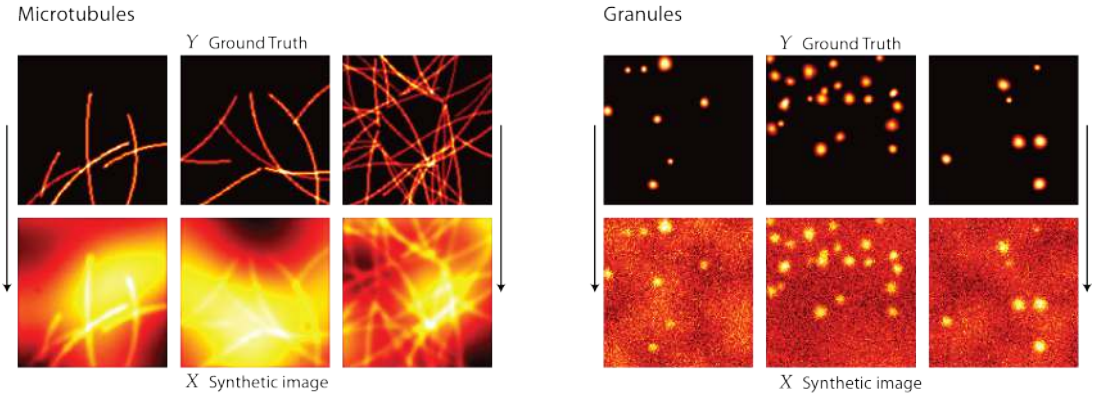


Figure 4.27: Synthetically generated training data for artificial microtubules structures and granules. We first generate the ground-truth images (Y) by tracing tubular bundles of different sizes and length (left) or simulating granular regions (right) on pixel grids of size 128×128 and pixel-size $0.08\mu m$. We then simulate the synthetic microscopy image (X) by adding low frequency Perlin noise resembling background fluorescence, convolving the resulting image with a microscope PSF, and adding Gaussian and Poisson camera noise.

Generation of synthetic microtubules and granules: For creating artificial images of a single microtubule structure, we first simulate two-dimensional trajectories $\{x_n\}_{n\in\mathbb{N}}$ on a pixel grid, with randomly changing orientation and given maximal curvature κ_{max}, mimicking the physical stiffness properties of microtubules (*e.g.* having a non-vanishing persistence length). To render the computation conceptually easier, we will describe a point in the Euclidian plane as given by a complex number $x_n \in \mathbb{C}$ and regard its real and imaginary parts as the corresponding coordinates in $\mathbb{R}^2$. For that we first generate a random walk on the curvatures $\kappa_n \in \mathbb{C}$:

$$\kappa_n = \exp\left(2\pi i \cdot \text{clip}_{\kappa_{max}}[\kappa_0 + \mathcal{W}_n(d\kappa)]\right), \quad \mathcal{W}_n(d\kappa) \sim \sum_i^n \mathcal{N}(0, d\kappa)$$

which we successively sum up to yield velocities and finally the trajectory itself:

$$v_n = v_0 \exp\left(2\pi i \cdot \sum_{i=0}^{n} \kappa_i\right) \qquad x_n = x_0 + \sum_{i=0}^{n} v_i$$

Many of these single tubules trajectories are then rasterized on a pixel-grid of given pixel-size to yield the final, ground-truth image (Fig. 4.27). To simulate ground-truth of granular objects, we use clipped Perlin noise whose intensity we vary randomly.

Image formation model: To synthetically generate the corresponding widefield image, we simulate auto-fluorescence by adding low-frequency Perlin noise as background, blur the result with the PSF of the microscope (we use the theoretical PSF for the respective NA of the considered experiment) and add Poisson and Gaussian noise, mimicking camera noise (Fig. 4.27).

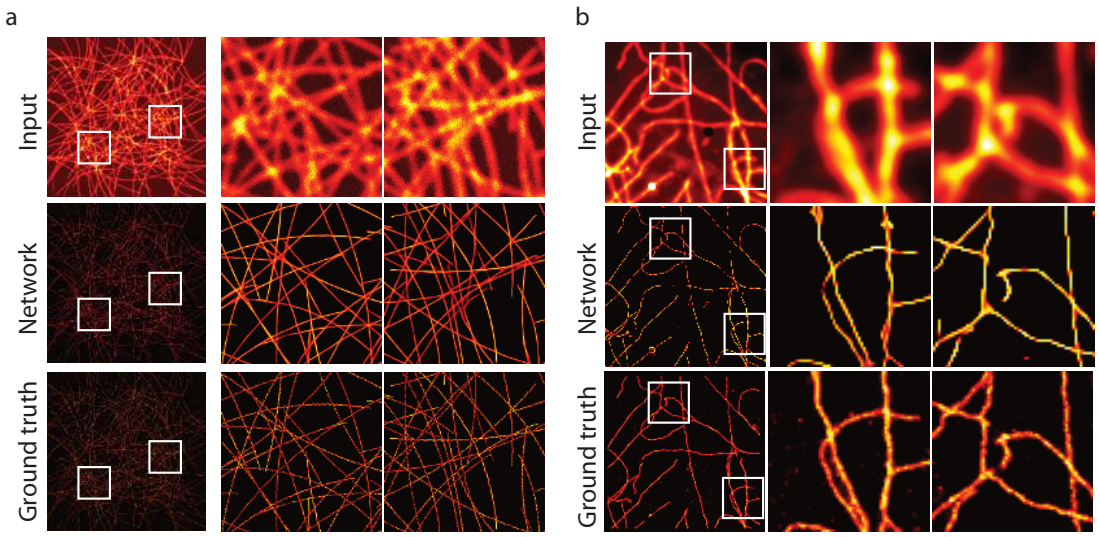


Figure 4.28: Evaluation on microtubule test data sets. a) Test image from the synthetic image creation pipeline with a synthetic PSF (FWHM = $6px$). Shown are the simulated widefield image (top), the network restoration (middle) and the synthetic ground-truth image (bottom). As the image degradations are almost identical to the training data, the restoration is almost perfect. b) The same network applied to an experimental Tubulin data set from a public SMLM challenge (Sage et al. 2015).

4.6.1 Microtubule Acquisitions and Comparison with PALM

We demonstrate the outlined CARE restoration strategy on images of microtubules alone. Here we use experimental stacks from the public SMLM challenge database (Sage et al. 2015), that show tubulin labeled microtubules, and for which super-resolution reconstructions via PALM is available (Manley et al. 2011)[22]. Specifically we generated tubular ground-truth data as described above, assuming a pixel-size of $0.08\mu m$ and a Gaussian PSF with an FWHM of $0.3\mu m$, from which we created 8000 synthetic pairs of patches of size 128×128. We finally trained a network to invert this degradation process. To that end, we used a 2D residual U-Net as in Section 4.5.1 with a depth 2, convolutional filters of size 7×7, and ReLU activation, resulting in a networks with $453,503$ parameters. After training, we applied the resulting model to a synthetically generated test image, where as expected the restoration is virtually perfect (Fig. 4.28a). We then applied the same network to the widefield SMLM data described above. As can be inferred from Fig. 4.28b, the restoration result agrees almost everywhere with the super-resolution image obtained with PALM. Importantly, the network reconstruction used only a single widefield image as input, whereas the PALM reconstruction needed $\approx 27,000$ single images of sparsely blinking emitters.

4.6.2 GFP-Tubulin in HeLa Cells and Comparison with SRRF

Next, we applied the described approach on time-lapse widefield images of microtubules in HeLa cells. Here we used a data set[23] of widefield images of GFP-labeled microtubules. In this case corresponding super-resolution images reconstructed via

[22] http://bigwww.epfl.ch/smlm/datasets/index.html?p=tubulin-conjal647
[23] Data from Sian Culley and Ricardo Henriques (UCL, London)

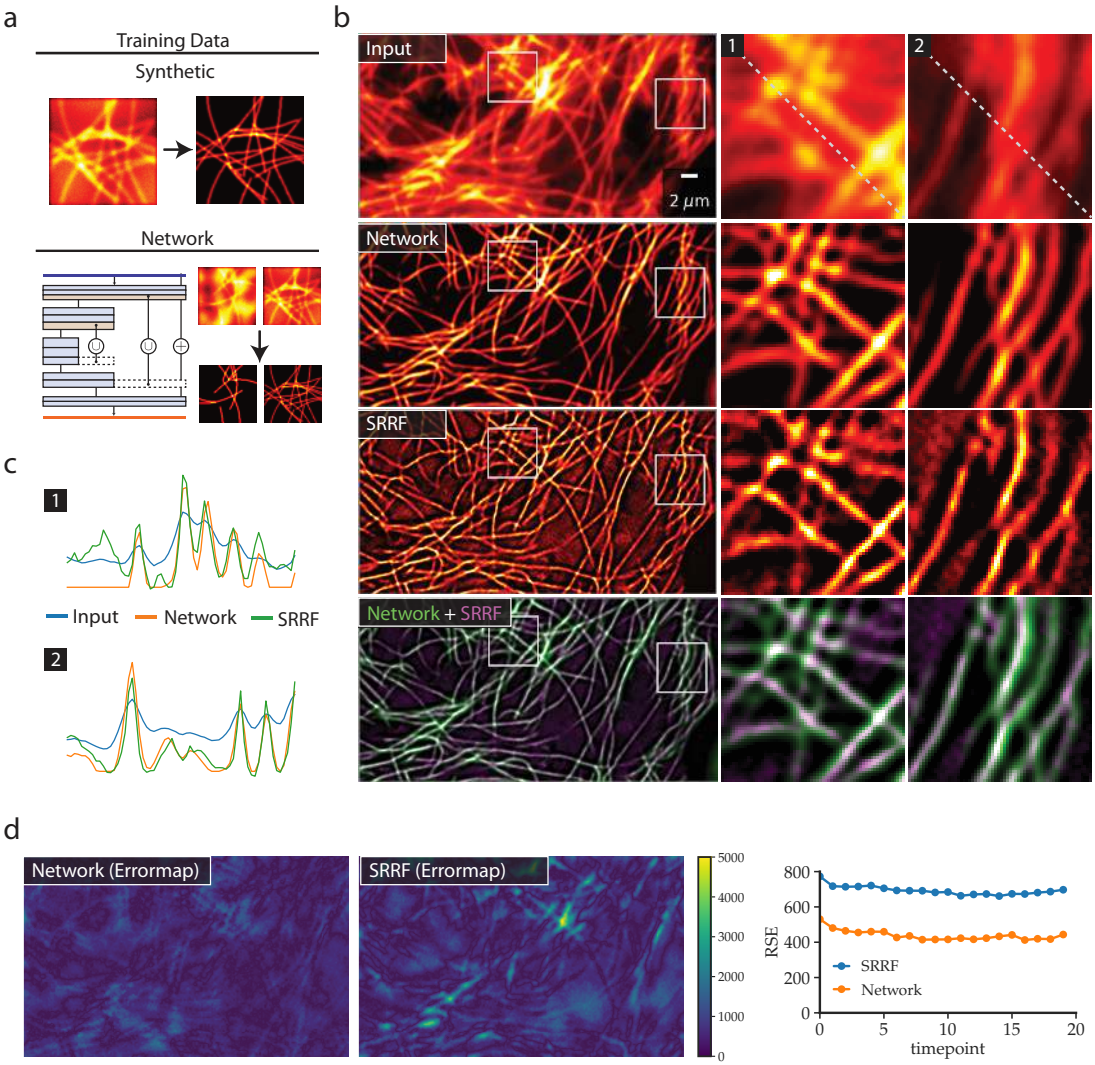


Figure 4.29: Resolving sub-diffraction structures in fluorescence microscopy images at very high frame rates with CARE. a) Schematic of the fully-synthetic generation of training data pairs as outlined in the main text. b) GFP-tagged microtubules in HeLa cells. Raw input image (top row), network restorations (second row), super-resolution images created by the state-of-the-art method SRRF (N. Gustafsson et al. 2016) (third row), and a superposition of our results with restorations by SRRF (bottom row). c) Line-plots along the diagonal of both insets. d) Quantitative assessment of reconstruction error via SQUIRREL (Culley et al. 2017). We show the error map (absolute error) for a single time point for the network and SRRF reconstruction (left) and the root mean squared error (RSE) over time (right). (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

SRRF (super-resolution radial fluctuations N. Gustafsson et al. 2016, Fig. 4.29b) are available for validation. We generated training data as described (while adapting experimental parameters, such as PSF and pixel size), resulting in a total of 5000 synthetic patch-pairs of size 128×128 (Fig. 4.29a). As can be seen in Fig. 4.29b & c, both network reconstruction and super-resolution (SRRF) results were able to recover detailed microtubule structures that are indistinguishable in the raw images (*cf.* Supp. Video 17). Additionally, the reconstruction quality of both methods can be quantitatively assessed by calculating error maps with SQUIRREL (Culley et al. 2017). As can be seen in Fig. 4.29d, the error of the network reconstruction is $\approx 30\%$ lower

Video 17

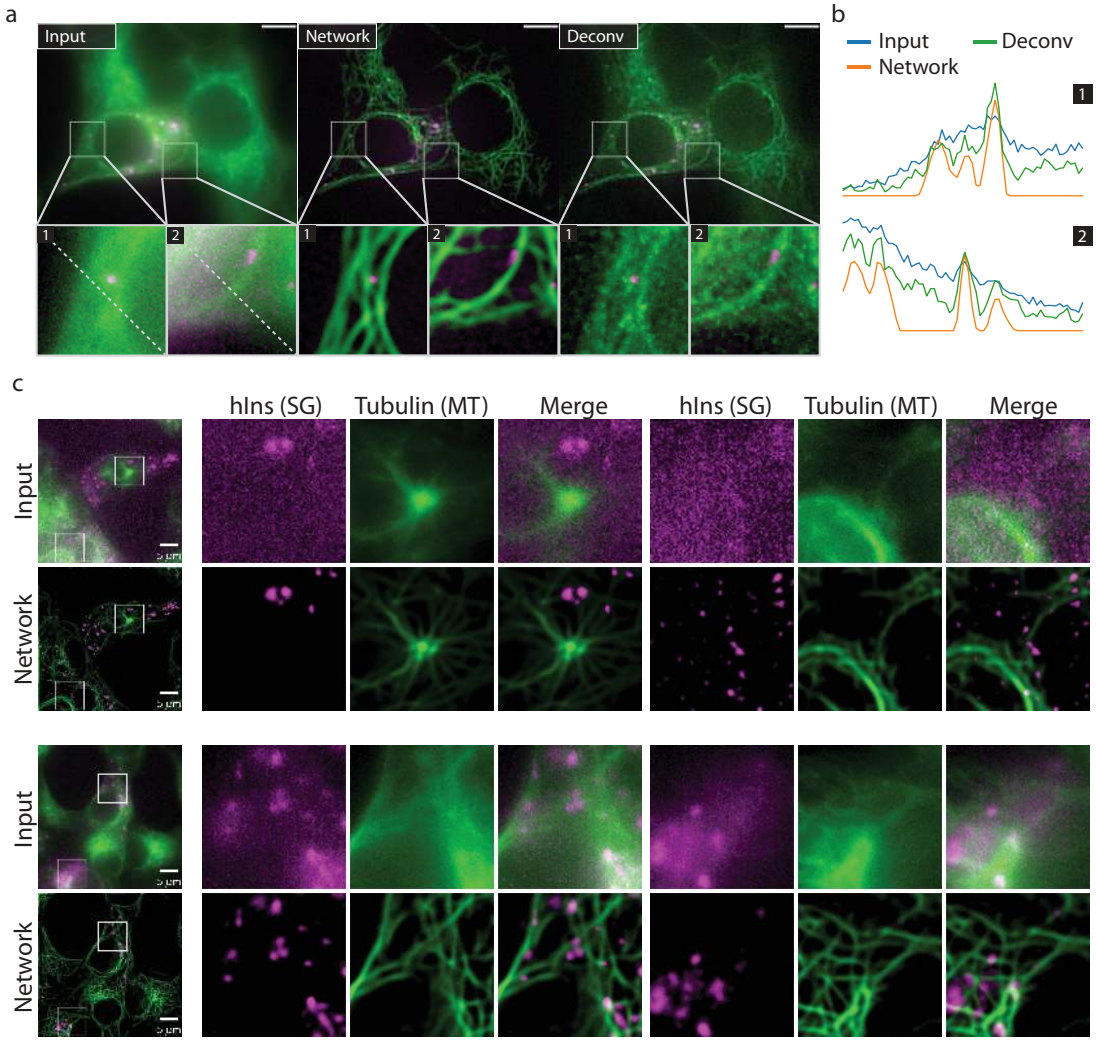


Figure 4.30: a) Live-cell two-channel imaging of rat secretory granules (pEG-hIns-SNAP, magenta) and microtubules (SiR-tubulin, green) in insulin-secreting INS-1 beta cells. Shown are the raw widefield images acquired with a DeltaVision OMX microscope (left), the corresponding network restorations (middle), and a deconvolution of the raw image as a baseline (right). Scalebar $5\mu m$. b) Line-plots along the diagonal of the insets shown in a). c) Reconstruction results for 2 additional cell acquisitions. (Adapted with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

than for SRRF for all time points. Interestingly, the regions of greatest error are different for both methods: Whereas SRRF exhibits the largest error in (high density) areas with crossing microtubules, the CARE network shows greatest discrepancy for background regions, which were suppressed in the reconstruction. This is expected, as the synthetic training data consisted of ground-truth without background. Note, that the input of SRRF consisted of 200 consecutive frames of widefield images (not shown), whereas the CARE result is based on a single image obtained from averaging only $5-10$ widefield frames, thus enabling 20 fold faster imaging.

4.6.3 Multi-channel INS-1 cells with Microtubules and Secretory Granules

We finally consider multi-channel images where tubular and additionally *granular* structures are present. Specifically, we will investigate images of insulin secretory granules (SG) and filamentous microtubule-networks (MTs) in insulin-secreting beta-cells (INS-1). The role of MTs in insulin SG exocytosis has been heavily debated recently (X. Zhu et al. 2015) and high resolution live-cell imaging is crucial for understanding the function of the MT network therein. Due to its resolution limit, widefield imaging only allows for restricted insights into this process. Live-cell super-resolution imaging (*e.g.* SIM), on the other hand, requires high laser power resulting in quick photo-bleaching of the fluorescent signal. We used 2-channel widefield time-lapse images of rat INS-1 cells[24] where SG and MT were labeled with 1 μM SNAP-Cell 505-Star and SiR-tubulin respectively. Imaging was done with the DeltaVision OMX (GE) microscope using an Olympus Plan Apochromat $60\times/1.43$ objective. Time-lapse movies were acquired in wide-field mode with 50 ms exposure-time for each channel resulting in a final speed of 2 frames per second (fps). To apply our method training data was firstly generated for each channel separately, by creating 8000 synthetic patch-pairs of size 128×128 with the tubular generative model and the granular model described earlier. Then a network was trained for each channel (SG/MT) independently, and finally applied on the respective channel of the widefield images. The network results are shown in Fig. 4.30. As can be observed, the restoration of both microtubules and secretory granules exhibit a dramatically improved resolution, revealing structures imperceptible in the widefield images (Fig. 4.30 & Supp. Video 16). We additionally compared the CARE restoration to the results obtained by deconvolution via SoftWorkx (shipped with the DeltaVision OMX microscope). Line profiles through the data show the improved performance of CARE network over deconvolution (Fig. 4.30b).

Video 16

Taken together, these results suggest that for structures that are straight-forward to model, such as microtubules or granules, CARE networks can enhance widefield images to a resolution usually only obtainable with super-resolution microscopy, yet at considerably higher frame rates.

4.7 Limitations

In the previous sections we have shown that with the right training data, CARE networks perform remarkably well on a wide range of image restoration tasks, opening new avenues for biological observations. CARE networks are generally trained for a specific biological organism, fluorescent marker and microscope setting. When

[24] Sample preparation and image acquisition done by Andreas Müller (Paul Langerhans Institute Dresden)

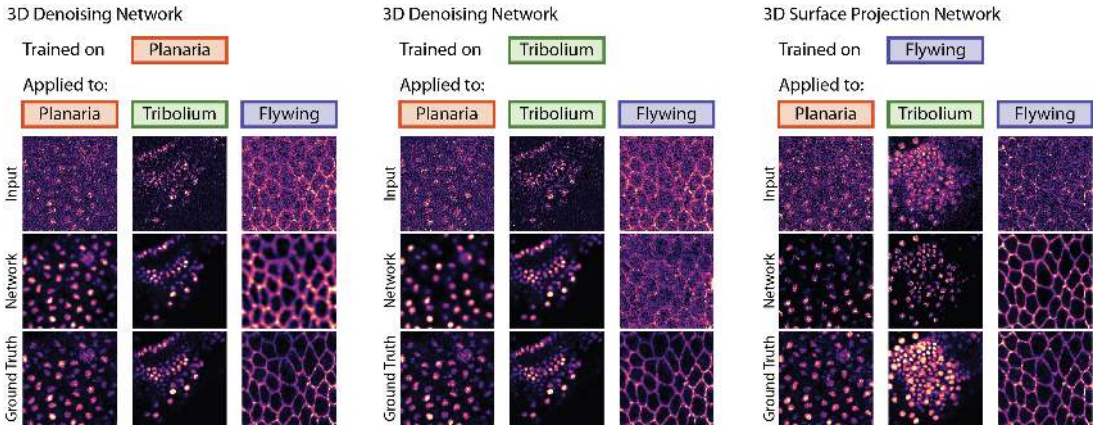


Figure 4.31: Cross-application of networks. For the planaria, tribolium, and fly wing data we applied the respective trained network on images of all three data sets. As can be seen, the restoration quality drops in cases of mismatching training and test data (*e.g.* planaria model on fly wing data). Interestingly, the amount of 'hallucinated' structures is negligible. Note that for the two denoising networks we show a slice through the resulting 3D volume, whereas the surface projection network results in a 2D projection of the whole input volume. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

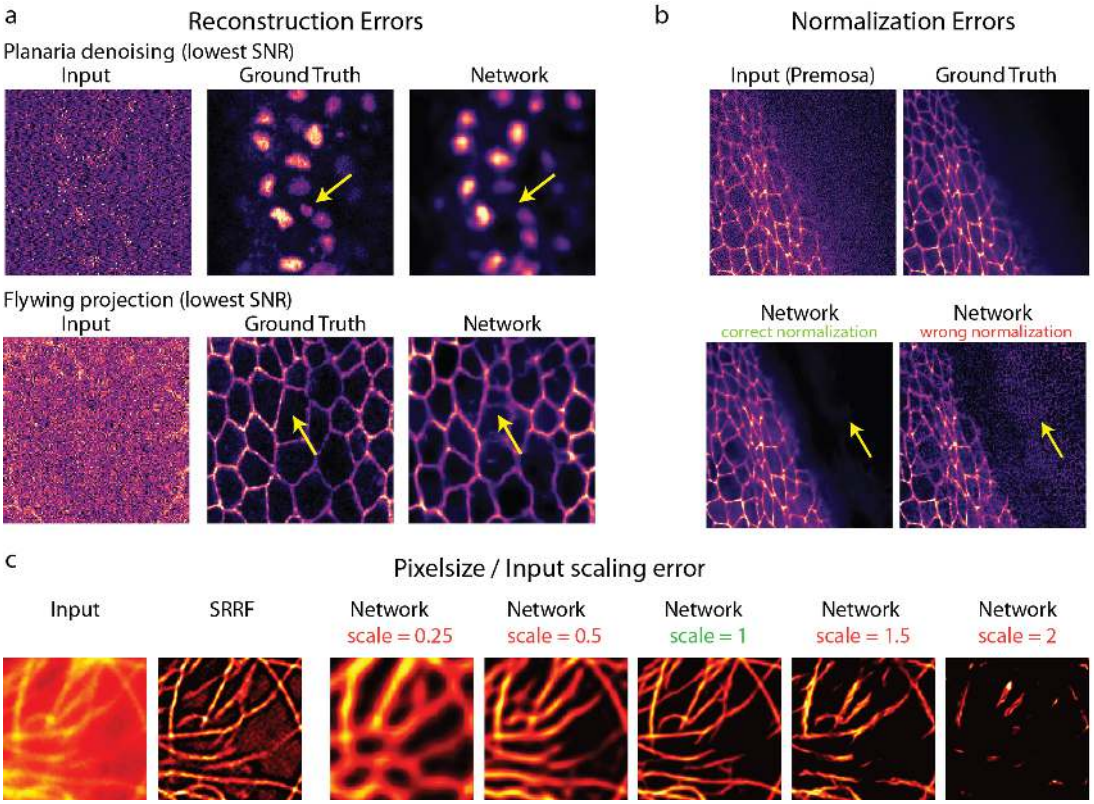


Figure 4.32: Failure cases of network reconstructions. a) Missing (top) or erroneously added (lower) structures for restorations of extremely challenging and noisy inputs. In both cases, the shown regions were selected as those that exhibit the largest relative error compared to the ground-truth over the entire test data set. b) Restoration errors due to wrong normalization of the input at test time. Here, the network was trained with inputs being normalized with percentiles $(p_{min}, p_{max}) = (2\%/99.7\%)$, but evaluated with percentile normalization of $(p_{min}, p_{max}) = (1\%/30\%)$. c) Influence of wrong pixel-sizes/scaling for tubulin restoration networks. Here, the network was trained with a specific assumed pixel-size and PSF width. Applying the network to wrongly scaled inputs results in erroneous results. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

applying a network to data it was not trained for, results are likely to suffer in quality. We demonstrate this by mutually applying the denoising and surface projection models trained earlier on test data of each category in Fig. 4.31. Here, a noticeable drop in visual quality is present for the two denoising models when applied to data it was not trained for. Nevertheless, we observed only minimal 'hallucination' effects, *i.e.* structures seen in the training data that would erroneously appear in the restored images Fig. 4.31. The strongest hallucination effects occurred when training and test data considerably differ in resolution while containing very specific structures that exhibit little variability (Fig. 4.32). Otherwise, we observe similar effects only in very rare cases where structures are either so dim that they are not any longer manifesting themselves in the input images, or in even rarer cases where the background noise can be interpreted as very low-SNR structure. Example are depicted in Fig. 4.32a, where we show the two most pronounced error cases across the entire body of available test image data.

Furthermore, the proposed isotropic restoration pipeline cannot be applied in every situation. In particular, it relies on the implicit assumption that the PSF is constant throughout the image volume and that structures are distributed isotropically in the latent ground-truth images. The former can be violated when imaging deep inside tissues, where PSFs are aberrated in a spatially varying way. The latter is violated for highly stratified or polarized tissues, such as cortical columns in the brain. Additionally, the appearance of strong anisotropic noise, such as for very low SNR acquisitions with sCMOS cameras, sometimes leads to ringing artefacts. We found that our method, due to its operating on 2D slices, works best for 3D or 2D structures (nuclei, membranes, sheets) and less well for 1D structures, which poses a limitation to certain biological structures. Further, it might be less effective for genuinely anisotropic biological tissues, such as strongly stratified layers of cells with a fixed, global elongation of membranes along one dimension.

Additionally, CARE cannot be used if ground-truth can neither be physically acquired nor synthetically generated. The synthetic generation of training data could, in general, benefit from recent advances in computer vision, such as generative adversarial networks (GANs) (Goodfellow et al. 2014), which were already successfully used for single molecule localization in super resolution microscopy Ouyang et al. 2018.

4.8 Summary

In this chapter, we investigated a *data-driven* approach to image restoration problems in fluorescence microscopy, which we termed content-aware image restoration (CARE). Importantly, we proposed several distinct strategies for generating the necessary training data either *physically*, *semi-synthetically*, or *synthetically* depending on experimental considerations. On a multitude of biological relevant examples

we demonstrated that CARE significantly expands the realm of observable biological phenomena: *E.g*. flatworms can be imaged avoiding unwanted muscle contractions, beetle embryos can be imaged much gentler and therefore longer and faster, large tiled scans of entire *Drosophila* wings can be imaged and simultaneously projected at dramatically increased temporal resolution, isotropic restorations of embryos and large organs can be computed from existing anisotropic data, and sub-diffraction structures can be restored from widefield systems at high frame rate. In all these examples, CARE allows one to invest the photon budget saved during imaging into improvement of acquisition parameters relevant for a given biological problem, such as speed of imaging, photo-toxicity, isotropy, or resolution. In the future, it would be interesting to apply described strategies for image restoration with more complex downstream task, such as segmentation and tracking. Ideally this would result in efficient methods to extract semantically rich information from vastly degraded images, and thus further broaden the applicability of CARE to complex biological imaging problems. Overall, our results show that fluorescence microscopes can, in combination with content-aware restorations, operate at higher frame-rates, shorter exposures, and lower light intensities, while reaching higher resolution, and thereby improving downstream analysis.

5 Conclusion and Outlook

In this thesis, I described novel computational strategies that address several challenges in volumetric fluorescence microscopy of developing organisms: *i)* How to process and analyze large data streams of imagery within the time constraints of acquisition, which is a prerequisite for the development of *smart, adaptive microscopy* (Chapter 2), *ii)* how to computational explore the interaction of light with biological tissue which currently limits the ability to *image deeper into tissues* (Chapter 3), and *iii)* how to extract information from *imperfect observations* that are a consequence of fundamental imaging constraints in microscopy of living organisms (Chapter 4). Together, the presented methods aim to advance the concept of *computational microscopy* which would augment the available optical hardware control with powerful, data-driven computational procedures.

In Chapter 2, I demonstrated how to visualize, analyze and process the constant stream of volumetric data acquired by light-sheet microscopes in *real-time*, *i.e*. within a time period that is compatible with the typical acquisition rate of experiments and the computational budget of a typical microscope workstation. I developed fast GPU-based rendering pipelines for microscopy volumes and introduced a *low-discrepancy sampling* strategy that enables the seamless display of large data sets during acquisition. To correct for the adverse effects of noise and intrinsic image blur, I implemented and extended GPU-accelerated denoising, convolution, and deconvolution methods. Furthermore, I proposed the use of *low-rank PSF decompositions* for situations where the blurring kernel (PSF) has small support, as is the case for light-sheet microscopy. Overall, this enables the denoising or deconvolution of *e.g*. a volume of size $1024 \times 1024 \times 64$ on a single consumer grade GPU in roughly $3s$, which is within the acquisition-rate of a typical experiment. I furthermore demonstrated how image sharpness measures can be integrated into microscopy acquisition software, providing instant image quality feedback to microscope controllers. The related open-source packages – *ClearVolume* (developed with Loic Royer), *spimagine*, and *gputools* – have already found many users within the research community and will provide the computational basis for further research. Naturally, the presented methods comprise only a small part of what would constitute a truly *smart* microscope, which would extract biological information directly during acquisition and be capable of autonomously optimizing its imaging parameters during the experiment. Additional desirable features include, for instance, the automatic detection of relevant parts of the sample, or adaptive illumination schemes that adjust for the necessary signal contrast and efficiently spend the available photon budget while

preserving sample health. On a higher level, the direct semantic classification of biological structures of interest (*e.g.* cells or vesicles) during acquisition would allow to compute and store only desired biological information (*e.g.* cell lineages or vesicle tracks), thus allowing for high-throughput experiments with small memory and storage footprints.

In Chapter 3, I investigated computationally tractable methods to accurately simulate the interaction of light fields with weakly scattering biological tissue. To that end, I adapted and implemented a GPU accelerated beam propagation method that I validated on several analytical and experimental examples. That way, the propagation of arbitrary light fields through a realistically large volume of *e.g.* $1024 \times 1024 \times 1024$ can be calculated in roughly 1 s. By proposing a multiplexing scheme for parallel point spread function (PSF) calculations inside of scattering media, I demonstrated that this enables the rigorous simulations of wave-optical image formation in light-sheet microscopes, allowing to study *in-silico* several tissue induced imaging distortions such as spatially varying PSF aberrations or the bending of the incident light sheet. Using this approach, I investigated a new correlative effect of scattering contributions in tissues - the *extended memory effect* - that is of potential use to adaptive optics. I finally demonstrated the utility of the developed methods for answering biological questions by simulating the light propagation through large refractive models of retinal tissue, thereby elucidating the optical properties of the inverted nuclei architecture in mouse retina. Due to the scalar approximation used in the underlying light propagation method, however, the accuracy of field computations is likely to suffer for highly scattering media and very oblique illumination. Furthermore the simulation of highly diffusive scattering - as often encountered in biological tissue - is beyond the current domain of application. Recently, several approaches have been proposed that partly avoid these problems while still retaining the computational efficiency that is necessary for large scale simulations (Osnabrugge et al. 2016; S. Schmidt et al. 2016). Adapting these methods to the problem of whole tissue simulations for microscopy would be an interesting future line of research. Another aspect that would merit further investigation is the use of the proposed wave-optical image-formation scheme as a *forward-model* to create synthetic microscopy images that can be used as training data for data-driven image restoration methods (*cf.* Kamilov et al. 2015). An interesting idea would be, for instance, to combine such a synthetic forward model with computational methods that aim to recover the latent fluorophore signal from images acquired with multiple random illumination pattern (Mudry et al. 2012; Negash et al. 2016), or from multiple aberrated acquisitions, such as done in *phase-diversity* (Gonsalves 1982; Kner 2013). In the latter case, one could envision a purely computational alternative to hardware-based adaptive optics imaging strategies that are currently limited by the achievable acquisition speed.

In Chapter 4, I described how the recovery of images from imperfect observation can be achieved via *machine-learning* based image restoration methods. Specifically, I demonstrated that the generation of sample and imaging specific training images and the subsequent application of *convolutional neural networks* allows to incorporate both the imaging process and the complex statistical structure of images as prior knowledge. This approach, that I termed *content-aware image restoration* (CARE), turned out to be a flexible and general method to solve fundamental restoration problems in fluorescence microscopy. To account for the diverse experimental situations found in microscopy imaging experiments, I proposed different strategies to create the necessary training data *physically*, *semi-synthetically*, or *fully synthetically*, and successfully applied CARE on a variety of problems, ranging from restoration of high quality images from low signal-to-noise-ratio acquisitions, to projection of noisy developing surfaces, isotropic recovery from anisotropic volumes, and the resolving of diffraction-limited structures from widefield images alone. In many cases the application of CARE allowed to extend the range of biological phenomena observable by microscopy beyond what classical restoration methods would allow. Examples include the restoration of extremely low signal-to-noise ratio images of the species *Schmidtea mediterranea* resulting in the first practical framework for live-cell imaging of this light sensitive organism, or the isotropic restoration of anisotropic imagery elucidating the dynamics of epithelia tissue folding in the *Drosophila* wing disc (Sui et al. 2018). As single CARE models are trained for a specific biological organism, fluorescent marker and microscope setting, applying them to images vastly different from the training distribution will likely result in poor performance as is expected for most machine-learning based inference methods[1]. Ideally, a restoration model would automatically detect such problematic situations and be able to discern inputs that are drawn from the expected distribution from those that are not. This problem of *out-of-distribution detection* (Hendrycks et al. 2016) is an active area of research in the field of machine-learning, and investigating it in the context of the presented method would be worthwhile. A related aspect is to provide estimates of the *uncertainty* of network predictions that could be a result of insufficient training data, uncertaintity of the model parameters, or the ambiguity of potential solutions of a restoration problem. Recent approaches include *approximative Bayesian inference* (Gal et al. 2016), *ensemble* methods for capturing *disagreement* between independently trained models (Guo et al. 2017; Lakshminarayanan et al. 2017; Weigert et al. 2018a), or the design of *invertible neural networks* that predict additional latent (hidden) variables relating to the several solutions of ambiguous problems (Ardizzone et al. 2018). In the context of restoration pipelines that involve the (*physical*) acquisition of corresponding pairs of corrupted and ground-truth images, it has recently been shown that - for certain types of degradations - the (sometimes arduous) assembly of ground-truth data can be avoided altogether, and models can be trained from multiple corrupted versions of the same image (*noise2noise*, Lehtinen et al. 2018). It would

[1] *Cf.* the so-called *no free lunch* theorem of learning algorithms (Wolpert 1996).

be interesting to investigate such a strategy for microscopy modalities where experimental constraints preclude the acquisition of high-quality images even in principle. Overall, the presented results show that fluorescence microscopes setups which are dedicated to a specific biological sample and imaging modality, can be operated at higher frame-rates, shorter exposures, and lower light intensity regimes while still maintaining the same image quality. Given the general trend in GPU accelerated computing power, it will soon be feasible to build custom sample-centric microscopes that are augmented with trained inference modules similar to the here presented. Ultimately, these could use the steady stream of acquired experimental data as continuous source of additional training data, thus transforming fluorescence microscopy into a data-driven discipline.

Bibliography

Abbe, E. (1873). "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung". In: *Archiv für mikroskopische Anatomie* 9.1 (cit. on p. 3).

Abdellah, M., Bilgili, A., Eilemann, S., Markram, H., and Schürmann, F. (2015). "Physically-based in silico light sheet microscopy for visualizing fluorescent brain models". In: *BMC Bioinformatics* 16.11 (cit. on p. 41).

Afshar, Y. and Sbalzarini, I. F. (2016). "A parallel distributed-memory particle method enables acquisition-rate segmentation of large fluorescence microscopy images". In: *PLoS ONE* 11.4 (cit. on p. 13).

Ahrens, J., Geveci, B., and Law, C. (2005). "ParaView : An End-User Tool for Large Data Visualization". In: *The Visualization Handbook*. Ed. by Hansen, C. R. and Johnson, C. D. Elsevier (cit. on p. 13).

Aigouy, B., Farhadifar, R., Staple, D. B., Sagner, A., Röper, J.-C., Jülicher, F., and Eaton, S. (2010). "Cell flow reorients the axis of planar polarity in the wing epithelium of Drosophila". In: *Cell* 142.5 (cit. on p. 97).

Alglave, J., Batty, M., Donaldson, A. F., Gopalakrishnan, G., Ketema, J., Poetzl, D., Sorensen, T., and Wickerson, J. (2015). "GPU concurrency: Weak behaviours and programming assumptions". In: *ACM SIGARCH Computer Architecture News* 43.1 (cit. on p. 14).

Amat, F., Höckendorf, B., Wan, Y., Lemon, W. C., McDole, K., and Keller, P. J. (2015). "Efficient processing and analysis of large-scale light-sheet microscopy data". In: *Nature Protocols* 10.11 (cit. on p. 5).

Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E. W., Klessen, R. S., Maier-Hein, L., Rother, C., and Köthe, U. (2018). "Analyzing Inverse Problems with Invertible Neural Networks". In: *arXiv preprint arXiv:1808.04730* (cit. on p. 127).

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., and Sebastian Seung, H. (2017). "Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification". In: *Bioinformatics* (cit. on p. 66).

Atela, P., Golé, C., and Hotton, S. (2002). "A dynamical system for plant pattern formation: a rigorous analysis". In: *Journal of Nonlinear Science* 12.6 (cit. on p. 19).

Azucena, O., Crest, J., Cao, J., Sullivan, W., Kner, P., Gavel, D., Dillon, D., Olivier, S., and Kubby, J. (2010). "Wavefront aberration measurements and corrections through thick tissue using fluorescent microsphere reference beacons". In: *Optics express* 18.16 (cit. on p. 38).

Balazs, B., Deschamps, J., Albert, M., Ries, J., and Hufnagel, L. (2017). "A real-time compression library for microscopy images". In: *bioRxiv* (cit. on p. 13).

Beier, H. T. and Ibey, B. L. (2014). "Experimental comparison of the high-speed imaging performance of an EM-CCD and sCMOS camera in a dynamic live-cell imaging test case". In: *PloS ONE* 9.1 (cit. on p. 21).

Beier, T., Pape, C., Rahaman, N., Prange, T., Berg, S., Bock, D. D., Cardona, A., Knott, G. W., Plaza, S. M., Scheffer, L. K., et al. (2017). "Multicut brings automated neurite segmentation closer to human performance". In: *Nature Methods* 14.2 (cit. on p. 72).

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., and Hess, H. F. (2006). "Imaging intracellular fluorescent proteins at nanometer resolution". In: *Science* 313.5793 (cit. on pp. 2, 4).

Blasse, C., Saalfeld, S., Etournay, R., Sagner, A., Eaton, S., and Myers, E. W. (2017). "PreMosa: extracting 2D surfaces from 3D microscopy mosaics". In: *Bioinformatics* (cit. on pp. 2, 70, 97–100, 154).

Błaszczak, Z., Kreysing, M., and Guck, J. (2014). "Direct observation of light focusing by single photoreceptor cell nuclei". In: *Optics Express* 22.9 (cit. on p. 64).

Bluestein, L. (1970). "A linear filtering approach to the computation of discrete Fourier transform". In: *IEEE Transactions on Audio and Electroacoustics* 18.4 (cit. on p. 28).

Bohren, C. F. and Huffman, D. R. (2008). *Absorption and scattering of light by small particles*. John Wiley & Sons (cit. on pp. 41, 48, 53).

Booth, M. J. (2014). "Adaptive optical microscopy: the ongoing quest for a perfect image". In: *Light: Science & Applications* 3.4 (cit. on p. 39).

Booth, M. J., Neil, M. A., and Wilson, T. (1998). "Aberration correction for confocal imaging in refractive-index-mismatched media". In: *Journal of microscopy* 192.2 (cit. on p. 41).

Boothe, T., Hilbert, L., Heide, M., Berninger, L., Huttner, W. B., Zaburdaev, V., Vastenhouw, N. L., Myers, E. W., Drechsel, D. N., and Rink, J. C. (2017). "A tunable refractive index matching medium for live imaging cells, tissues and model organisms". In: *eLife* 6 (cit. on p. 89).

Born, M. and Wolf, E. (1999). *Principles of Optics*. 7th ed. Cambridge University Press (cit. on pp. 3, 42, 49, 156).

Boulanger, J., Kervrann, C., Bouthemy, P., Elbau, P., Sibarita, J.-B., and Salamero, J. (2010). "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences". In: *IEEE Transactions on Medical Imaging* 29.2 (cit. on p. 23).

Boyd, N., Jonas, E., Babcock, H. P., and Recht, B. (2018). "Deeploco: Fast 3D localization microscopy using neural networks". In: *BioRxiv* (cit. on p. 75).

Breiman, L. (2001). "Random forests". In: *Machine learning* 45.1 (cit. on p. 72).

Brenner, K.-H. and Singer, W. (1993). "Light propagation through microlenses: a new simulation method". In: *Applied optics* 32.26 (cit. on p. 42).

Brock, D. C. and Moore, G. E. (2006). *Understanding Moore's law: four decades of innovation*. Chemical Heritage Foundation (cit. on p. 13).

Brown, S. J., Shippy, T. D., Miller, S., Bolognesi, R., Beeman, R. W., Lorenzen, M. D., Bucher, G., Wimmer, E. A., and Klingler, M. (Aug. 2009). "The Red Flour Beetle, Tribolium castaneum (Coleoptera): A Model for Studies of Development and Pest Biology". In: *Cold Spring Harbor Protocols* 2009.8 (cit. on p. 93).

Bruce, M. A. and Butte, M. J. (2013). "Real-time GPU-based 3D deconvolution". In: *Optics Express* 21.4 (cit. on pp. 13, 27, 28).

Buades, A., Coll, B., and Morel, J.-M. (2005). "A non-local algorithm for image denoising". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2. IEEE (cit. on pp. 2, 23, 24, 70, 82, 83, 88, 90, 99, 100).

Burger, H. C., Schuler, C. J., and Harmeling, S. (2012). "Image denoising: Can plain neural networks compete with BM3D?" In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE (cit. on pp. 23, 74).

Caicedo, J. C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., Vasilevich, A. S., Barry, J. D., Bansal, H. S., Kraus, O., et al. (2017). "Data-analysis strategies for image-based cell profiling". In: *Nature Methods* 14.9 (cit. on p. 72).

Carlton, P. M., Boulanger, J., Kervrann, C., Sibarita, J.-B., Salamero, J., Gordon-Messer, S., Bressan, D., Haber, J. E., Haase, S., Shao, L., Winoto, L., Matsuda, A., Kner, P., Uzawa, S., Gustafsson, M., Kam, Z., Agard, D. A., and Sedat, J. W. (Sept. 2010). "Fast live simultaneous multiwavelength four-dimensional optical microscopy." In: *Proc Natl Acad Sci U S A* 107 (cit. on p. 23).

Carroll, J. D. and Chang, J.-J. (1970). "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition". In: *Psychometrika* 35.3 (cit. on p. 29).

Chambolle, A. (2004). "An algorithm for total variation minimization and applications". In: *Journal of Mathematical Imaging and Vision* 20.1-2 (cit. on pp. 90, 92).

Cheeseman, B. L., Günther, U., Susik, M., Gonciarz, K., and Sbalzarini, I. F. (2018). "Forget Pixels: Adaptive Particle Representation of Fluorescence Microscopy Images". In: *bioRxiv* (cit. on p. 13).

Chen, B.-C., Legant, W. R., Wang, K., Shao, L., Milkie, D. E., Davidson, M. W., Janetopoulos, C., Wu, X. S., Hammer 3rd, J. A., Liu, Z., English, B. P., Mimori-Kiyosue, Y., Romero, D. P., Ritter, A. T., Lippincott-Schwartz, J., Fritz-Laylin, L., Mullins, R. D., Mitchell, D. M., Bembenek, J. N., Reymann, A.-C., Böhme, R., Grill, S. W., Wang, J. T., Seydoux, G., Tulu, U. S., Kiehart, D. P., and Betzig, E. (Oct. 2014). "Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution." In: *Science* 346.6208 (cit. on pp. 5, 11, 16, 21, 47, 48).

Chhetri, R. K., Amat, F., Wan, Y., Höckendorf, B., Lemon, W. C., and Keller, P. J. (Dec. 2015). "Whole-animal functional and developmental imaging with isotropic spatial resolution." In: *Nature Methods* 12.12 (cit. on pp. 11, 101).

Chi, W., Wang, D., Wang, Y., Zhao, J., Rong, L., and Yuan, Y. (2018). "Measurement of 3D refractive index distribution by optical diffraction tomography". In: *2017*

International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology. Vol. 10620. International Society for Optics and Photonics (cit. on p. 67).

Choi, W., Fang-Yen, C., Badizadegan, K., Oh, S., Lue, N., Dasari, R. R., and Feld, M. S. (2007). "Tomographic phase microscopy". In: *Nature Methods* 4 (cit. on pp. 39, 40, 43, 58, 67).

Choromanska, A., Henaff, M., Mathieu, M., Arous, G. B., and LeCun, Y. (2015). "The loss surfaces of multilayer networks". In: *Artificial Intelligence and Statistics* (cit. on p. 78).

Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., Mount, E., O'Neil, A., Shah, K., Lee, A. K., et al. (2018). "In silico labeling: Predicting fluorescent labels in unlabeled images". In: *Cell* 173.3 (cit. on p. 72).

Chung, K. and Deisseroth, K. (2013). "CLARITY for mapping the nervous system". In: *Nature Methods* 10.6 (cit. on p. 38).

Cireşan, D., Meier, U., Masci, J., and Schmidhuber, J. (2012). "Multi-column deep neural network for traffic sign classification". In: *Neural networks* 32 (cit. on p. 72).

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). "Fast and accurate deep network learning by exponential linear units (elus)". In: *arXiv preprint arXiv:1511.07289* (cit. on p. 87).

Cooley, J. W. and Tukey, J. W. (1965). "An algorithm for the machine calculation of complex Fourier series". In: *Mathematics of computation* 19.90 (cit. on p. 28).

Cox, c. I. and Sheppard, C. (1986). "Information capacity and resolution in an optical system". In: *JOSA A* 3.8 (cit. on p. 70).

Culley, S., Albrecht, D., Jacobs, C., Pereira, P. M., Leterrier, C., Mercer, J., and Henriques, R. (2017). "Quantitative mapping and minimization of super-resolution optical imaging artifacts". In: *Nature Methods* (cit. on p. 118).

Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function". In: *Mathematics of control, signals and systems* 2.4 (cit. on pp. 71, 87).

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2009). "BM3D image denoising with shape-adaptive principal component analysis". In: *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations* (cit. on pp. 23, 70, 82, 83, 90, 92).

Dagum, L. and Menon, R. (1998). "OpenMP: an industry standard API for shared-memory programming". In: *IEEE Computational Science and Engineering* 5.1 (cit. on p. 13).

Darbon, J., Cunha, A., Chan, T. F., Osher, S., and Jensen, G. J. (2008). "Fast nonlocal filtering applied to electron cryomicroscopy". In: *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*. IEEE (cit. on p. 25).

De Medeiros, G., Norlin, N., Gunther, S., Albert, M., Panavaite, L., Fiuza, U.-M., Peri, F., Hiiragi, T., Krzic, U., and Hufnagel, L. (2015). "Confocal multiview light-sheet microscopy". In: *Nature Communications* 6 (cit. on p. 11).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Ieee (cit. on pp. 72, 74).

Dong, C., Loy, C. C., He, K., and Tang, X. (2014). "Learning a deep convolutional network for image super-resolution". In: *European Conference on Computer Vision*. Springer (cit. on pp. 103–105).

— (2016). "Image super-resolution using deep convolutional networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.2 (cit. on p. 95).

Drebin, R. A., Carpenter, L., and Hanrahan, P. (1988). "Volume rendering". In: *ACM Siggraph Computer Graphics*. Vol. 22. 4. ACM (cit. on p. 16).

Economo, M. N., Clack, N. G., Lavis, L. D., Gerfen, C. R., Svoboda, K., Myers, E. W., and Chandrashekar, J. (2016). "A platform for brain-wide imaging and reconstruction of individual neurons". In: *eLife* 5 (cit. on p. 101).

Eigen, D. and Fergus, R. (2015). "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture". In: *Proceedings of the IEEE International Conference on Computer Vision* (cit. on p. 78).

Ertürk, A., Becker, K., Jährling, N., Mauch, C. P., Hojer, C. D., Egen, J. G., Hellal, F., Bradke, F., Sheng, M., and Dodt, H.-U. (2012). "Three-dimensional imaging of solvent-cleared organs using 3DISCO". In: *Nature Protocols* 7.11 (cit. on p. 38).

Etournay, R., Merkel, M., Popović, M., Brandl, H., Dye, N. A., Aigouy, B., Salbreux, G., Eaton, S., and Jülicher, F. (2016). "TissueMiner: A multiscale analysis toolkit to quantify how cellular processes create tissue dynamics". In: *eLife* 5 (cit. on p. 101).

Etournay, R., Popović, M., Merkel, M., Nandi, A., Blasse, C., Aigouy, B., Brandl, H., Myers, G., Salbreux, G., Jülicher, F., et al. (2015). "Interplay of cell dynamics and epithelial tension during morphogenesis of the Drosophila pupal wing". In: *eLife* 4 (cit. on p. 97).

Fahrbach, F. O., Simon, P., and Rohrbach, A. (2010). "Microscopy with self-reconstructing beams". In: *Nature Photonics* 4.11 (cit. on pp. 38, 41).

Feit, M. and Fleck, J. (1978). "Light propagation in graded-index optical fibers". In: *Applied optics* 17.24 (cit. on p. 41).

Fertig, M. and Brenner, K.-H. (2010). "Vector wave propagation method". In: *J. Opt. Soc. Am.* 27.4 (cit. on p. 42).

Firasta, N., Buxton, M., Jinbo, P., Nasri, K., and Kuo, S. (2008). "Intel AVX: New frontiers in performance improvements and energy efficiency". In: *Intel white paper* 19 (cit. on p. 13).

Foreman, M. R. and Török, P. (2011). "Computational methods in vectorial imaging". In: *Journal of Modern Optics* 58.5-6 (cit. on pp. 47, 156).

Fujimoto, J. G., Pitris, C., Boppart, S. A., and Brezinski, M. E. (2000). "Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy". In: *Neoplasia (New York, NY)* 2.1-2 (cit. on p. 52).

Fukushima, K. and Miyake, S. (1982). "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition". In: *Competition and cooperation in neural nets*. Springer (cit. on p. 71).

Gal, Y. and Ghahramani, Z. (2016). "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning". In: *International Conference on Machine Learning* (cit. on p. 127).

Gao, L., Shao, L., Chen, B.-C., and Betzig, E. (May 2014). "3D live fluorescence imaging of cellular dynamics using Bessel beam plane illumination microscopy." In: *Nature Protocols* 9.5 (cit. on p. 38).

Glaser, A. K., Chen, Y., and Liu, J. T. (2016). "Fractal propagation method enables realistic optical microscopy simulations in biological tissues". In: *Optica* 3.8 (cit. on p. 40).

Gong, Y. and Sbalzarini, I. F. (2017). "Curvature filters efficiently reduce certain variational energies". In: *IEEE Transactions on Image Processing* 26.4 (cit. on p. 23).

Gonsalves, R. A. (1982). "Phase retrieval and diversity in adaptive optics". In: *Optical Engineering* 21.5 (cit. on p. 126).

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*. Vol. 1. MIT press Cambridge (cit. on pp. 76–78).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems* (cit. on pp. 75, 122).

Goodman, J. (1996). *Introduction to Fourier Optics*. 2nd ed. MaGraw-Hill (cit. on pp. 3, 42).

Goossens, B., Luong, H., Aelterman, J., Pižurica, A., and Philips, W. (2010). "A GPU-accelerated real-time NLMeans algorithm for denoising color video sequences". In: *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer (cit. on p. 13).

Gouesbet, G., Lock, J., and Gréhan, G. (2011). "Generalized Lorenz–Mie theories and description of electromagnetic arbitrary shaped beams: localized approximations and localized beam models, a review". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 112.1 (cit. on p. 53).

Gräf, R., Rietdorf, J., and Zimmermann, T. (2005). "Live cell spinning disk microscopy". In: *Microscopy techniques*. Springer (cit. on p. 4).

Gropp, W., Lusk, E., Doss, N., and Skjellum, A. (1996). "A high-performance, portable implementation of the MPI message passing interface standard". In: *Parallel computing* 22.6 (cit. on p. 13).

Grotjohann, T., Testa, I., Reuss, M., Brakemann, T., Eggeling, C., Hell, S. W., and Jakobs, S. (2012). "rsEGFP2 enables fast RESOLFT nanoscopy of living cells". In: *Elife* 1 (cit. on p. 4).

Guirao, B., Rigaud, S. U., Bosveld, F., Bailles, A., Lopez-Gay, J., Ishihara, S., Sugimura, K., Graner, F., and Bellaiche, Y. (2015). "Unified quantitative characterization of epithelial tissue development". In: *eLife* 4 (cit. on p. 97).

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). “On Calibration of Modern Neural Networks”. In: *34th International Conference on Machine Learning (ICML)* (cit. on p. 127).

Gustafsson, M. G. (2000). “Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy”. In: *Journal of Microscopy* 198.2 (cit. on pp. 2, 4, 70).

Gustafsson, N., Culley, S., Ashdown, G., Owen, D. M., Pereira, P. M., and Henriques, R. (2016). “Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations”. In: *Nature Communications* 7 (cit. on pp. 70, 118, 155).

Hadley, G. R. (1992). “Transparent boundary condition for the beam propagation method”. In: *IEEE Journal of Quantum Electronics* 28.1 (cit. on p. 52).

Harshman, R. A. (1970). “Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis”. In: (cit. on p. 29).

Hart, M. (2010). “Recent advances in astronomical adaptive optics”. In: *Applied optics* 49.16 (cit. on p. 67).

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR (cit. on p. 77).

He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep Residual learning for image recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (cit. on pp. 72, 78).

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Scholkopf, B. (1998). “Support vector machines”. In: *IEEE Intelligent Systems and their applications* 13.4 (cit. on p. 72).

Heinrich, L., Bogovic, J. A., and Saalfeld, S. (2017). “Deep Learning for Isotropic Super-Resolution from Non-isotropic 3D Electron Microscopy”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention* (cit. on pp. 75, 95).

Heintzmann, R. and Cremer, C. G. (1999). “Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating”. In: *Optical Biopsies and Microscopic Techniques III*. Vol. 3568. International Society for Optics and Photonics (cit. on pp. 2, 4).

Hell, S. W., Lindek, S., Cremer, C., and Stelzer, E. H. (1994). “Confocal microscopy with an increased detection aperture: type-B 4Pi confocal microscopy”. In: *Optics letters* 19.3 (cit. on pp. 4, 101).

Hell, S. and Stelzer, E. H. (1992). “Properties of a 4Pi confocal fluorescence microscope”. In: *JOSA A* 9.12 (cit. on p. 4).

Helmchen, F. and Denk, W. (2005). “Deep tissue two-photon microscopy”. In: *Nature methods* 2.12 (cit. on p. 39).

Hendrycks, D. and Gimpel, K. (2016). “A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks”. In: *CoRR* abs/1610.02136. eprint: 1610.02136 (cit. on p. 127).

Hergert, W. and Wriedt, T. (2012). *The Mie Theory: basics and applications*. Vol. 169. Springer (cit. on p. 53).

Hochreiter, S. (1991). "Untersuchungen zu dynamischen neuronalen Netzen". In: *Diploma, Technische Universität München* 91.1 (cit. on p. 72).

Hornik, K., Stinchcombe, M., and White, H. (1989). "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5 (cit. on p. 71).

Huang, W. and Xu, C. (1993). "Simulation of three-dimensional optical waveguides by a full-vector beam propagation method". In: *IEEE journal of quantum electronics* 29.10 (cit. on p. 41).

Huang, Z.-L., Zhu, H., Long, F., Ma, H., Qin, L., Liu, Y., Ding, J., Zhang, Z., Luo, Q., and Zeng, S. (2011). "Localization-based super-resolution microscopy with an sCMOS camera". In: *Optics Express* 19.20 (cit. on p. 21).

Hubel, D. H. and Wiesel, T. N. (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of physiology* 160.1 (cit. on p. 71).

Huisken, J. and Stainier, D. Y. R. (Sept. 2007). "Even fluorescence excitation by multidirectional selective plane illumination microscopy (mSPIM)." In: *Opt Lett* 32.17 (cit. on p. 37).

Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J., and Stelzer, E. H. (2004). "Optical sectioning deep inside live embryos by selective plane illumination microscopy". In: *Science* 305.5686 (cit. on pp. 5, 47, 48, 51).

Icha, J., Kunath, C., Rocha-Martins, M., and Norden, C. (2016). "Independent modes of ganglion cell translocation ensure correct lamination of the zebrafish retina". In: *Journal of Cell Biology* 215.2 (cit. on p. 111).

Icha, J., Weber, M., Waters, J. C., and Norden, C. (2017). "Phototoxicity in live fluorescence microscopy, and how to avoid it". In: *Bioessays* 39.8 (cit. on pp. 7, 21, 70).

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (July 2017). "Image-To-Image Translation With Conditional Adversarial Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (cit. on p. 75).

Jacques, S. L. (2013). "Optical properties of biological tissues: a review". In: *Physics in Medicine and Biology* 58.11 (cit. on p. 40).

Jain, V. and Seung, S. (2008). "Natural Image Denoising with Convolutional Networks". In: *Advances in Neural Information Processing Systems 21*. Ed. by Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L. (cit. on pp. 23, 74).

Ji, N. (2017). "Adaptive optical fluorescence microscopy". In: *Nature Methods* 14.4 (cit. on pp. 38, 39).

Jin, K. H., McCann, M. T., Froustey, E., and Unser, M. (2017). "Deep Convolutional Neural Network for Inverse Problems in Imaging". In: *IEEE Transactions on Image Processing* 26.9 (cit. on p. 75).

Jog, A., Carass, A., and Prince, J. L. (2016). "Self Super-Resolution for Magnetic Resonance Images". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (cit. on p. 103).

Johnsen, S. and Widder, E. A. (1999). "The physical basis of transparency in biological tissue: ultrastructure and the minimization of light scattering". In: *Journal of Theoretical Biology* 199.2 (cit. on pp. 43, 52).

Judkewitz, B., Horstmeyer, R., Vellekoop, I. M., Papadopoulos, I. N., and Yang, C. (2015). "Translation correlations in anisotropically scattering media". In: *Nature Physics* 11.8 (cit. on pp. 39, 40, 61, 62).

Kamilov, U., Papadopoulos, I., Shoreh, M., Goy, A., Vonesch, C., Unser, M., and Psaltis, D. (2015). "Learning Approach to Optical Tomography". In: *Optica* 2.6 (cit. on p. 126).

Karas, P. and Svoboda, D. (2011). "Convolution of large 3D images on GPU and its decomposition". In: *EURASIP Journal on Advances in Signal Processing* 2011.1 (cit. on p. 27).

Kawano, K. and Kitoh, T. (2001). *Introduction to optical waveguide analysis*. New York: John Wiley & Sons (cit. on p. 41).

Keller, P. J. and Stelzer, E. H. K. (Dec. 2008a). "Quantitative in vivo imaging of entire embryos with Digital Scanned Laser Light Sheet Fluorescence Microscopy." In: *Curr Opin Neurobiol* 18.6 (cit. on p. 47).

Keller, P. J., Schmidt, A. D., Wittbrodt, J., and Stelzer, E. H. K. (Nov. 2008b). "Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy." In: *Science* 322 (5904) (cit. on p. 5).

Kim, J., Kwon Lee, J., and Mu Lee, K. (2016). "Accurate Image Super-Resolution Using Very Deep Convolutional Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (cit. on p. 95).

Kim, K., Park, W. S., Na, S., Kim, S., Kim, T., Do Heo, W., and Park, Y. (2017). "Correlative three-dimensional fluorescence and refractive index tomography: bridging the gap between molecular specificity and quantitative bioimaging". In: *Biomedical Optics Express* 8.12 (cit. on p. 67).

Kingma, D. and Ba, J. (2015). "Adam: A Method for Stochastic Optimization". In: *International Conference on Learning Representations (ICLR)* (cit. on pp. 79, 82).

Kner, P. (Oct. 2013). "Phase diversity for three-dimensional imaging." In: *J Opt Soc Am A Opt Image Sci Vis* 30.10 (cit. on p. 126).

Kolda, T. G. and Bader, B. W. (2009). "Tensor decompositions and applications". In: *SIAM review* 51.3 (cit. on p. 29).

Kreysing, M., Boyde, L., Guck, J., and Chalut, K. J. (2010). "Physical insight into light scattering by photoreceptor cell nuclei". In: *Optics Letters* 35.15 (cit. on pp. 39, 63, 64).

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems* (cit. on pp. 72, 82).

Krotkov, E. (1988). "Focusing". In: *International Journal of Computer Vision* 1.3 (cit. on p. 33).

Kubby, J. A. (2013). *Adaptive Optics for Biological Imaging*. CRC press (cit. on p. 39).

Kuipers, L. and Niederreiter, H. (2012). *Uniform Distribution of Sequences*. Dover Books on Mathematics. Dover Publications (cit. on p. 18).

Kumar, A., Wu, Y., Christensen, R., Chandris, P., Gandler, W., McCreedy, E., Bokinsky, A., Colón-Ramos, D. A., Bao, Z., McAuliffe, M., Rondeau, G., and Shroff, H. (Nov. 2014). "Dual-view plane illumination microscopy for rapid and spatially isotropic imaging." In: *Nature Protocols* 9.11 (cit. on p. 5).

Kwan, K. M., Otsuna, H., Kidokoro, H., Carney, K. R., Saijoh, Y., and Chien, C.-B. (2012). "A complex choreography of cell movements shapes the vertebrate eye". In: *Development* 139.2 (cit. on p. 111).

Laissue, P. P., Alghamdi, R. A., Tomancak, P., Reynaud, E. G., and Shroff, H. (2017). "Assessing phototoxicity in live fluorescence imaging". In: *Nature Methods* 14.7 (cit. on pp. 7, 70).

Lakshminarayanan, B., Pritzel, A., and Blundell, C. (2017). "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles". In: *Advances in Neural Information Processing Systems 30 (NIPS)* (cit. on p. 127).

LeCun, Y. et al. (1989). "Generalization and network design strategies". In: *Connectionism in perspective* (cit. on p. 87).

LeCun, Y., Bengio, Y., and Hinton, G. (2015). "Deep learning". In: *Nature* 521.7553 (cit. on p. 70).

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., and Jackel, L. D. (1990). "Handwritten digit recognition with a back-propagation network". In: *Advances in Neural Information Processing Systems* (cit. on p. 71).

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (cit. on pp. 70, 77).

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T. (2018). "Noise2Noise: Learning Image Restoration without Clean Data". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018* (cit. on pp. 75, 127).

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. (1959). "What the frog's eye tells the frog's brain". In: *Proceedings of the IRE* 47.11 (cit. on p. 71).

Li, K., Wu, X., Chen, D. Z., and Sonka, M. (2006). "Optimal Surface Segmentation in Volumetric Images – A Graph-Theoretic Approach". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.1 (cit. on pp. 99, 100).

Li, L., Hou, W., Zhang, X., and Ding, M. (2013). "GPU-Based Block-Wise Nonlocal Means Denoising for 3D Ultrasound Images". In: *Computational and Mathematical Methods in Medicine* 2013 (cit. on p. 25).

Linnainmaa, S. (1970). "The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors". In: *Master's Thesis (in Finnish), Univ. Helsinki* (cit. on p. 79).

Liu, T.-l., Upadhyayula, S., Milkie, D. E., Singh, V., Wang, K., Swinburne, I. A., Mosaliganti, K. R., Collins, Z. M., Hiscock, T. W., Shea, J., et al. (2018). "Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms". In: *Science* 360.6386 (cit. on pp. 5, 38, 39).

Ljosa, V., Sokolnicki, K. L., and Carpenter, A. E. (2012). "Annotated high-throughput microscopy image sets for validation". In: *Nature Methods* 9.7 (cit. on pp. 23, 81).

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully Convolutional Networks for Semantic Segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (cit. on p. 74).

Lorensen, W. E. and Cline, H. E. (1987). "Marching cubes: A high resolution 3D surface construction algorithm". In: *ACM siggraph computer graphics*. Vol. 21. 4. ACM (cit. on pp. 16, 20).

Loupas, T., McDicken, W., and Allan, P. L. (1989). "An adaptive weighted median filter for speckle suppression in medical ultrasonic images". In: *IEEE transactions on Circuits and Systems* 36.1 (cit. on p. 23).

Lucy, L. B. (1974). "An iterative technique for the rectification of observed distributions". In: *The Astronomical Journal* 79 (cit. on pp. 27, 30, 101, 104, 105).

MacKay, R., Meiss, J., and Percival, I. (1984). "Transport in Hamiltonian systems". In: *Physica D: Nonlinear Phenomena* 13.1-2 (cit. on p. 19).

Maggioni, M., Katkovnik, V., Egiazarian, K., and Foi, A. (2013). "Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction". In: *IEEE Transactions on Image Processing* 22.1 (cit. on pp. 23, 90, 92).

Mäkitalo, M. and Foi, A. (2012). "Poisson-gaussian denoising using the exact unbiased inverse of the generalized anscombe transformation". In: *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE (cit. on p. 21).

Manley, S., Gunzenhäuser, J., and Olivier, N. (2011). "A starter kit for point-localization super-resolution imaging". In: *Current opinion in chemical biology* 15.6 (cit. on p. 117).

Mao, X., Shen, C., and Yang, Y.-B. (2016). "Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections". In: *Advances in Neural Information Processing Systems* (cit. on pp. 23, 74, 95).

Márques, A. and Pardo, A. (2013). "Implementation of non local means filter in gpus". In: *Iberoamerican Congress on Pattern Recognition*. Springer (cit. on p. 13).

Martin, D., Fowlkes, C., Tal, D., and Malik, J. (July 2001). "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics". In: *Proc. 8th Int'l Conf. Computer Vision*. Vol. 2 (cit. on pp. 72, 74).

Matejčić, M., Salbreux, G., and Norden, C. (2018). "A non-cell-autonomous actin redistribution enables isotropic retinal growth". In: *PLoS biology* 16.8 (cit. on p. 111).

Mätzler, C. (2002). "MATLAB functions for Mie scattering and absorption, version 2". In: *IAP Res. Rep* 8.1 (cit. on p. 53).

Mazziotta, J. C. and Huang, H. (1976). "THREAD (three-dimensional reconstruction and display) with biomedical applications in neuron ultrastructure and computerized tomography". In: *Proceedings of the June 7-10, 1976, national computer conference and exposition*. ACM (cit. on p. 16).

McConnell, G., Trägårdh, J., Amor, R., Dempster, J., Reid, E., and Amos, W. B. (2016). "A novel optical microscope for imaging large embryos and tissue volumes with sub-cellular resolution throughout". In: *Elife* 5 (cit. on p. 11).

McCulloch, W. S. and Pitts, W. (1943). "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4 (cit. on p. 71).

Miller, G. (1994). "Efficient algorithms for local and global accessibility shading". In: *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. ACM (cit. on p. 20).

Minsky, M. (1961). *Microscopy apparatus*. US Patent 3,013,467 (cit. on p. 4).

Minsky, M. and Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA: MIT Press (cit. on p. 71).

Moore, G. E. (1965). *Cramming more components onto integrated circuits* (cit. on p. 13).

Morales-Navarrete, H., Segovia-Miranda, F., Klukowski, P., Meyer, K., Nonaka, H., Marsico, G., Chernykh, M., Kalaidzidis, A., Zerial, M., and Kalaidzidis, Y. (2015). "A versatile pipeline for the multi-scale digital reconstruction and quantitative analysis of 3D tissue architecture". In: *eLife* 4 (cit. on pp. 2, 101, 106, 107).

Mudry, E., Belkebir, K., Girard, J., Savatier, J., Le Moal, E., Nicoletti, C., Allain, M., and Sentenac, A. (2012). "Structured illumination microscopy using unknown speckle patterns". In: *Nature Photonics* 6.5 (cit. on p. 126).

Müller, M., Mönkemöller, V., Hennig, S., Hübner, W., and Huser, T. (2016). "Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ". In: *Nature Communications* 7 (cit. on p. 70).

Nagy, J. G. and O'Leary, D. P. (1998). "Restoring Images Degraded by Spatially Variant Blur". In: *SIAM Journal on Scientific Computing* 19.4 (cit. on p. 52).

Nair, V. and Hinton, G. E. (2010). "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th international conference on machine learning (ICML-10)* (cit. on p. 87).

Nasrabadi, N. M. (2007). "Pattern recognition and machine learning". In: *Journal of electronic imaging* 16.4 (cit. on p. 72).

Negash, A., Labouesse, S., Sandeau, N., Allain, M., Giovannini, H., Idier, J., Heintzmann, R., Chaumet, P. C., Belkebir, K., and Sentenac, A. (2016). "Improving the axial and lateral resolution of three-dimensional fluorescence microscopy using random speckle illuminations". In: *JOSA A* 33.6 (cit. on p. 126).

Nehme, E., Weiss, L. E., Michaeli, T., and Shechtman, Y. (2018). "Deep-STORM: super-resolution single-molecule microscopy by deep learning". In: *Optica* 5.4 (cit. on pp. 72, 75, 115).

Nickolls, J., Buck, I., Garland, M., and Skadron, K. (2008). "Scalable parallel programming with CUDA". In: *ACM SIGGRAPH 2008 classes*. ACM (cit. on p. 14).

O'shea, D. C. and C'Shea, D. C. (1985). *Elements of modern optical design*. Vol. 51. Wiley New York (cit. on p. 41).

Oktay, O., Bai, W., Lee, M., Guerrero, R., Kamnitsas, K., Caballero, J., Marvao, A. de, Cook, S., O'Regan, D., and Rueckert, D. (2016). "Multi-input Cardiac Image Super-Resolution Using Convolutional Neural Networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (cit. on p. 75).

Oskooi, A. F., Roundy, D., Ibanescu, M., Bermel, P., Joannopoulos, J. D., and Johnson, S. G. (Mar. 2010). "Meep: A flexible free-software package for electromagnetic simulations by the FDTD method". In: *Computer Physics Communications* 181.3 (cit. on pp. 40, 42, 45).

Osnabrugge, G., Leedumrongwatthanakun, S., and Vellekoop, I. M. (2016). "A convergent Born series for solving the inhomogeneous Helmholtz equation in arbitrarily large media". In: *Journal of Computational Physics* 322 (cit. on pp. 40, 42, 67, 126).

Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F., and Johnson, G. (2018). "Label-free prediction of three-dimensional fluorescence images from transmitted light microscopy". In: *bioRxiv* (cit. on p. 72).

Ouyang, W., Aristov, A., Lelek, M., Hao, X., and Zimmer, C. (2018). "Deep learning massively accelerates super-resolution localization microscopy". In: *Nature biotechnology* (cit. on pp. 72, 75, 115, 122).

Paik, J. K. and Katsaggelos, A. K. (1992). "Image restoration using a modified Hopfield network". In: *IEEE Transactions on Image Processing* 1.1 (cit. on p. 74).

Pawley, J. (2010). *Handbook of biological confocal microscopy*. Springer Science & Business Media (cit. on p. 4).

Pawley, J. B. (2006). "Fundamental limits in confocal microscopy". In: *Handbook of Biological Confocal Microscopy*. Springer (cit. on pp. 7, 70).

Pelt, D. M. and Sethian, J. A. (2017). "A mixed-scale dense convolutional neural network for image analysis". In: *Proceedings of the National Academy of Sciences* (cit. on p. 85).

Pelt, D. M. and Batenburg, K. J. (2013). "Fast tomographic reconstruction from limited data using artificial neural networks". In: *IEEE Transactions on Image Processing* 22.12 (cit. on p. 75).

Perlin, K. (1985). "An Image Synthesizer". In: *Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '85. New York, NY, USA: ACM (cit. on p. 59).

Perona, P. and Malik, J. (1990). "Scale-space and edge detection using anisotropic diffusion". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.7 (cit. on p. 23).

Petráň, M., Hadravský, M., Egger, M. D., and Galambos, R. (1968). "Tandem-scanning reflected-light microscope". In: *JOSA* 58.5 (cit. on p. 4).

Pharr, M. and Fernando, R. (2005). *GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation*. Addison-Wesley Professional (cit. on p. 14).

Pharr, M., Jakob, W., and Humphreys, G. (2016). *Physically based rendering: From theory to implementation*. Morgan Kaufmann (cit. on p. 18).

Phong, B. T. (1975). "Illumination for computer generated pictures". In: *Communications of the ACM* 18.6 (cit. on p. 20).

Pietzsch, T., Saalfeld, S., Preibisch, S., and Tomancak, P. (2015). "BigDataViewer: visualization and processing for large image data sets". In: *Nature Methods* 12.6 (cit. on p. 13).

Pitrone, P. G., Schindelin, J., Stuyvenberg, L., Preibisch, S., Weber, M., Eliceiri, K. W., Huisken, J., and Tomancak, P. (2013). "OpenSPIM: an open-access light-sheet microscopy platform". In: *Nature Methods* 10.7 (cit. on p. 5).

Planchon, T. A., Gao, L., Milkie, D. E., Davidson, M. W., Galbraith, J. A., Galbraith, C. G., and Betzig, E. (May 2011). "Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination." In: *Nature Methods* 8.5 (cit. on pp. 38, 47).

Power, R. M. and Huisken, J. (2017). "A guide to light-sheet fluorescence microscopy for multiscale imaging". In: *Nature Methods* 14.4 (cit. on p. 5).

Preibisch, S., Amat, F., Stamataki, E., Sarov, M., Singer, R. H., Myers, E., and Tomancak, P. (2014). "Efficient Bayesian-based multiview deconvolution." In: *Nature Methods* 11 (6) (cit. on pp. 2, 5, 27, 28, 31, 70, 101, 109).

Rajaram, S., Pavie, B., Hac, N. E., Altschuler, S. J., and Wu, L. F. (2012). "SimuCell: a flexible framework for creating synthetic microscopy images". In: *Nature Methods* 9.7 (cit. on p. 115).

Richardson, W. H. (1972). "Bayesian-based iterative method of image restoration". In: *Journal of the Optical Society of America* 62.1 (cit. on pp. 2, 27, 30, 70, 101, 104, 105).

Rigamonti, R., Sironi, A., Lepetit, V., and Fua, P. (2013). "Learning separable filters". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE (cit. on p. 29).

Ringler, M. (2008). "Plasmonische Nahfeldresonatoren aus zwei biokonjugierten Goldnanopartikeln". PhD thesis. LMU (cit. on pp. 41, 45, 53).

Rivenson, Y., Göröcs, Z., Günaydin, H., Zhang, Y., Wang, H., and Ozcan, A. (2017). "Deep learning microscopy". In: *Optica* 4.11 (cit. on pp. 72, 75).

Robbins, M. S. and Hadwen, B. J. (2003). "The noise performance of electron multiplying charge-coupled devices". In: *IEEE transactions on Electron Devices* 50.5 (cit. on p. 21).

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (cit. on pp. 74, 78, 82, 85, 102, 160).

Rosenblatt, F. (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (cit. on p. 71).

Roth, S. D. (1982). "Ray casting for modeling solids". In: *Computer Graphics and Image Processing* 18.2 (cit. on p. 16).

Royer, L. A., Lemon, W. C., Chhetri, R. K., Wan, Y., Coleman, M., Myers, E. W., and Keller, P. J. (2016). "Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms". In: *Nature Biotechnology* 34.12 (cit. on pp. 5, 6, 11, 12, 19, 21, 31, 33, 37, 109, 110, 154).

Rudin, L. I., Osher, S., and Fatemi, E. (1992). "Nonlinear total variation based noise removal algorithms". In: *Physica D: Nonlinear Phenomena* 60.1-4 (cit. on pp. 23, 82, 83).

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). "Learning representations by back-propagating errors". In: *nature* 323.6088 (cit. on p. 79).

Rust, M. J., Bates, M., and Zhuang, X. (2006). "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)". In: *Nature Methods* 3.10 (cit. on p. 4).

Sage, D., Kirshner, H., Pengo, T., Stuurman, N., Min, J., Manley, S., and Unser, M. (2015). "Quantitative evaluation of software packages for single-molecule localization microscopy". In: *Nature methods* 12.8 (cit. on p. 117).

Sarrazin, A. F., Peel, A. D., and Averof, M. (2012). "A segmentation clock with two-segment periodicity in insects". In: *Science* 336.6079 (cit. on p. 93).

Sbalzarini, I. F. (2016). "Seeing Is believing: quantifying is convincing: computational image analysis in biology". In: *Focus on Bio-Image Informatics*. Springer (cit. on pp. 23, 27).

Schawinski, K., Zhang, C., Zhang, H., Fowler, L., and Santhanam, G. K. (2017). "Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit". In: *Monthly Notices of the Royal Astronomical Society: Letters* 467.1 (cit. on p. 75).

Scherf, N. and Huisken, J. (2015). "The smart and gentle microscope". In: *Nature Biotechnology* 33.8 (cit. on pp. 6, 7, 12).

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (July 2012). "Fiji: an open-source platform for biological-image analysis." In: *Nature Methods* 9.7 (cit. on pp. 25, 66).

Schmid, B. and Huisken, J. (2015). "Real-time multi-view deconvolution". In: *Bioinformatics* 31.20 (cit. on pp. 13, 27, 28, 31).

Schmid, B., Schindelin, J., Cardona, A., Longair, M., and Heisenberg, M. (2010). "A high-level 3D visualization API for Java and ImageJ". In: *BMC Bioinformatics* 11.1 (cit. on pp. 13, 16).

Schmidt, S., Tiess, T., Schröter, S., Hambach, R., Jäger, M., Bartelt, H., Tünnermann, A., and Gross, H. (2016). "Wave-optical modeling beyond the thin-element-approximation". In: *Optics Express* 24.26 (cit. on pp. 42, 52, 67, 126).

Schmidt, U. and Roth, S. (2014). "Shrinkage fields for effective image restoration". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (cit. on pp. 23, 72, 92).

Schürmann, M., Cojoc, G., Girardo, S., Ulbricht, E., Guck, J., and Müller, P. (2017). "Three-dimensional correlative single-cell imaging utilizing fluorescence and refractive index tomography". In: *Journal of Biophotonics* (cit. on p. 67).

Schwertner, M., Booth, M., and Wilson, T. (2004). "Simulation of specimen-induced aberrations for objects with spherical and cylindrical symmetry". In: *Journal of microscopy* 215.3 (cit. on p. 41).

Shettigar, N., Joshi, A., Dalmeida, R., Gopalkrishna, R., Chakravarthy, A., Patnaik, S., Mathew, M., Palakodeti, D., and Gulyani, A. (2017). "Hierarchies in light sensing and dynamic interactions between ocular and extraocular sensory networks in a flatworm". In: *Science Advances* 3.7 (cit. on p. 88).

Shihavuddin, A., Basu, S., Rexhepaj, E., Delestro, F., Menezes, N., Sigoillot, S. M., Del Nery, E., Selimi, F., Spassky, N., and Genovesio, A. (2017). "Smooth 2D manifold extraction from 3D image stack". In: *Nature Communications* 8 (cit. on pp. 2, 70, 97, 99, 100).

Shreiner, D. (2009). *OpenGL programming guide: the official guide to learning OpenGL, versions 3.0 and 3.1*. Pearson Education (cit. on p. 17).

Sidhaye, J. and Norden, C. (2017). "Concerted action of neuroepithelial basal shrinkage and active epithelial migration ensures efficient optic cup morphogenesis". In: *eLife* 6 (cit. on p. 111).

Siedentopf, H. and Zsigmondy, R. (1902). "Uber sichtbarmachung und größenbestimmung ultramikoskopischer teilchen, mit besonderer anwendung auf goldrubingläser". In: *Annalen der Physik* 315.1 (cit. on p. 5).

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). "Mastering the game of Go with deep neural networks and tree search". In: *nature* 529.7587 (cit. on p. 72).

Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., Guck, J., and Joffe, B. (2009). "Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution". In: *Cell* 137.2 (cit. on pp. 39, 63, 64).

Starck, J.-L., Murtagh, F. D., and Bijaoui, A. (1998). *Image processing and data analysis: the multiscale approach*. Cambridge University Press (cit. on p. 21).

Stegmaier, J., Amat, F., Lemon, W. C., McDole, K., Wan, Y., Teodoro, G., Mikut, R., and Keller, P. J. (2016). "Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos". In: *Developmental cell* 36.2 (cit. on p. 13).

Steinbach, P. and Werner, M. (2017). "gearshifft–The FFT Benchmark Suite for Heterogeneous Platforms". In: *International Supercomputing Conference*. Springer (cit. on p. 28).

Stone, J. E., Gohara, D., and Shi, G. (2010). "OpenCL: A parallel programming standard for heterogeneous computing systems". In: *Computing in science & engineering* 12.3 (cit. on p. 14).

Sutter, H. (2005). "The free lunch is over: A fundamental turn toward concurrency in software". In: *Dr. Dobb's journal* 30.3 (cit. on p. 13).

Svoboda, D. and Ulman, V. (2017). "Mitogen: A framework for generating 3D synthetic time-lapse sequences of cell populations in fluorescence microscopy". In: *IEEE Transactions on Medical Imaging* 36.1 (cit. on p. 115).

Tao, X., Crest, J., Kotadia, S., Azucena, O., Chen, D. C., Sullivan, W., and Kubby, J. (2012). "Live imaging using adaptive optics with fluorescent protein guide-stars". In: *Optics express* 20.14 (cit. on p. 61).

Tomasi, C. and Manduchi, R. (1998). "Bilateral filtering for gray and color images". In: *Computer Vision, 1998. Sixth International Conference on*. IEEE (cit. on p. 23).

Tomer, R., Khairy, K., Amat, F., and Keller, P. J. (July 2012). "Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy." In: *Nature Methods* 9.7 (cit. on pp. 5, 11, 16).

Truong, T. V., Supatto, W., Koos, D. S., Choi, J. M., and Fraser, S. E. (Sept. 2011). "Deep and fast live imaging with two-photon scanned light-sheet microscopy." In: *Nature Methods* 8.9 (cit. on p. 39).

Uddin, M. S., Lee, H. K., Preibisch, S., and Tomancak, P. (Aug. 2011). "Restoration of uneven illumination in light sheet microscopy images." In: *Microsc Microanal* 17.4 (cit. on pp. 40, 41).

Ulyanov, D., Vedaldi, A., and Lempitsky, V. (June 2018). "Deep Image Prior". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (cit. on p. 75).

Van Roey, J., Van der Donk, J., and Lagasse, P. (1981). "Beam-propagation method: analysis and assessment". In: *J. Opt. Soc. Am.* 71.7 (cit. on p. 41).

Vellekoop, I. M. and Aegerter, C. M. (2010). "Scattered light fluorescence microscopy: imaging through turbid layers". In: *Optics letters* 35.8 (cit. on p. 39).

Vellekoop, I. M. and Mosk, A. (2007). "Focusing coherent light through opaque strongly scattering media". In: *Optics Letters* 32.16 (cit. on pp. 39, 61).

Venkataramani, V., Herrmannsdörfer, F., Heilemann, M., and Kuner, T. (2016). "SuReSim: simulating localization microscopy experiments from ground truth models". In: *Nature Methods* 13.4 (cit. on pp. 40, 115).

Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A. (2008). "Extracting and composing robust features with denoising autoencoders". In: *Proceedings of the 25th international conference on Machine learning*. ACM (cit. on p. 77).

Voie, A. H., Burns, D., and Spelman, F. (1993). "Orthogonal-plane fluorescence optical sectioning: three-dimensional imaging of macroscopic biological specimens". In: *Journal of microscopy* 170.3 (cit. on p. 5).

Walt, S. van der, Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., and contributors, the scikit-image (June 2014). "scikit-image: image processing in Python". In: *PeerJ* 2 (cit. on pp. 25, 66).

Wang, K., Milkie, D. E., Saxena, A., Engerer, P., Misgeld, T., Bronner, M. E., Mumm, J., and Betzig, E. (Apr. 2014). "Rapid adaptive optical recovery of optimal resolution over large volumes." In: *Nature Methods* 11.6 (cit. on pp. 5, 39).

Wang, K., Sun, W., Richie, C. T., Harvey, B. K., Betzig, E., and Ji, N. (2015). "Direct wavefront sensing for high-resolution in vivo imaging in scattering tissue". In: *Nature Communications* 6 (cit. on p. 39).

Wang, S., Su, Z., Ying, L., Peng, X., Zhu, S., Liang, F., Feng, D., and Liang, D. (2016). "Accelerating magnetic resonance imaging via deep learning". In: *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE (cit. on p. 75).

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). "Image quality assessment: from error visibility to structural similarity". In: *IEEE Transactions on Image Processing* 13.4 (cit. on pp. 24, 80, 83).

Westphal, V. and Hell, S. W. (2005). "Nanoscale resolution in the focal plane of an optical microscope". In: *Physical review letters* 94.14 (cit. on p. 4).

Wolpert, D. H. (1996). "The lack of a priori distinctions between learning algorithms". In: *Neural computation* 8.7 (cit. on p. 127).

Wong, T.-T., Luk, W.-S., and Heng, P.-A. (1997). "Sampling with Hammersley and Halton points". In: *Journal of graphics tools* 2.2 (cit. on p. 18).

Wu, X. and Chen, D. Z. (2002). "Optimal net surface problems with applications". In: *International Colloquium on Automata, Languages, and Programming*. Springer (cit. on pp. 99, 100).

Yee, K. (1966). "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media". In: *IEEE Transactions on antennas and propagation* 14.3 (cit. on p. 42).

Young, L. K., Jarrin, M., Saunter, C. D., Quinlan, R. A., and Girkin, J. M. (2018). "Non-invasive in vivo quantification of the developing optical properties and graded index of the embryonic eye lens using SPIM". In: *Biomedical optics express* 9.5 (cit. on p. 40).

Zeiler, M. D. (2012). "ADADELTA: An Adaptive Learning Rate Method". In: *arXiv preprint arXiv:1212.5701* (cit. on p. 79).

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017a). "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising". In: *IEEE Transactions on Image Processing* 26.7 (cit. on pp. 74, 92).

Zhang, K., Zuo, W., Gu, S., and Zhang, L. (2017b). "Learning Deep CNN Denoiser Prior for Image Restoration". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2 (cit. on p. 75).

Zhang, Q., Eagleson, R., and Peters, T. M. (2011). "Volume visualization: a technical overview with a focus on medical applications". In: *Journal of digital imaging* 24.4 (cit. on p. 16).

Zhang, X., Bloch, S., Akers, W., and Achilefu, S. (2012). "Near-infrared molecular probes for in vivo imaging". In: *Current protocols in cytometry* 60.1 (cit. on p. 39).

Zhou, Y.-T., Chellappa, R., Vaid, A., and Jenkins, B. K. (1988). "Image restoration using a neural network". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36.7 (cit. on p. 74).

Zhu, C. and Liu, Q. (2013). "Review of Monte Carlo modeling of light transport in tissues". In: *Journal of Biomedical Optics* 18.5 (cit. on p. 40).

Zhu, X., Hu, R., Brissova, M., Stein, R. W., Powers, A. C., Gu, G., and Kaverina, I. (2015). "Microtubules negatively regulate insulin secretion in pancreatic β cells". In: *Developmental Cell* 34.6 (cit. on p. 120).

Zoran, D. and Weiss, Y. (2011). "From Learning Models of Natural Image Patches to Whole Image Restoration". In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE (cit. on p. 23).

List of Publications

The following publications contain material presented in this thesis:

1. Royer, L. A., **Weigert**, **M.**, Günther, U., Maghelli, N., Jug, F., Sbalzarini, I. F., and Myers, E. W. (2015). "ClearVolume: open-source live 3D visualization for light-sheet microscopy". In: *Nature Methods* 12.6

2. Patel, A., Lee, H. O., Jawerth, L. M., Maharana, S. K., Jähnel, M., Hein, M. Y., Stoynov, S. S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., **Weigert**, **M.**, Myers, E. W., Grill, S. W., Drechsel, D. N., Hyman, A. A., and Alberti, S. (2015). "A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation". In: *Cell* 162

3. **Weigert**, **M.**, Royer, L., Jug, F., and Myers, G. (2017). "Isotropic Reconstruction of 3D Fluorescence Microscopy Images Using Convolutional Neural Networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*

4. **Weigert**, **M.**, Subramanian, K., Bundschuh, S. T., W, E., and Kreysing, M. (2018b). "Biobeam – Multiplexed wave-optical simulations of light-sheet microscopy". In: *PLoS Computational Biology*

5. Sui, L., Alt, S., **Weigert**, **M.**, Dye, N., Eaton, S., Jug, F., Myers, E. W., Jülicher, F., Salbreux, G., and Dahmann, C. (2018). "Differential lateral and basal tension drive folding of Drosophila wing discs through two distinct mechanisms". In: *Nature Communications* 9.1

6. **Weigert**, **M.**, Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., Rocha-Martins, M., Segovia-Miranda, F., Norden, C., Henriques, R., Zerial, M., Solimena, M., Rink, J., Tomancak, P., Royer, L., Jug, F., and Myers, E. W. (2018a). "Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy". In: *Nature Methods* 15.12

Erklärungen zur Eröffnung des Promotionsverfahrens

1. Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.
2. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten:
3. Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen fuïr Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.
4. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehoörde vorgelegt und ist auch noch nicht veröffentlicht worden.
5. Ich bestätige, dass ich die geltende Promotionsordnung der Fakultät Informatik der Technischen Universität Dresden anerkenne.

Dresden, November 7, 2019

A Appendix

A.1 Supplementary Videos

The supplementary videos can be found at http://mpi-cbg.de/~mweigert/thesis.

Supplementary Video 1
Real-time 3D rendering of early* C. elegans *embryo acquisitions
Real-time 3D rendering of a developing *C. elegans* embryo (His-GFP+BTub-GFP) during acquisition on a light-sheet microscope. The video shows the embryo at 12 cells, two of which are undergoing mitosis (MS and E founder cells). Rendering is done in real-time via *ClearVolume* running on the high-speed light sheet microscope capable of acquiring 6 stacks per second. In a second part the video shows a full screen view of the same embryo where we center, zoom and rotate around a cell division. In a third part the video shows the same embryo imaged at a later stage.

Supplementary Video 2
***Rendering of time-lapse data sets with* spimagine**
The video demonstrates offline rendering of time-lapse data sets with *spimagine*, a Python tool for visualization and processing of time-lapse microscopy volumes.

Supplementary Video 3
Real-time denoising and deconvolution
Demonstration of real-time denoising and deconvolution applied during acquisition of HeLa cells (FUS labeling) at a custom light-sheet microscope. Shown is the processing interface of *ClearVolume* that integrates tightly with the microscope control software.

Supplementary Video 4
Real-time image sharpness computational on the microscope
Shown is the image sharpness measure computed by *ClearVolume* during calibration of a custom light-sheet microscope. Visualizing the computed image quality in real-time (blue curve) provides instant feedback to the microscope controller regarding the effects of adjustments on the optical quality of the imaging process in 3D.

Supplementary Video 5
Wave-optical simulation of image-formation in light-sheet microscopy

The tissue model represents a multicellular (760 nuclei) organism of size ($100\mu m \times 200\mu m \times 100\mu m$) in an aqueous medium with $n = 1.33$. The refractive index distribution is in the range $(1.35, 1.42)$ comprising reference values for cell nuclei, eggshell and the cytoplasm. Weak absorption is homogenously present, but could also be localized (e.g. a spherical absorbing compartment in the center). The simulations of both the illumination and detection processes were carried out on a computational grid of $(1024 \times 2048 \times 1024)$ voxels with a spacing of $100nm$ along each dimension. The illumination field is a cylindrical light sheet with $NA_{illum} = 0.1$ focused laterally at the center and the detection system was assumed to have $NA_{detect} = 0.6$. For generating the final stack both illumination and detection fields were simulated at 200 different axial positions. The deterioration of both resolution and intensity at regions where photons along either the illumination detection path had to travel through large inhomogeneities can clearly be seen.

Supplementary Video 6
Illustration of a single PSF calculation

Illustration of a single PSF calculation insight the tissue via the propagation of analytically defined diffraction-limited input fields. Due to the linearity of wave-optics, these PSF calculations can be highly multiplexed.

Supplementary Video 7
Simulation of microscopy images with biobeam

A video illustrating rigorous wave-optical mimicry of a wide-field microscope as simulated with *biobeam*. The imaging of a $100\mu m^2$ test chart is simulated while a refractive sphere is continuously introduced into the microscope's optical path. *Biobeam* generated the underlying wave-optical simulations in 30 seconds.

Supplementary Video 8
Screencast of an interactive biobeam session

Screencast of an interactive command line session of *biobeam*, demonstrating its capabilities and speed. Shown are the generation of Gaussian and Bessel input fields, the propagation of a light sheet through a synthetic refractive sphere, and the multiplexed PSF grid calculations. The video is played in real time.

Supplementary Video 9
Different illumination modes available in biobeam
Wave-optical simulation of different illumination modes and simulated light sheets available in *biobeam*. Video showing the predefined illumination modes and simulated light sheets being scanned through a biological plausible tissue model. Both coherent (cylindrical lens SPIM) illumination and partially-incoherent illumination modes (time scanned Gaussian/Bessel beams) are simulated.

Supplementary Video 10
Challenges in live-cell imaging of flatworm Schmidtea mediterranea
Challenges in time-lapse imaging of flatworm *Schmidtea mediterranea*. Image stacks of RedDot1-labeled, anesthetized specimen were acquired every 2 min with a Spinning disk confocal microscope ($NA = 1.05$), at high and low SNR (illumination) conditions (10% laser, 30ms exposure per plane vs 0.5% laser, 10ms per plane). Whereas in the high-SNR case the specimen shows illumination induced twitching, the image quality in the low-SNR case is insufficient for further analysis. Network restoration enables us to recover high-SNR images from images acquired at low-SNR conditions without twitching of the specimen, thus providing a practical framework for live-cell imaging of *Schmidtea mediterranea*.

Supplementary Video 11
Restoration of low-SNR acquisitions and comparison with ground-truth
Restoration results of low-SNR acquisitions of *Schmidtea mediterranea* and comparison to ground-truth. Shown are a 3D rendering of the results on a multi-tiled acquisition ($8192 \times 3072 \times 100$ pixels) and the comparison with ground-truth.

Supplementary Video 12
Restoration of low-SNR acquisitions of developing Tribolium
Restoration of low SNR volumetric time-lapses of developing *Tribolium castaneum* embryos (EFA::nGFP labeled nuclei). Acquisition was done on a Zeiss LSM 710 NLO Multiphoton Laser Scanning Microscope with a time-step of 8 min, and stack-size of a single time-point $760 \times 760 \times 100$ pixels. Shown are maximum-intensity projection and single slices of the raw stacks, the network prediction and the high-SNR ground-truth.

Supplementary Video 13
Surface projection of developing Drosophila wing epithelia
Joint surface projection and denoising of developing *Drosophila melanogaster* wing epithelia. 3D image stacks of the developing wing of a membrane labeled (Ecad::GFP) fly pupa were acquired with a spinning disk confocal ($63\times$, $NA = 1.3$) microscope. We show the projected epithelial surface obtained by a conventional method (PreMosa Blasse et al. 2017), the restoration network, and the projected ground-truth. We applied a random-forest based cell segmentation pipeline and show segmentation/tracking results, demonstrating vastly improved accuracy for the restoration when compared to the conventionally processed raw stacks.

Supplementary Video 14
Isotropic restoration of anisotropic acquisitions of Drosophila embryos
Isotropic restoration of anisotropic time-lapse acquisitions of hisGFP-tagged developing *Drosophila melanogaster* embryos. We used the original, pre-processed data set of (Royer et al. 2016), acquired with a light-sheet microscope ($NA = 1.1$) with 5-fold axially undersampled resolution (lateral/axial pixel size: $0.39\mu m/1.96\mu m$). The video shows different axial (xz) regions of a single time-point from both the original (input) stacks and the isotropic restoration.

Supplementary Video 15
Isotropic restoration of anisotropic acquisitions of zebrafish retina
Isotropic restoration of anisotropic dual-color acquisitions of developing *Danio rerio* retina. The data was acquired with a Spinning disk confocal microscope (Olympus $60\times, NA = 1.1$) and exhibits a 10-fold axial anisotropy (lateral/axial pixel size: $0.2\mu m/2.0\mu m$); labeled structures are nuclei (DRAQ5, magenta) and nuclear envelope (GFP+LAP2b, green). The video shows a rendering of the dual-color input stack and its isotropic reconstruction.

Supplementary Video 16
Reconstruction of diffraction-limited structures in secreting beta cells
Enhancement of diffraction-limited structures in widefield images of rat INS-1 (beta) cells. The video shows time-lapses of several INS-1 cells, acquired with the widefield mode of a DeltaVision OMX microscope ($63\times$, $NA = 1.43$). Labeled are secretory granules (pEG-hIns-SNAP, magenta) and microtubules (SiR-tubulin, green). Next to the time-lapse of the widefield images we show the output of the reconstruction networks.

Supplementary Video 17
Reconstruction of diffraction-limited structures in HeLa cells

Enhancement of diffraction-limited widefield images of GFP-labeled microtubules in HeLa cells and comparison with SRRF (super-resolution radial fluctuations N. Gustafsson et al. 2016). Images were acquired with a Zeiss Elyra PS.1 microscope in TIRF mode ($100\times$, $NA = 1.46$). The video shows the widefield input sequence, the network restoration and the corresponding SRRF images. Note that the time-resolution of the SRRF image sequence is 20 times less then the network restoration, as 20 times more images have to be processed for the same restoration quality.

A.2 Focus Field Calculations

Here we want to calculate the three-dimensional light distribution in the focus of a lens of given numerical aperture $NA = n_0 \sin\alpha$. The following assumptions are made:

- light incident at the back aperture is polarized in x, *i.e.* only has components $\vec{E} = E_x^0 \vec{e}_x$
- lens satisfies sine condition

then the general vectorial light field $\vec{E}(\rho, \varphi, z)$ around the focus is given by (Foreman et al. 2011)

$$\begin{aligned} E_i(\rho,\varphi,z) &= \frac{-ikfE_x^0}{4\pi}\int_0^\alpha\int_0^{2\pi} d\theta d\phi\, B_i(\theta,\phi)P(\theta,\phi)\sqrt{\cos\theta}\sin\theta\cdot e^{ik\rho\sin\theta\cos(\phi-\varphi)}e^{ikz\cos\theta} \\ B_x(\theta,\phi) &= (\cos\theta+1)+(\cos\theta-1)\cos 2\phi \\ B_y(\theta,\phi) &= (\cos\theta-1)\sin 2\phi \\ B_z(\theta,\phi) &= -2\sin\theta\cos\phi \end{aligned} \tag{A.1}$$

For a constant pupil function $P(\theta,\phi) = 1$ the inner integration with respect to ϕ can be carried out explicitly to yield Bessel functions of different order $J_i(z)$ and one gets

$$\begin{aligned} \vec{E}(\rho,\varphi,z) &= \frac{-ikfE_x^0}{2}\begin{pmatrix} I_0 + I_2\cos 2\varphi \\ I_2\sin 2\varphi \\ -2iI_1\cos\varphi \end{pmatrix} \\ I_0 &= \int_0^\alpha d\theta\sqrt{\cos\theta}\sin\theta(\cos\theta+1)J_0(k\rho\sin\theta)e^{ikz\cos\theta} \\ I_1 &= \int_0^\alpha d\theta\sqrt{\cos\theta}\sin^2\theta J_1(k\rho\sin\theta)e^{ikz\cos\theta} \\ I_2 &= \int_0^\alpha d\theta\sqrt{\cos\theta}\sin\theta(\cos\theta-1)J_2(k\rho\sin\theta)e^{ikz\cos\theta} \end{aligned} \tag{A.2}$$

For small numerical apertures we have $\theta < 1$ and one can approximate the integrands up to first order ($\cos \approx 1, \sin\theta \approx \theta \ldots$) to get

$$\begin{aligned} I_0 &\approx 2\int_0^\alpha d\theta J_0(k\rho\sin\theta)e^{ikz\cos\theta} \\ I_1 &\approx I_2 \approx 0 \end{aligned} \tag{A.3}$$

which reduces to the Born-Wolff model of the scalar PSF (Born et al. 1999).

A.3 Zernike Polynomials

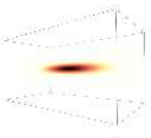
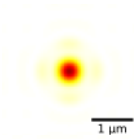
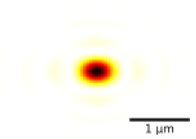
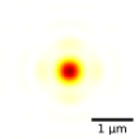
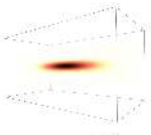
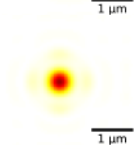
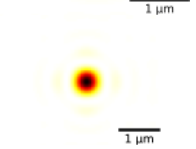
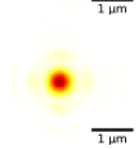
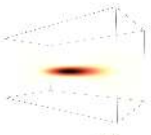
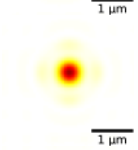
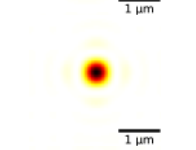
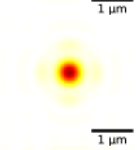
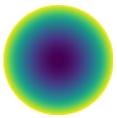
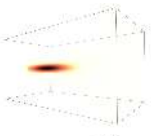
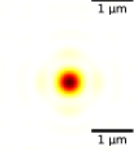
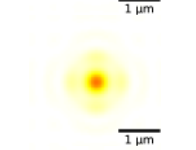
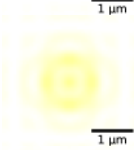
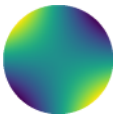
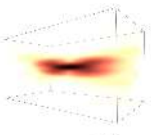
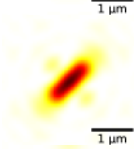
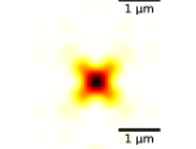
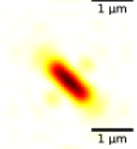
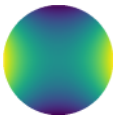
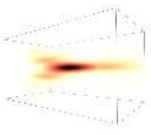
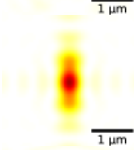
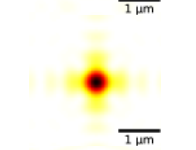
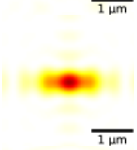
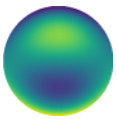
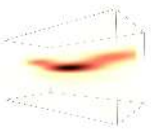
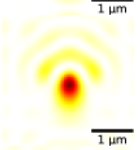
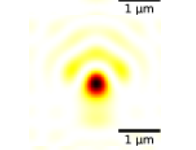
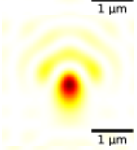
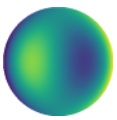
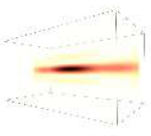
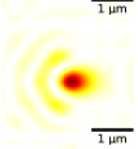
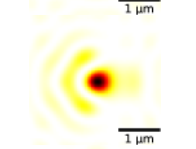
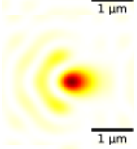
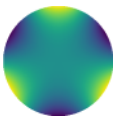
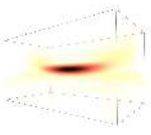
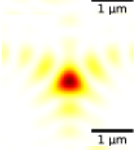
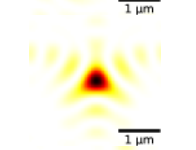
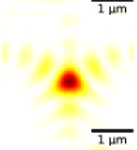
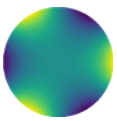
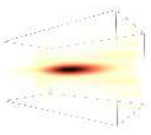
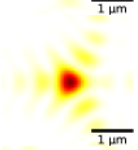
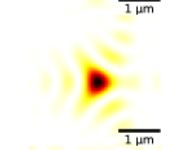
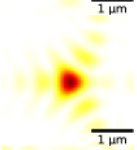
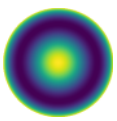
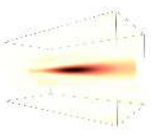
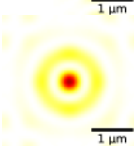
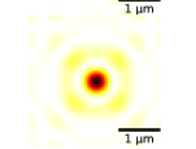
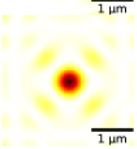
Noll index	Wavefront	PSF (3D)	$-1\mu m$	$0\mu m$	$+1\mu m$
1 piston					
2 tip					
3 tilt					
4 defocus					
5 oblique astigmatism					
6 vertical astigmatism					
7 vertical coma					
8 horizontal coma					
9 vertical trefoil					
10 oblique trefoil					
11 primary spherical					

Table A.1: Zernike modes and resulting PSF

A.4 Receptive Field of Convolutional Networks

The receptive field of an output pixel $y_{i,j}$ within neural network is simply the set of all pixels in the input image $x_{i,j}$ that contribute to the output value of y, which *e.g.* for a 2D fully convolutional network with isotropic kernel sizes and simple max-pooling is simply a square $b \times b$.

Let the network $\mathcal{N}$ be given by its successive layers $\mathcal{N} := [L_1, L_2, \ldots, L_n]$:

$$L_i = \begin{cases} C(k) & \text{Convolution with a kernel } k \\ P(\sigma) & \text{Pooling by a factor } \sigma \\ U(\sigma) & \text{Upsampling by a factor } \sigma \end{cases} \tag{A.4}$$

Let b_m be the receptive field after layer L_m, thus b_n the receptive field of the full network. Than we have by induction:

$$b_0, s_0 = 1, 1 \tag{A.5}$$

$$b_{m+1}, s_{m+1} = \begin{cases} b_m + s_m(k-1), s_m & \text{if } L_m = C(k) \\ s_m(\sigma - 1)b_m, \sigma s_m & \text{if } L_m = P(\sigma) \\ b_m, \frac{s_m}{\sigma} & \text{if } L_m = U(s) \end{cases} \tag{A.6}$$

A.5 Scaled Normalization

For two given images x and y we want to find the affine parameters a, b such that the $L = \text{MSE}(x - (ay + b))^2$ is minimized. Both a and b can be computed explicitly:

$$L(a,b) = \langle (x - a(y-b))^2 \rangle \tag{A.7}$$

$$= \langle x^2 \rangle - 2a\langle x(y-b) \rangle + a^2 \langle (y-b)^2 \rangle \tag{A.8}$$

$$= \langle x^2 \rangle - 2a\langle xy \rangle + 2ab\langle x \rangle + a^2\langle y^2 \rangle - 2a^2 b\langle y \rangle + a^2 b^2 \tag{A.9}$$

$$\frac{\partial L}{\partial a} = -2\langle xy \rangle + 2b\langle x \rangle + 2a\langle y^2 \rangle - 4ab\langle y \rangle + 2ab^2 = 0 \tag{A.10}$$

$$\frac{\partial L}{\partial b} = 2a\langle x \rangle - 2a^2\langle y \rangle + 2a^2 b = 0 \tag{A.11}$$

yielding finally

$$a = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle y^2 \rangle - \langle y \rangle^2} = \frac{\text{Cov}(x,y)}{\text{Var}(y)} \qquad b = \langle y \rangle - \frac{\langle x \rangle}{a} = \langle y \rangle - \frac{\langle x \rangle \text{Var}(y)}{\text{Cov}(x,y)} \tag{A.12}$$

With $\text{Var}(y)$ and $\text{Cov}(x,y)$ being the classical variance of y and covariance between x and y respectively.

A.6 Additional Material for Care Experiments

	Planaria	Tribolium	Flywing	Retina	INS-1 (tubules)	INS-1 (granules)	HeLa cells
Pooling layers	2	2	3	2	2	2	2
Filter size	3	3	3	5	5	5	5
Init. conv. feat.	32	32	32	32	32	32	32
No. of param.	996802	996802	1386347	923876	923010	923010	923010
Input size	(16, 64, 64, 1)	(16, 64, 64, 1)	(50, 64, 64, 1)	(128, 128, 2)	(128, 128, 1)	(128, 128, 1)	(128, 128, 1)
Output size	(16, 64, 64, 2)	(16, 64, 64, 2)	(64, 64, 2)	(128, 128, 4)	(128, 128, 2)	(128, 128, 2)	(128, 128, 2)
No. of images	17005	14725	16891	24121	5578	2728	5578
Batch size	16	16	16	64	32	32	32
No. of epochs	200	200	100	100	100	100	100
Learning rate	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
Training time	30h 06m	26h 12m	10h 17m	06h 15m	01h 23m	00h 39m	01h 23m

Table A.2: Details for network architecture and training. The top part of the table provides details w.r.t. the encoder-decoder architecture. Specifically, the number of pooling/upsampling layers (indicating the number of resolution levels), the size of the convolution filters in all image dimensions, and the number of initial convolution filters, which is doubled after each pooling layer, and halved after each upsampling layer. It also refers to the figure where the network topology is shown and lists the total number of network parameters. The bottom part pertains to learning details, first showing the input size of the training images and the corresponding output size of the network; note that a network can be applied to different image sizes after training. Further shown are the number of training images, batch size, number of epochs, learning rate, and total training time on a single Nvidia 1080 GPU.

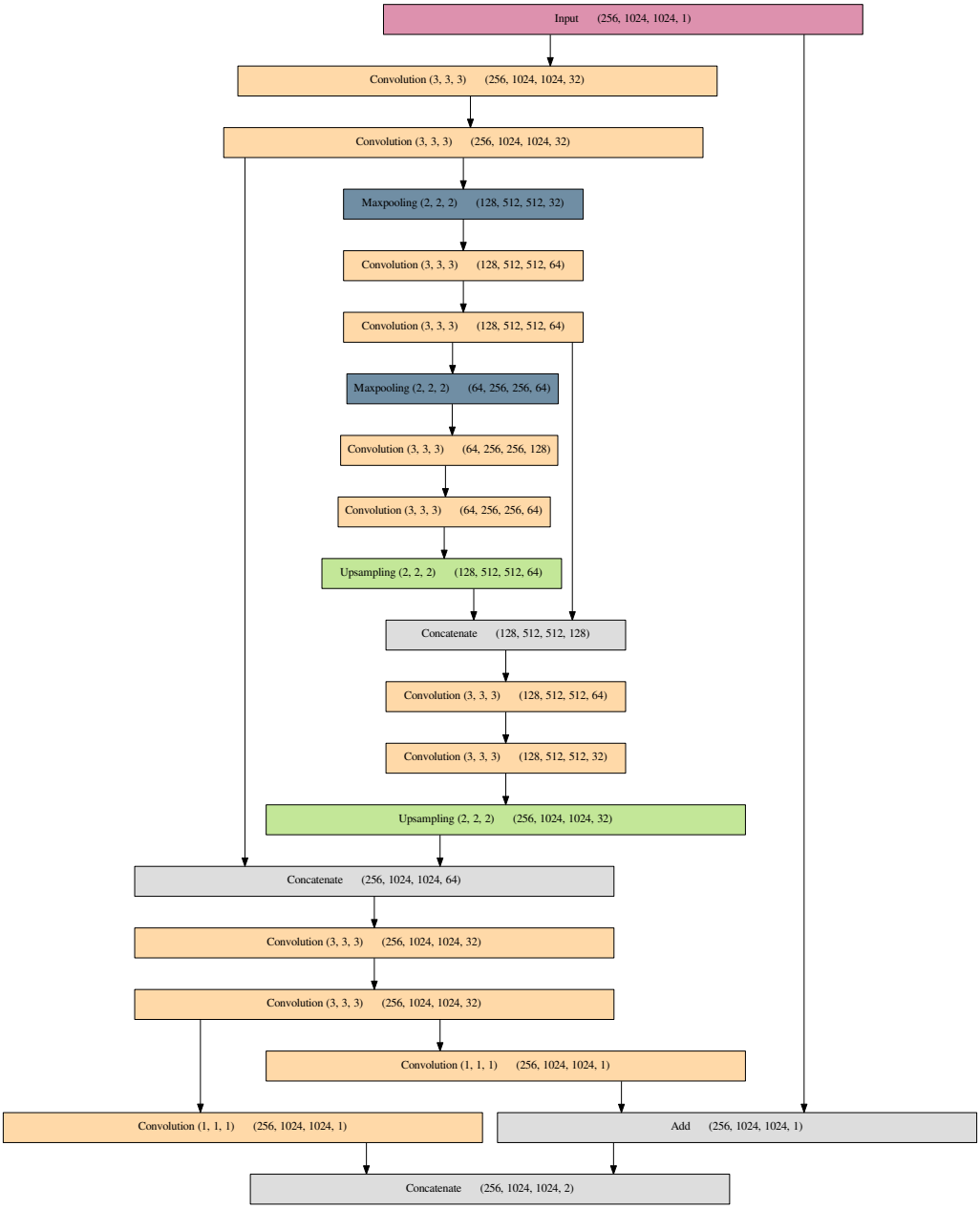


Figure A.1: Network architecture used for low-SNR restoration (shown for 3D inputs). We use an architecture based on the U-Net (Ronneberger et al. 2015), but predict a per-pixel distribution (parameterized by mean and scale) resulting in a 2-channel output. The mean is calculated as the sum of the input and the internal predictions (*residual* addition layer near the end). (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

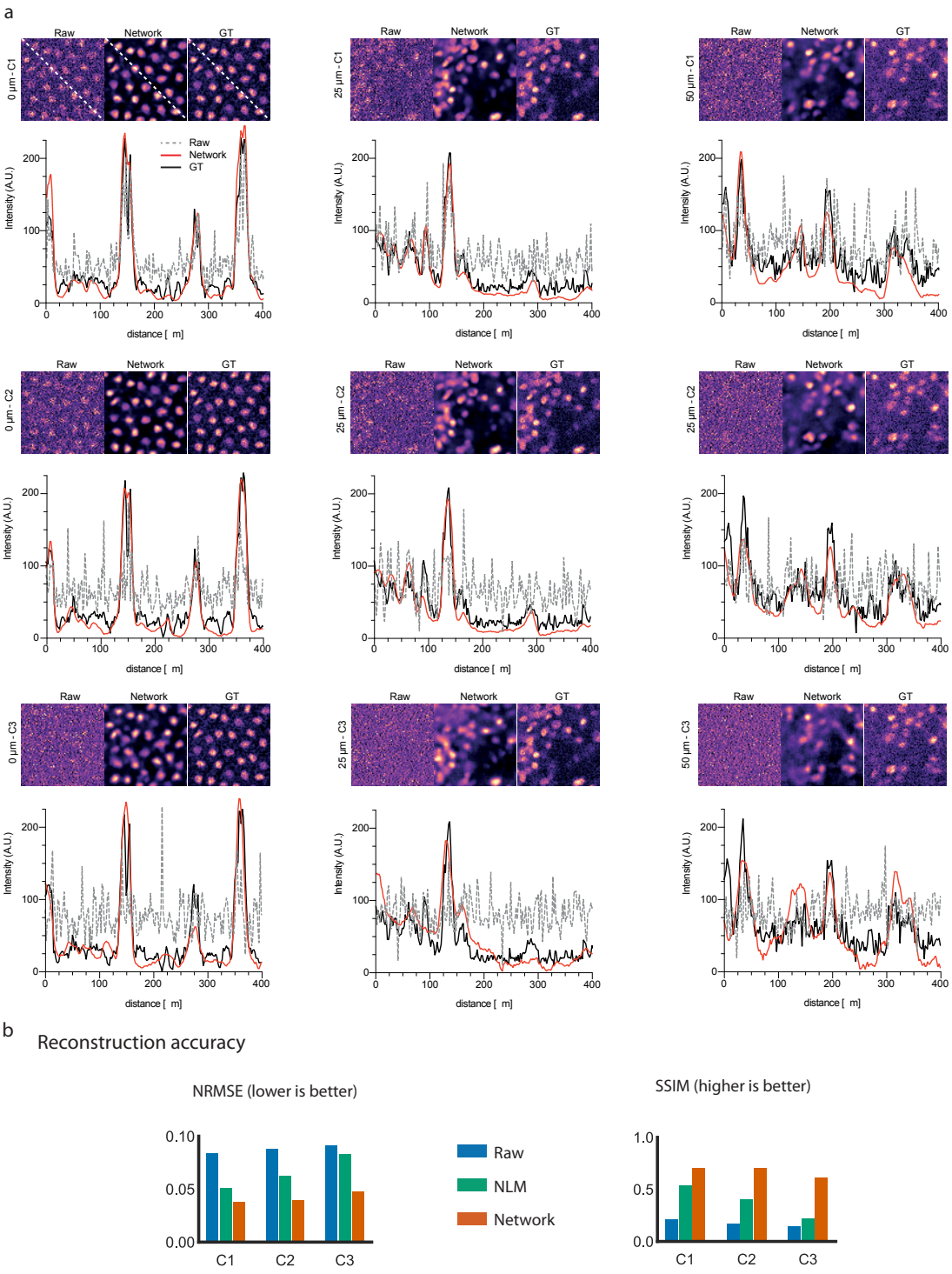


Figure A.2: Restoration of low SNR volumes of *Schmidtea mediterranea*. a) Reconstructions for different SNR conditions (C1-C3, top to bottom) at different tissue depths ($0 - 50\mu m$, left to right) and corresponding lineplots. As SNR decreases with increasing depth, both the input and the reconstruction quality decreases.b) Reconstruction accuracy measured on held out test set (comprising 22 volumetric stacks for each condition). Shown are normalized root mean squared error (NRMSE, lower is better) and structural similarity (SSIM, higher is better). We compare the error between ground-truth *vs*. raw stacks, the network prediction and a strong denoising baseline (Non-local-means with parameters tuned to the test set). (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)

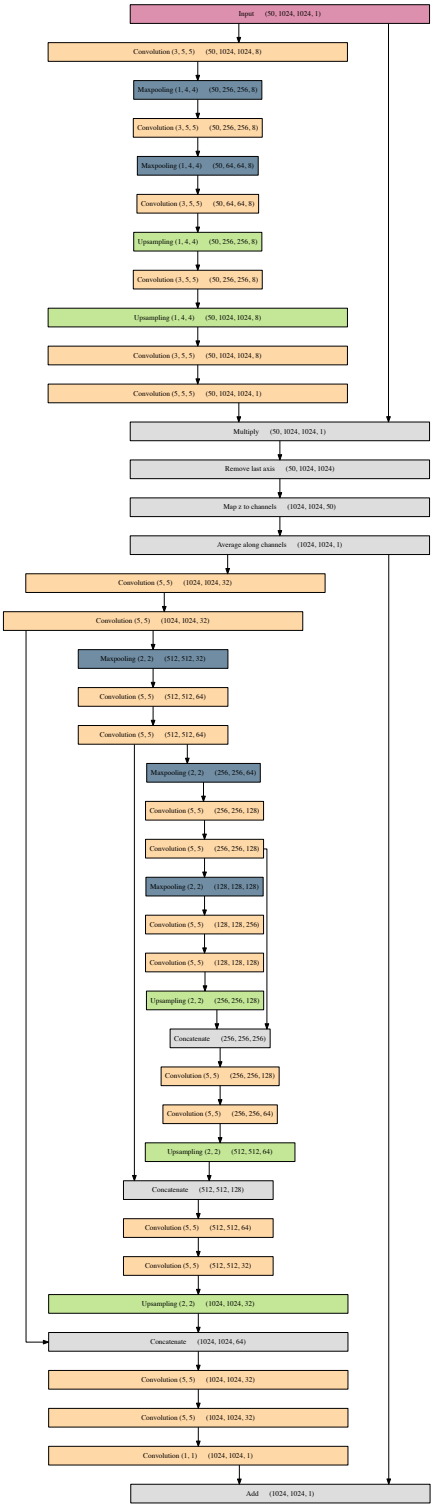


Figure A.3: Network architecture used for joint surface projection and denoising. The first encoder-decoder part (until "Multiply"-layer) estimates the probability of a certain pixel to belong to the surface, whereas the later parts restore and denoise the projected signal. We depict the non-probabilistic version here, for the probabilistic version, the pen-ultimate convolutional layer is split into the mean and scale parameter layers as in Fig. A.1. (Reproduced with permission from Weigert et al. 2018a. Copyright 2018, Springer Nature)