Computational Methods for 3D Imaging of Neural Activity in Light-Field Microscopy

Herman Isaac Verinaz Jadan

Supervised by Professor Pier Luigi Dragotti

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Electrical and Electronic Engineering of Imperial College London
and the Diploma of Imperial College London

Department of Electrical and Electronic Engineering
Imperial College London
2022

Declaration of Originality

I herewith declare that the content of this thesis is the product of my own work under the guidance of my PhD supervisor Prof. Pier Luigi Dragotti. All material that is not my own work has been properly acknowledged.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Light Field Microscopy (LFM) is a 3D imaging technique that captures spatial and angular information from light in a single snapshot. LFM is an appealing technique for applications in biological imaging due to its relatively simple implementation and fast 3D imaging speed. For instance, LFM can help to understand how neurons process information, as shown for functional neuronal calcium imaging. However, traditional volume reconstruction approaches for LFM suffer from low lateral resolution, high computational cost, and reconstruction artifacts near the native object plane. Therefore, in this thesis, we propose computational methods to improve the reconstruction performance of 3D imaging for LFM with applications to imaging neural activity.

First, we study the image formation process and propose methods for discretization and simplification of the LF system. Typical approaches for discretization are performed by computing the discrete impulse response at different input locations defined by a sampling grid. Unlike conventional methods, we propose an approach that uses shift-invariant subspaces to generalize the discretization framework used in LFM. Our approach allows the selection of diverse sampling kernels and sampling intervals. Furthermore, the typical discretization method is a particular case of our formulation.

Moreover, we propose a description of the system based on filter banks that fit the physics of the system. The periodic-shift invariant property per depth guarantees that the system can be accurately described by using filter banks. This description leads to a novel method to reduce the computation time using singular value decomposition (SVD). Our simplification method capitalizes on the inherent low-rank behaviour of the system. Furthermore, we propose rearranging our filter-bank model into a linear convolution neural network (CNN) that allows exploiting existing deep-learning software for computation.

Then, we study the problem of 3D reconstruction from a single LF image. We pro-

pose the shift-invariant-subspace assumption as a prior for volume reconstruction under ideal conditions. We experimentally show that artifact-free reconstruction (aliasing-free) is achievable under these settings. Furthermore, the tools developed to study the forward model are exploited to design a reconstruction algorithm based on ADMM that allows artifact-free 3D reconstruction for real data. Contrary to traditional approaches, our method includes additional priors for reconstruction without dramatically increasing the computational complexity. We extensively evaluate our approach on synthetic and real data and show that our approach performs better than conventional model-based strategies in computation time, image quality, and artifact reduction.

Finally, we exploit deep-learning techniques for reconstruction. Specifically, we propose to use two-photon imaging to enhance the performance of LFM when imaging neurons in brain tissues. The architecture of our network is derived from a sparsity-based algorithm for reconstruction named Iterative Shrinkage and Thresholding Algorithm (ISTA). Furthermore, we propose a semi-supervised training based on Generative Adversarial Neural Networks (GANs) that exploits the knowledge of the forward model to achieve remarkable reconstruction quality. We propose efficient architectures to compute the forward model using linear CNNs. This description allows fast computation of the forward model and complements our reconstruction approach. Our method is tested under adverse conditions: lack of training data, background noise, and non-transparent samples. We experimentally show that our method performs better than model-based reconstruction strategies and typical neural networks for imaging neuronal activity in mammalian brain tissue. Our approach enjoys both the robustness of the model-based methods and the reconstruction speed of deep learning.

Acknowledgment

I want to thank my family for their constant and invaluable support throughout my entire life. Thanks, mom, dad, and sister, for your help in every project I undertook and for cheering me up when I needed it.

Thank Prof. Dragotti for his guidance and support throughout this long journey. He trusted and guided me during my research, and he supported me during other difficulties in my life. I am very grateful to have had an excellent supervisor and mentor.

I want to express my gratitude to the LFM research group. Working in the group helped me to grow in my career. Thanks, Pingfan, Carmel, Peter, and Amanda, for your guidance and interesting discussion during our meetings.

I want to thank all my friends during my stay in London. Thanks to all the people in the CSP group who helped me enjoy the journey towards my Ph.D. even more. Special thanks to my friends from Pier Luigi's research group for the funny chats and your support. Thanks to my old friends from Ecuador for our conversations that helped me when I needed a break from work. I am pleased to have met all of you guys. You are incredible people.

Finally, I would like to thank SENESCYT and the Walter Valdano program from my alma mater, ESPOL, that funded my MSc at Imperial College, allowing me to start this adventure. I would also like to thank the Electrical and Electronic Engineering Department of Imperial College for funding my Ph.D.

Abbreviations

2P Two-Photon

ADMM Alternating direction method of multipliers

ANN Artificial neural networks

CFT Continuous Time Fourier Transform

CMOS Complementary Metal Oxide Semiconductor

CNN Convolutional neural network

CSC Convolutional Sparse Coding

DFT Discrete Fourier Transform

EPI Epipolar plane Image

GAN Generative adversarial network

GECI Genetically encoded calcium indicators

ISRA Iterative Image Space Reconstruction Algorithm

ISTA Iterative shrinkage and thresholding algorithm

LASSO Least Absolute Shrinkage and Selection Operator

LF Light Field

LFM Light Field Microscopy

LFP Light Field Photography

LISTA Learned Iterative shrinkage and thresholding algorithm

LSGAN Least squares generative adversarial network

MLA Micro-lens Array

MSE Mean Square Error

nD n-dimensional

NIP Native Image Plane

NIR Near-infrared

NOP Native Object Plane

PSF Point Spread Function

PSNR Peak Signal-to-Noise Ratio

RL Richardson-Lucy algorithm

RMSE Root Mean Square Error

RNN Recurrent neural network

SIS Shift-Invariant Subspace

SP Sensor Plane

SPIM Single plane selective-plane illumination microscopy

SSIM Structural SIMilarity

STFT Short-time Fourier Transform

SVD Singular Value Decomposition

WDF Wigner Distribution Function

WF Wide Field

Notations

$\mathbf{A}_{\varphi}$**:** Analysis operator with $\varphi[\cdot]$ as template

$\mathbf{B}_{\mathbf{H}}$**:** Elementary matrix that circularly repeated forms $\mathbf{H}_i$

$g(x)$**:** Continuous LF output signal

$f(x, z)$**:** Continuous volumetric input signal

$\mathbf{H}_{\delta}, \mathbf{H}$**:** Conventional LFM measurement matrix

Δx_1**:** Lateral Sampling interval for $f(x, z)$

Δx_2**:** Sampling interval for $g(x)$

Δz**:** Axial sampling interval for $f(x, z)$

$g[k]$**:** Discrete LF output signal

$f[k, l], f_{k,l}$**:** Discretized volumetric input signal

$\tilde{\varphi}(x, z)$**:** Dual basis function of $\varphi(x, z)$

$h(x, \mathbf{p})$**:** Impulse response of $\mathcal{H}$

$\mathbf{H}_i$**:** LFM measurement matrix for the i-th depth

$\mathcal{H}_d$**:** LFM operator

$\mathcal{H}$**:** LFM operator with continuous output

T**:** Microlens Pitch

N**:** Number of pixels under each microlens

$d(x)$**:** Output sampling filter

$\mathbf{S}_i$**:** Operator that extracts the depth i

$\varphi(x, z)$**:** Synthesis filter

$\mathbf{S}_{\varphi}$**:** Synthesis operator with $\varphi[\cdot]$ as template

s**:** Upsampling factor

$\mathbf{U}_m$**:** Upsampling matrix of factor m

f: Vectorized input volume

g: Vectorized light field

Contents

Declaration of Originality ii

Copyright Declaration iii

Abstract iv

Acknowledgment vi

Abbreviations vii

Notations ix

Contents xi

List of Figures xiv

List of Tables xvi

Chapter 1. Introduction 1

1.1 Motivation and Objectives 1

1.2 Original Contribution and Outline of Thesis 3

1.3 Publications 5

Chapter 2. Background 8

2.1 Imaging Neural Activity 8

2.2 The Concept of LF and LF Devices 11

2.2.1 Ray-Optics Perspective 12

2.2.2 Scalar-Wave-Optics Perspective 19

2.2.3 Phase-Space-Optics Perspective 23

2.3 Computational Methods for 3D Reconstruction from LFM Data 27

2.3.1 Model-based Methods 28

2.3.2 Learning-based Methods 29

Chapter 3. Shift-Invariant-Subspace Discretization and Model Simplifica-

tion **30**
3.1 Forward Model Analysis . . . 31
3.1.1 A General Discretization Framework . . . 34
3.1.2 Filter-Bank Description and Model Simplification . . . 36
3.1.3 Experiments and Results . . . 39
3.2 Forward Model as a Linear CNN . . . 42
3.2.1 4D Representation of LF . . . 42
3.2.2 Linear CNN . . . 42
3.3 Summary . . . 44

Chapter 4. Model-Based Reconstruction for LFM **46**
4.1 Ideal Reconstruction Under Shift-invariant-subspace Assumption . . . 46
4.1.1 Experiments . . . 48
4.2 General Scenario and Additional priors . . . 50
4.3 Experiments and Results . . . 52
4.3.1 Synthetic Data . . . 53
4.3.2 Additional Datasets . . . 59
4.3.3 Real Data . . . 63
4.4 Summary . . . 70

Chapter 5. Physics-based Deep Learning for Imaging Neuronal Activity via Two-photon and Light Field Microscopy **71**
5.1 Problem Formulation . . . 73
5.2 Efficient Implementations of the Forward Model . . . 75
5.3 3D Reconstruction . . . 76
5.3.1 Deep Neural Network Architecture for Volume Reconstruction . . . 77
5.3.2 Training Strategy . . . 79
5.4 Experiments and Results . . . 81
5.4.1 Experimental Setup . . . 81
5.4.2 Deep-learning Setup . . . 82
5.4.3 Reconstruction of Structural 3D Images from Ligh Field Images . . . 85
5.4.4 Reconstruction of Volume Time Series from LF Images . . . 87
5.5 Summary . . . 91

Chapter 6. Conclusions **92**
6.1 Summary of Thesis Achievements . . . 92
6.2 Future Work . . . 94
6.2.1 Forward Model Calibration or Learning . . . 94
6.2.2 Exploitation of Temporal Correlation in LF Stacks . . . 95

6.2.3 Design of Discriminators . . . 95
6.2.4 Exploration of Alternative LFM designs . . . 96

Appendix A. **97**
A.1 Fresnel-Kirchhoff Diffraction . . . 97
A.1.1 Fesnel-Kirchoff Diffraction Formula . . . 98
A.1.2 Rayleigh-Sommerfeld Solutions . . . 98
A.2 Angular Spectrum Representation . . . 99
A.3 Lenses . . . 100
A.3.1 Simple Lens . . . 100
A.3.2 4f System . . . 102
A.4 Sampling the WDF . . . 103

Appendix B. **107**
B.1 Stopping Criteria . . . 107

Appendix C. **110**
C.1 Experimental Setup . . . 110
C.1.1 Light-field Modality Specifications . . . 110
C.1.2 Two-photon Modality Specifications . . . 111
C.1.3 Fluorophore Transfection . . . 111
C.1.4 Brain Slice Preparation . . . 112

Bibliography **113**

List of Figures

2.1 Imaging modalities 9
2.2 An optical diagram of LFM 11
2.3 Illustration of LF imaging 12
2.4 Definition of the LF function 13
2.5 Light field geometry for a point source. 14
2.6 Light field camera diagram. 16
2.7 Sampling pattern and kernel 18
2.8 LF microscope diagram. 20
2.9 Impulse response simulation 23
2.10 Sampling kernel predicted by phase-space optics 25
2.11 Block diagram of a LF microscope based on the WDF. 27

3.1 Description of a LF System 32
3.2 Discretization of the LF system using shift-invariant subspaces. 34
3.3 Filter bank representation. 38
3.4 Optimal analysis filters. 40
3.5 Model simplification example. 41
3.6 Linear forward model 43

4.1 Shift-invariant subspace assumption. 47
4.2 Ideal reconstruction using the pseudoinverse on synthetic data. 49
4.3 Reconstruction using synthetic LF data. 54
4.4 Reconstruction using noisy synthetic LF data. 57
4.5 Reconstruction using synthetic LF images computed from labelled blood vessels captured using confocal microscopy 60
4.6 Reconstruction using synthetic LF images computed from mouse brain slices captured using our two-photon microscope. 61
4.7 Reconstruction using real LF data from a brain slice imaged from an EGFP tagged triple transgenic mouse line. 64

4.8 Reconstruction using real LF data from acute mouse brain slices expressing the calcium indicator NIR-GECO2G 65
4.9 Reconstruction using real LF data from fluorescent beads 66
4.10 Reconstruction using ADMM with deep prior. 69

5.1 Overview of our approach 72
5.2 Forward model as a CNN 76
5.3 CNN architecture 78
5.4 Training of our GAN architecture 79
5.5 Reconstruction using real LF data from acute mouse brain slices expressing tdTomato fluorophore 84
5.6 Reconstruction using real LF data from acute mouse brain slices expressing the calcium indicator jGCaMP8f 88
5.7 Visual comparison of sub-aperture images 89
5.8 Reconstruction of the temporal evolution of neuron activity in brain tissue . 90

A.1 Scalar wave optics. 98
A.2 Microlens arrays. 101
A.3 4F system diagram. 102

B.1 Fidelity loss. 108
B.2 Reconstruction using a loss-based stopping criterion. 109

List of Tables

3.1 Forward model computation time. 44

4.1 Performance of model-based methods on synthetic data. 55
4.2 Performance of iterative methods with Poisson noise. 56
4.3 ADMM hyper-parameters. 59
4.4 Hyper-parameter study (PSNR). 59
4.5 Performance of iterative methods with different microscope settings. 62
4.6 Performance of learning methods on synthetic data. 67

5.1 PSNR and SSIM for real LF data of neurons imaged using tdTomato fluorophore. 86
5.2 Computation time. 86

Chapter 1

Introduction

1.1 Motivation and Objectives

UNDERSTANDING how the brain processes information is a critical goal of neuroscience. This ambitious task poses multiple challenges for scientists in many fields. In particular, understanding neuron communication mechanisms requires imaging the dynamics of large populations of neurons at high spatial and temporal resolution over large areas. Optical technology needed for this task has to simultaneously capture every neuron in the 3D space through highly scattering brain tissue.

Even though great efforts have been made to push forward fast optical techniques for 3D imaging, conventional modalities are still insufficient to monitor populations of neurons in the mammalian brain. For instance, wide-field microscopy increases photon generation rates and imaging speeds; however, it produces background interference since out-of-focus light appears in the in-focus image, reducing contrast. Two-photon microscopy explores near-infrared illumination, which implies deep tissue penetration, and reduced scattering. However, it requires sequential acquisition (point scanning), which limits the imaging speed. Multiple alternatives have been proposed to increase acquisition speed such as engineered beam trajectories [1–4], spatial or temporal multiplexing of multiple foci [5], [6–12], as well as sculpting fluorescence excitation into an extended point spread function [13–16], either scanned or targeted statically onto neurons of interest. However, these alternatives usually imply constrained sample geometries or are not well-suited for scattering samples.

LFM is an attractive technique for monitoring 3D neural activity with a fast frame

rate. LF microscopes record the light intensity together with the location and direction of arrival of the rays in a single snapshot. This capability can be achieved with a simple modification of a standard wide-field (WF) microscope: a microlens array is placed at the original imaging plane, and the imaging sensor is moved to the rear focal plane of the microlenses. Several works have reported the use of LFM for monitoring neural activity [17–19]. However, the benefits of LF imaging imply limited spatial resolution due to the recording of spatial and angular information with a single camera sensor and a MLA. Specifically, if the LF function is sampled with a high angular sampling density, it is necessarily sampled with a low spatial sampling density due to the MLA and camera sensor arrangement. Furthermore, LFM also suffers from image degradation due to scattering in brain tissue.

LFM requires fast and robust algorithms to accurately reconstruct 3D images from single LF images. Existing model-based reconstruction methods for LFM offer a degree of robustness when imaging scattering tissue. However, they need high computational resources and introduce reconstruction artifacts, reducing the merit of LFM. Studying large brain regions implies reconstructing thousands of 3D images from LF images. Since conventional reconstruction methods may require hours for single 3D reconstruction, reconstruction from large LF sequences becomes a bottleneck for LFM. New emerging learning-based techniques offer improved quality under specific scenarios but cannot address challenging acquisition conditions when imaging brain tissue, such as a lack of training data or highly scattering samples. Therefore, new computational methods which are faster, more accurate, and more robust are needed.

Model-based reconstruction methods in LFM use simplified frameworks to avoid further increases in computational requirements. Specifically, LFM needs to handle huge 3D images, to perform reconstruction from large LF stacks, and the system is not shift-invariant (it cannot be described with a single convolution). Thus, LFM poses a complex scenario that makes it challenging to use sophisticated methods for reconstruction. Practical reconstruction approaches for LFM must consider computational complexity since time-consuming methods make studying large LF sequences challenging. Existing techniques use modified versions of the Richardson-Lucy (RL) algorithm since it offers a fea-

sible reconstruction method. This work aims to investigate approaches that help manage data dimensionality. In general, we seek efficient methods for the computation of the forward model. Furthermore, we aim to explore perfect (aliasing-free) reconstruction from LF images under ideal settings. A broader goal is to study model-based reconstruction under real acquisition conditions such as noise, modelling imperfections, and lack of training data.

Deep-learning methods often challenge model-based methods in terms of performance and speed in many applications. For instance, they have achieved remarkable results in image restoration, super-resolution, inpainting, etc. New deep-learning approaches have recently been proposed for volume reconstruction in LFM. However, in LFM, deep-learning methods usually require idealized settings that can be hard to achieve in many practical situations. For instance, they rely on the availability of huge high-quality datasets to achieve good reconstruction performance and generalization ability. In LFM, it is often impossible to capture huge labelled datasets when imaging neural activity in mammalian brain tissue. The neural activity cannot be captured accurately with current 3D imaging techniques since they are not sufficiently fast. Furthermore, the quality of the captured data is often poor since mammalian brain tissue is highly scattering. Even though training with purely synthetic data may alleviate these issues, it requires very accurate simulations of the system and noise, which is often non-trivial.

We aim to study hybrid approaches combining learning and model-based methods that exploit current knowledge of the forward model and use very small training datasets for reconstruction. The acquisition of small datasets can be performed by exploiting multi-modal imaging combining LFM and two-photon microscopy. Two-photon microscopy can be used to acquire structural data offline since this task does not require high acquisition rates. Furthermore, we aim to design the neural network architecture by unfolding sparsity-based reconstruction algorithms. The unfolding technique allows embedding the prior knowledge of the system into the network architecture and better interpretability.

1.2 Original Contribution and Outline of Thesis

The remainder of this thesis is organised as follows:

Chapter 2 reviews basic principles and techniques for imaging neuronal activity in brain tissue. It also describes the idea of LF under three perspectives: ray optics, wave optics, and phase-space optics. Moreover, this chapter explains how the LF camera and the LF microscope work under these three perspectives. Finally, it reviews previous work on 3D reconstruction for LFM, providing context for the following chapters.

Chapter 3 studies the discretization and simplification of the forward model. This chapter proposes a framework for discretization based on generalized sampling theory. Unlike standard methods, the proposed approach allows diverse sampling kernels and different sampling rates for the input and output. Furthermore, it shows that the system can be accurately described using filter banks due to the periodic shift-invariance property. In this chapter, we also propose a simplification method that uses SVD to accelerate the computation of the forward model. We show that it is possible to accelerate the computation without impairing the accuracy of the model. Finally, we propose to describe the system using a linear CNN by adapting our filter-bank model. This novel description allows for efficient implementations of the forward model.

Chapter 4 studies model-based reconstruction techniques for 3D imaging in LFM. First, we exploit the shift-invariance assumption to achieve aliasing-free reconstruction under ideal settings. We experimentally show that signals lying in a shift-invariant subspace can be perfectly reconstructed if the sampling rates and sampling kernels are chosen so that the equivalent discrete system is invertible. Furthermore, we study reconstruction approaches under real acquisition conditions. We propose using optimization approaches incorporating additional priors to perform artifact-free 3D reconstruction from real LF images. Due to the simplification of the forward model, our method is faster than conventional approaches used for 3D reconstruction. We experimentally show that our approach performs better than Richardson-Lucy-based strategies in computation time, image quality, and artifact reduction.

Chapter 5 studies learning-based reconstruction for imaging mammalian brain tissues in LFM. This chapter analyses multi-modal imaging modalities combining 2P and LFM to improve reconstruction quality. First, the architecture of our CNN is obtained by unfolding the Iterative Shrinkage and Thresholding Algorithm (ISTA) and is based on the

observation that neurons in tissue are sparse. Then, we propose a semi-supervised training based on Generative Adversarial Neural Networks (GANs) that exploits the knowledge of the forward model to achieve remarkable reconstruction quality. Furthermore, we propose efficient descriptions of the imaging system using linear convolutional neural networks that fit the physics of the system. These descriptions allow online computing of the forward model, complementing our reconstruction approach. We experimentally show that our method provides better reconstruction quality than state-of-the-art model-based and learning-based methods under adverse conditions: lack of training data, background noise, and non-transparent samples.

Finally, Chapter 6 summarizes the main contributions and discusses possible future work directions.

1.3 Publications

The work developed during my Ph.D. was part of a broader project run jointly with the Department of Bioengineering (BB/R009007/1). The following paragraphs mention all the material I contributed as an author. However, this thesis is mainly based on the papers where I am the first author.

Submitted

- H. Verinaz-Jadan, P. Song, C. L. Howe, A. J. Foust and P. L. Dragotti, "Physics-based Deep Learning for Imaging Neural Activity via Two-photon and Light Field Microscopy", submitted to IEEE Transactions on Computational Imaging, 2022 (bioRxiv).

Peer-Reviewed Journals

- H. Verinaz-Jadan, P. Song, C. L. Howe, A. J. Foust and P. L. Dragotti, "Shift-invariant-subspace discretization and volume reconstruction for light field microscopy", IEEE Transactions on Computational Imaging, 2022.

- Carmel L. Howe, Peter Quicke, Pingfan Song, Herman Verinaz-Jadan, Pier Luigi Dragotti, Amanda J. Foust, "Comparing synthetic refocusing to deconvolution for the

extraction of neuronal calcium transients from light fields," Neurophotonics, March 2022.

- P. Song, H. Verinaz-Jadan, C. L. Howe, A. J. Foust and P. L. Dragotti, "Light-field microscopy for optical imaging of neuronal activity: when model-based methods meet data-driven approaches", IEEE Signal Processing Magazine, March 2022.

- Peter Quicke, Carmel L. Howe, Pingfan Song, Herman Verinaz-Jadan, Chenchen Song, Thomas Knöpfel, Mark Neil, Pier Luigi Dragotti, Simon R. Schultz, Amanda J. Foust, "Subcellular resolution three-dimensional light-field imaging with genetically encoded voltage indicators," Neurophotonics, August 2020.

- P. Song, H. Verinaz-Jadan, C. L. Howe, P. Quicke, A. J. Foust and P. L. Dragotti, "3D Localization for Light-Field Microscopy via Convolutional Sparse Coding on Epipolar Images", in IEEE Transactions on Computational Imaging, 2020.

Peer-Reviewed Conferences

- P. Song, H. Verinaz-Jadan, C. L. Howe, P. Quicke, A. J. Foust and P. Luigi Dragotti, "Model-Inspired Deep Learning for Light-Field Microscopy with Application to Neuron Localization," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.

- H. Verinaz-Jadan, P. Song, C. L. Howe, P. Quicke, A. J. Foust and P. L. Dragotti, "Deep learning for light field microscopy using physics-based models", IEEE International Symposium on Biomedical Imaging (ISBI), 2021.

- H. Verinaz-Jadan, P. Song, C. L. Howe, A. J. Foust and P. L. Dragotti, "Volume Reconstruction for Light Field Microscopy", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020.

- C. L. Howe, P. Quicke, P. Song, H. Verinaz-Jadan, P. L. Dragotti, and A. J. Foust, "Comparing wide field to 3D light field for imaging red calcium transients in mammalian brain," Biophotonics Congress: Biomedical Optics 2020 (Translational, Microscopy, OCT, OTS, BRAIN), 2020.

- P. Quicke, C. L. Howe, P. Song, H. Verinaz-Jadan, P. L. Dragotti, T. Knopfel, A. J. Foust, S. R. Schultz, and M. Neil, "Calculation of high numerical aperture lightfield microscope point spread functions," Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP), 2019.

- P. Song, H. Verinaz-Jadan, P. Quicke, C. L. Howe, A. J. Foust, and P. L. Dragotti, "Location Estimation for Light Field Microscopy based on Convolutional Sparse Coding," in Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP), 2019.

Chapter 2

Background

2.1 Imaging Neural Activity

IMAGING a large population of neurons in the brain and their electrical activity up to cellular resolution or even sub-cellular resolution is necessary to reveal the functioning mechanism of neural circuits [20]. The observation of this electrical activity is known as functional cellular neuroimaging while observing the morphology or anatomy of neural circuits is called structural imaging. There are various viable techniques for functional imaging, such as functional MRI or electrophysiological techniques [21], [20]. However, temporal resolution in functional MRI is limited to the sub-Hz regime and spatial resolution is in the millimeter range [21]. Furthermore, electrophysiological techniques need invasive insertion of electrodes and the number of neurons observed is limited to the electrodes vicinity. On the other hand, optical imaging with fluorescent indicators is more attractive owing to the simultaneous high spatial resolution and non-invasive nature. Optical imaging with fluorescent indicators has, in part, enabled the study of networks of neurons in biological systems [22–24]. However, high-speed 3D functional imaging of a large number of neurons in brain tissue is still challenging for current imaging technology.

Fluorescence is a tool in biological imaging that offers high sensitivity, contrast, and the possibility of obtaining cell type-specific labeling [21]. Typical fluorescent indicators for functional neuroimaging usually respond to changes in membrane potential or detects changes in concentration of calcium ions such as Ca^{2+} [20]. Although calcium indicators monitor neuronal membrane potential indirectly, they are popular because of their slow

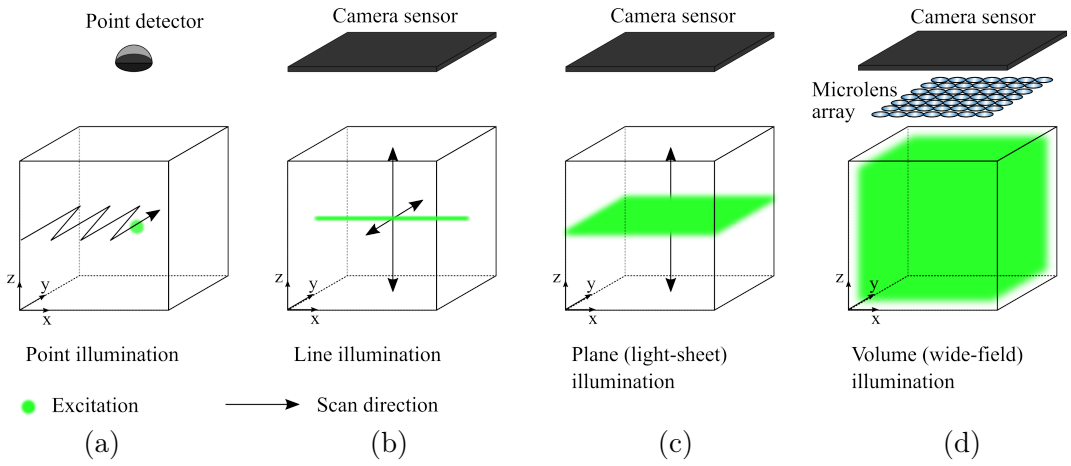


Figure 2.1: The imaging modalities categorized based on the acquisition mode: (a) point, (b) line, (c) plane (light-sheet), and (d) volume (wide-field) illumination. In a sequential acquisition, shown in (a), a focal spot is scanned across one or more dimensions in space to cover the entire volume. Consequently, the location of the signal is determined by the instantaneous position of the excitation beam, and a point detector can be used to collect the emitted fluorescence photons irrespective of the path with which they reach the detector. Such imaging modalities confer robustness to light scattering and are, therefore, well-suited for deep tissue imaging. This comes at the cost of a reduced temporal resolution. In parallel acquisition modes, as shown in (b)-(d), some or all voxels are recorded simultaneously on a camera sensor array, which enables higher volume rates. In particular, light-sheet imaging simultaneously scans the light-sheet excitation plane to quickly build up a volume plane by plane. In this way, it enables the imaging of the whole volume plane by plane. In contrast, LF imaging enables the simultaneous recording of all voxels within a volume. Fluorescence generated throughout the volume is captured through a microlens array to simultaneously encode both the position and angular information. However, unlike point scanning modalities, parallel acquisition modes are inherently vulnerable to scatter-induced crosstalk between neighboring camera pixels. See also [21] for a description of different imaging modalities.

transients and high signal amplitude, which makes them easier to detect than membrane potential indicators. However, recent developments show that genetically encoded voltage indicators (GEVIs), which respond to membrane potential changes, also efficiently monitor neural activity [25]. Even though approaches, such as detecting sodium or potassium ions, or pH, are possible, they produce signals many orders below the sensitivity required for the study of entire networks of neurons [20].

Calcium indicators can be used with diffraction-limited point scanning techniques and single-pixel detectors to identify the precise location of activity-related signals [20]. In particular, two-photon microscopy is an imaging approach robust to scattering, with a small point spread function, and is well suited for imaging deep in tissue. Even though two-photon microscopy has the potential to facilitate the understanding of neural circuits,

it results in a low temporal resolution due to the point scanning nature. In two-photon microscopy, a point-like laser scans the whole volume, and the emitted fluorescence is collected with a point-like detector, as shown in Figure 2.1 (a). Two-photon microscopy offers appealing tissue penetration and spatial resolution, but its temporal resolution is ill-suited to capture fast biological dynamics on a large population of neurons.

The study of neural information coding requires high-temporal-resolution imaging methods. In recent years, we have witnessed substantial progress in fluorescence microscopy imaging due to improved fluorescent indicators of neuronal activity and alternative scanning strategies for faster acquisition [21], [20]. For example, scanning with lines or sheets instead of points speeds up the acquisition through spatial parallelization. These scanning schemes are depicted in Figure 2.1 (b) and (c). However, parallel scanning usually implies constrained sample geometries or is not well-suited for scattering samples.

Scanning-based microscopy is proposed as an alternative to scan-less whole-volume microscopy (called wide-field). Wide-field imaging is one of the most basic microscopy techniques. It uses the most efficient illumination scheme, maximizing photon generation rates and imaging speed. However, this speed comes at the cost of increased background interference since out-of-focus light appears in the in-focus image, reducing contrast. Thus, typical wide-field microscopy is unsuitable for efficient imaging of neuron activity. Nevertheless, it is possible to take advantage of the out-of-focus fluorescence and reassign photons to their correct 3D locations computationally, as depicted in Figure 2.1 (d). Light field microscopy (LFM) is a technique that exploits this idea by using a point spread function (PSF) designed to promote diversity between impulse responses located at different depths.

LFM is a wide-field imaging technique that has become an attractive candidate for high-speed 3D bioimaging. In LFM, a microlens array (MLA) is inserted at the native image plane (NIP) of a wide-field microscope to capture a 4D LF (including the 2D lateral position and 2D angular information), as shown in Figure 2.2(a). The angular information, in turn, relays depth information for volumetric reconstruction. Figure 2.2 (b),(c) and (d) shows how the impulse response changes as the axial position of the light source changes. A more detailed description of this behaviour is given in the following sections. In this way, LFM can capture volumetric information of the scene in a single snapshot, allowing

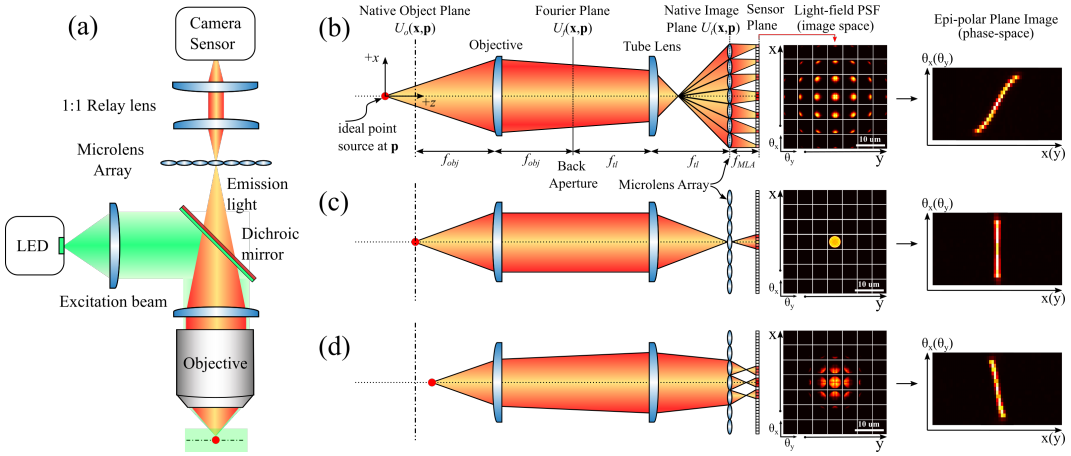
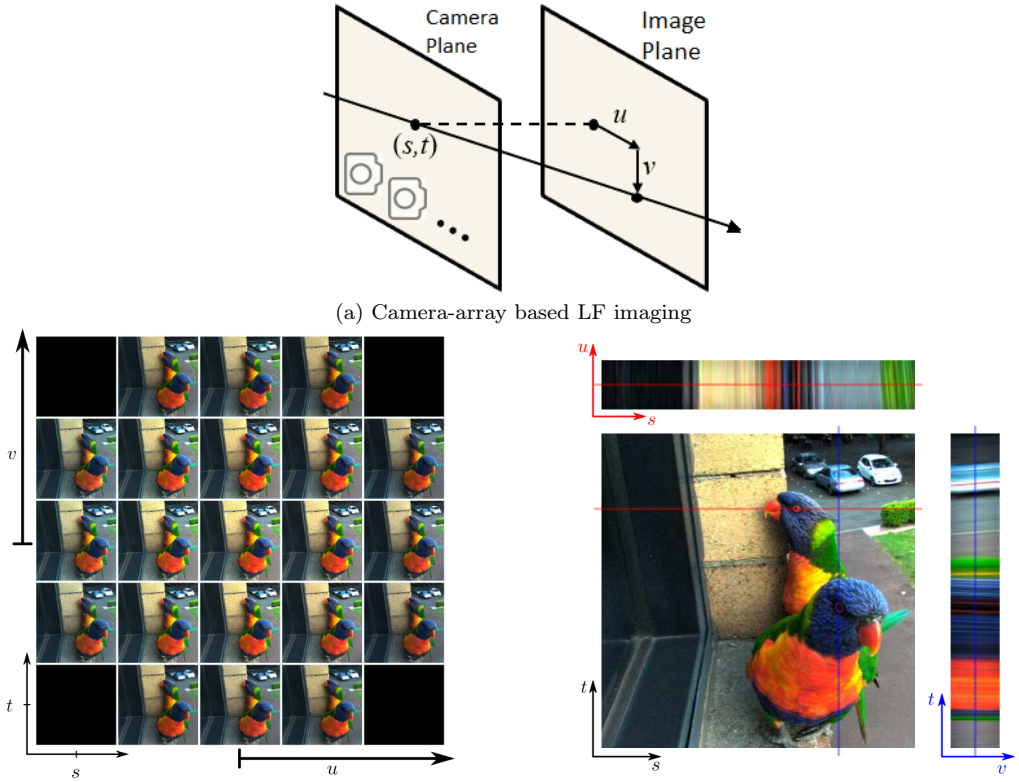


Figure 2.2: An optical diagram of LFM. (b)-(d) The LFM optical path and LF patterns generated by one point source at different depths (i.e., axial positions). A point source below or above the native object plane, as shown in (b) and (d), generates a more complicated intensity pattern than a source in focus, as shown in (c). White lines in the LFM images depict the virtual profile of the microlens in the image space, and each square represents a subimage associated with a specific lenslet. LFM simultaneously captures spatial and angular information, which can be revealed in an epipolar plane image. A deeper source leads to a more tilted line, whose center indicates the lateral position.

3D imaging at video frame rates. Various works have explored LFM for imaging neuronal circuits. For instance, whole-brain calcium imaging in small organisms like C. elegans and zebrafish larvae [13], [26], [27] or large-volume functional recording in awake, behaving mice [17] have been demonstrated. However, a limitation of LFM is the trade-off between spatial and angular resolution, computational complexity, and reconstruction artifacts. Furthermore, deep in the tissue, scattering events may lead to a complete loss of directional information from emitted fluorescence [21].

2.2 The Concept of LF and LF Devices

LF is a function that maps each ray in the space to a non-negative value called radiance. As mentioned by Ren Ng et al. in [30], the modern concept of LF was first introduced by Levoy and Hanrahan [31], and by Gortler et al. [32] in 1996. However, these ideas come from even earlier work. The pioneer was Lippman in 1908 with his contribution to the concept of "Integral Photography". Even though the LF is traditionally described with ray optics, it can be studied with different levels of detail. This section describes the LF from the ray optics, scalar wave optics, and phase-space optics perspectives.



(a) Camera-array based LF imaging

(b) An array of multi-view images.

(c) EPIs in $u-s$ and $v-t$ space.

Figure 2.3: Illustration of LF imaging. (a) In a LF system, a light ray that propagates from the scene is uniquely determined by the intersections with two parallel planes, leading to a relative two-plane parametrization of the LF. The $s-t$ plane is usually related to the spatial information of the rays, and the $u-v$ plane encodes the angular information of the rays. A pinhole camera placed in the camera plane captures light rays with directions determined by the image plane inside the camera. The collection of captured images in (b) shows the multi-view images or sub-aperture images obtained when the angles $u-v$ and location $s-t$ change. The variables $s-t$ index the pixel of each sub-aperture image, while $u-v$ index the view angle. (c) A EPI (up) in $u-s$ space for fixed (v,t) and an EPI (right) in $v-t$ space for fixed (u,s). (Images are modified from [28, 29].)

2.2.1 Ray-Optics Perspective

LF imaging attempts to record the intensity of every light ray from a scene of interest. A simple method to achieve this goal is to place a pinhole camera at every position (s,t) in the camera plane. A pixel (u,v) in the image plane of the pinhole camera indexes the angle of arrival of a light ray, while the camera position (s,t) indexes its spatial location. See Figure 2.3 (a) for clarification. The set of captured images is the observable LF. It can be reordered into a set of sub-aperture images as shown in Figure 2.3 (b) or forming epipolar plane images (EPIs), as in part (c). These concepts are clarified in the following paragraphs.

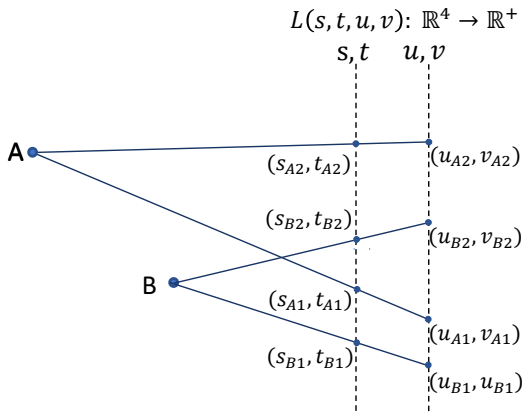


Figure 2.4: LF definition. The LF is a function that maps every ray in the space to a non-negative value (radiance). Using the two-plane parametrization, each ray in the 3D space can be described with four scalars (s, t, u, v). The two-plane parametrization exploits the observation that a ray in space intersects two parallel planes in two points. The four scalars defining the two points of intersection are the parameters that define a ray.

Since the LF maps a light ray to a positive number, it can be represented by a function $L(s, t, u, v) : \mathbb{R}^4 \mapsto \mathbb{R}^+$. The four scalars (s, t, u, v) uniquely represent a light ray using the two-plane parameterization, as shown in Figure 2.3 (a). Several assumptions must hold to make the definition of LF consistent. The LF must be monochromatic to remove the wavelength dependence, the temporal dimension is removed by integration over time, and rays propagate through a vacuum in a space without any external object such that the radiance is constant along the ray. The latter statement is called free-space assumption. Also, diffraction effects should be negligible. An exhaustive explanation of these concepts can be found in [33], [31].

Ideally, the knowledge of the LF allows for finding the location of every light source in the 3D space. For instance, in Figure 2.4, from the intersections defined by s_{A1} and s_{A2} with directions defined by u_{A1} and u_{A2}, it is possible to retrieve the position of the point source A in the 2D space. Similarly, the point source B located at a different position emits a different set of rays, say s_{B1} and s_{B2} with directions defined by u_{B1} and u_{B2}. By considering the four coordinates (s, t, u, v) for each ray, one can localize point sources in the 3D space. Therefore, the LF is a tool that can be potentially used for reconstructing 3D scenes.

The LF is inherently redundant. The LF corresponding to the rays emitted from a

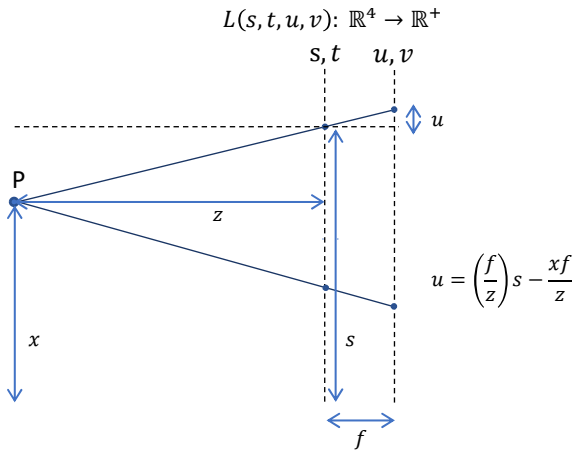


Figure 2.5: LF geometry for a point source. This figure shows the geometric relationship satisfied by all the rays coming from a given point source. The two-plane parametrization of the rays shows that the LF of a single point source is nonzero on a 2D plane immersed in the 4D space.

single point source in the 3D space is non-zero in a 2D plane immersed in the 4D space. To see this, note that from the diagram shown in Figure 2.5 and using triangle similarities, it can be written that:

$$\begin{aligned} u &= \left(\frac{f}{z}\right) s - \frac{xf}{z} \\ v &= \left(\frac{f}{z}\right) t - \frac{yf}{z}, \end{aligned} \tag{2.1}$$

where positions u and v are measured with respect to the s and t locations, respectively. Measuring the relative positions u and v allows a simple interpretation of the LF, which will become clear in the following lines. For a given point source at (x, y, z), the LF $L(s, t, u, v)$ is nonzero on the 2D plane defined by Equations (2.1), and zero elsewhere. Note that if the parameter f is known, the slope in Equation (2.1) defines the z-coordinate of the point source, and the intercepts define the x-y location. To distinguish between spatial (s, t) and angular (u, v) coordinates, we can redefine u and v as $u := u/f$ and $v := v/f$, which are the tangents of the arrival angle and do not depend on the plane separation.

The LF carries spatial and angular information from light. If we compute the LF for fixed angular coordinates (u, v), we obtain an image $L_{u,v}(s, t)$ containing only spatial information. This image shows a view of the scene (all the rays arrive with the same angle) and is known as a sub-aperture image, as shown in Figure 2.3 (b). Similarly, when the

coordinates (s, u) are fixed, the obtained image $L_{s,u}(t, v)$ is named epipolar plane image (EPI) and it is relevant in many applications since it carries depth information of the scene [34–37]. See Figure 2.3 (c) for clarification.

LF Camera

A LF camera records ray intensities with their spatial location and direction of arrival simultaneously. In contrast, a standard camera records only the intensity and location of the rays reaching the sensor. Thus, a LF camera allows for digital refocusing, for reconstruction of super-resolved images, and also for changing digitally the aperture [28], [38].

Various forms to capture the LF function have been proposed over the years. As mentioned in the previous section, one can use an array with multiple cameras or a moving camera to capture the LF. However, these implementations are impractical and challenging due to synchronization problems. It is more efficient to record the LF with a single device that captures multiple views of the scene in a single snapshot [39] such as the array of mirrors camera [40], the coded aperture imaging technique [41], or the microlens-based camera. The latter is the most widespread LF device since it is very effective and simple to implement. Ren Ng et al. reported the first successful implementation of a modern LF camera using MLAs in [30]. Now, this camera design is known as plenoptic 1.0. Later, other designs, such as the plenoptic 2.0 [42], were also proposed.

A microlens-based LF camera is implemented by interposing a MLA between the pixel sensor and the main lens of a standard camera. The distance between the MLA and the sensor equals the microlens focal length in the plenoptic 1.0. In the plenoptic 2.0, this distance is not necessarily the focal length. Contrary to the plenoptic 1.0, the plenoptic 2.0 configuration ensures that the plane in focus at the image sensor is not located at infinity. Thus, regions near the in-focus plane can be captured with a higher resolution than in the original plenoptic 1.0. However, when the spatial resolution over a wide depth range is evaluated, the original plenoptic 1.0 configuration gives better average performance [43]. It has been found that the plenoptic 1.0 creates a more diverse point spread function (PSF) between adjacent points than the plenoptic 2.0 design [43]. In this work, we focus on the plenoptic 1.0 configuration since it is more common in LFM. A diagram of a plenoptic 1.0 camera is depicted in Figure 2.6.

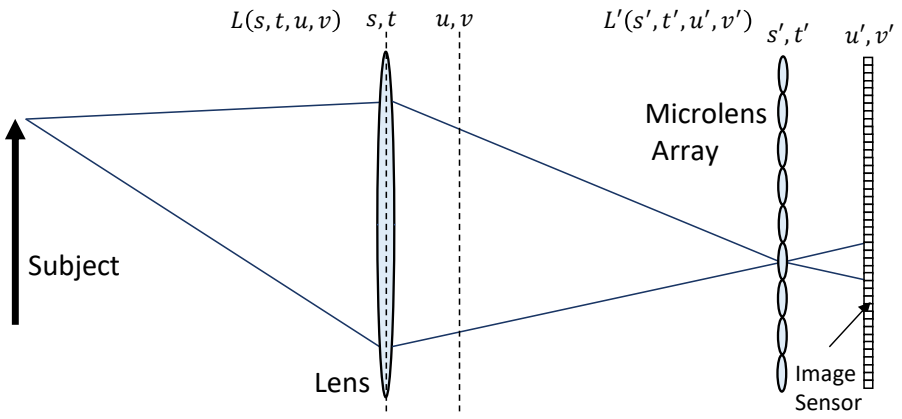


Figure 2.6: LF camera diagram. A microlens-based LF camera is designed by interposing a MLA between the main lens and the image sensor of a standard camera. Each microlens in the camera tells the location of the light ray, while the pixel behind the microlens tells the direction of arrival of the ray. If the the main lens and microlenses are matched [30], the image at the image sensor can be interpreted as a sample version of the LF $L'(s',t',u',v')$.

The idea of LF allows a simple interpretation of the behaviour of the microlens-based camera. For simplicity, we consider that the microlens plane corresponds to the $s-t$ plane mentioned in previous section, and the sensor plane corresponds to the $u-v$ plane. They are re-named $s'-t'$ and $u'-v'$, respectively, as shown in Figure 2.6. Similarly, the LF inside the camera is named $L'(s',t',u',v')$ to distinguish it from the external LF. Suppose each microlens is modelled as a pinhole and each pixel sensor is a point-like detector [28]. In that case, one can find that the indexation of the microlenses tells the location of the ray, while the indexation of the pixels tells the direction of arrival of the ray. Thus, the MLA together with the sensor pixels act as a sampling device of the function $L'(s',t',u',v')$. For this to hold, we assume that rays from adjacent microlenses do not overlap, which is ensured by matching the main lens and microlenses [30]. For a plenoptic 1.0 configuration, the final output of the camera is defined as follows:

$$L[m,n,p,q] = L'(mT_s, nT_t, pT_u, qT_v), \tag{2.2}$$

where T_s and T_t are the horizontal and vertical microlens pitch, which are usually the same $T_s = T_t = T$, the angular sampling intervals are $T_u = T_v = T/(Nf_\mu)$, T is the microlens pitch, N is the one-dimensional number of pixels under the microlens, and f_μ is

the microlens focal length. Note that since the raw output of the camera is a 2D image, one must first calibrate and reshape the 2D array into a 4D structure to properly recover $L[m, n, p, q]$, as explained in [29].

The internal LF $L'(s', t', u', v')$ is not the same as the external LF $L(s, t, u, v)$ due to the optical system of the camera. This system, usually modelled with a single lens, aims to bring to focus any object of interest but also limits the spatial and angular information of the LF. For simplicity, the internal and external LFs are usually related with a linear transformation as follows:

$$L(s, t, u, v) = L'((s, t, u, v)\mathbf{M}), \tag{2.3}$$

where $L(s, t, u, v)$ is the external LF, and the transformation matrix $\mathbf{M}$ models both propagation and lens effect [44]. See Figure 2.6.

A single lens cannot describe the optical system of a LF microscope. Unlike a LF camera, a LF microscope uses a 4f-system with a telecentric stop, which differs from the optical system used for cameras (see Appendix 6.2.4 for a description of the 4f-system). A LF microscope is an orthographic system, as explained in [45]. This optical design implies that, under ideal conditions, the internal LF $L'(s', t', u', v')$ only differs from $L(s, t, u, v)$ due to the magnification factor as follows:

$$L(s, t, u, v) = L'(sM, tM, u/M, v/M) \tag{2.4}$$

where M is a scalar representing the magnification factor of the microscope. In this thesis, we are interested in the behaviour of the LF microscope, which is described in more detail in the following sections.

Sampling Pattern and Kernel in Microlens-Array Based Devices

In the previous section, each microlens in the array is a pinhole, and each pixel in the sensor is a point-like detector. Under these assumptions, and considering only one spatial and one angular dimension for simplicity, the predicted sampling pattern is a 2D comb function, shown with red dots in Figure 2.7 (a). However, ray optics allows for a more accurate description of the system. Suppose each pixel is rectangular, each microlens is

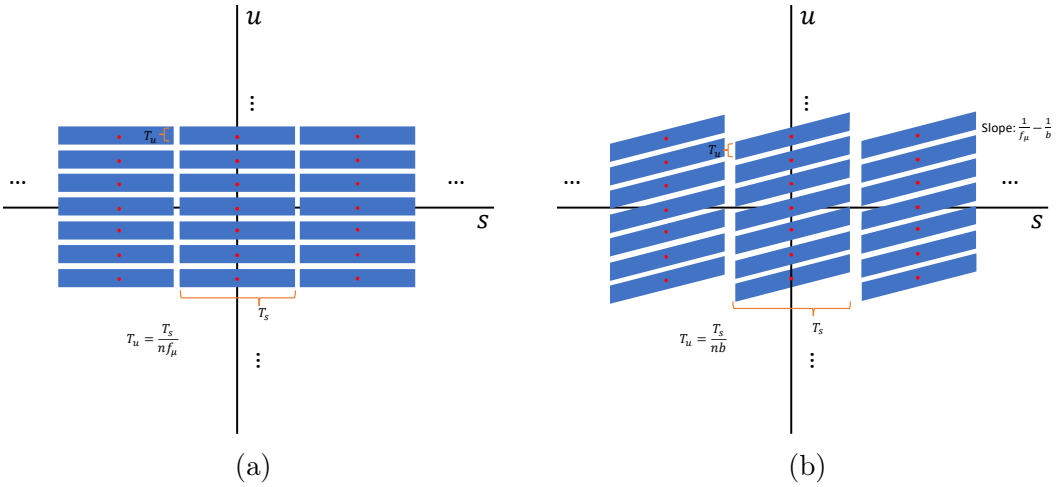


Figure 2.7: Sampling pattern and kernel. The microlens array, followed by propagation and a camera sensor, can be a sampling device for the LF function. In part (a), we show the sampling pattern and kernel of a plenoptic 1.0 configuration. If each microlens is a pinhole and each pixel in the sensor is a point-like detector, the predicted sampling pattern is a 2D comb function shown in (a) with red dots. If each pixel is a rectangular sensor and each microlens is a thin lens. The sampling kernel is a rectangle shown in blue in part (a). In part (b), we show the sampling pattern and kernel of the plenoptic 2.0. Here, the distance between the microlens array and the sensor is not the same as the focal length of the microlenses. Thus the shape of the sampling kernel changes to a sheared rectangle. In this figure, n is the one-dimensional number of pixels under each microlens, $T_s = T$ is the microlens pitch, f_μ is the microlens focal length and b is the distance between the microlens array and the sensor.

a rectangular thin lens, and that light from adjacent microlenses does not overlap. In that case, the predicted sampling kernel is rectangular, with dimensions defined by the geometry of the camera sensor and MLA, as specified in Figure 2.7 (a). If the distance between the MLA and the sensor is not the same as the focal length of the microlenses, as in the plenoptic 2.0 configuration, the shape of the sampling kernel changes to a sheared rectangle, as shown in Figure 2.7 (b). For more details, see [46].

If rays from adjacent microlenses do not overlap, a microlens array followed by propagation and a camera sensor is a sampling device for the LF function. The geometry of the elements in the arrangement, such as the focal length, the microlens pitch, the sensor pixel pitch, and the separation between the MLA and the sensor, determines the device sampling pattern and sampling kernel. For instance, the plenoptic 1.0 configuration forms a rectangular kernel, and the plenoptic 2.0 yields a rectangle with slanted sides, as shown in Figure 2.7. Also, note that a rectangular packaged MLA and pixels sensor always form

a rectangular-like sampling grid. A different MLA and pixel sensor package can create other sampling patterns. For instance, one can use hexagonal packaged MLAs described in Appendix 6.2.4. In this thesis, we use a given plenoptic 1.0 LF microscope with a square shaped and square packaged MLA and sensor.

2.2.2 Scalar-Wave-Optics Perspective

The description of light based on the ray model is considered insufficient to describe a microscope since it ignores diffraction, which is a dominant effect on a microscopic scale. Thus, one can study the complex field directly instead of studying the 4D LF mentioned in the previous section. Therefore, microscopes are usually described using scalar or vector wave optics depending on the scenario [47]. Scalar wave optics does not take into account the vector nature of light. It assumes that the electromagnetic field can be described by a single time-dependent scalar field, namely $u(t)$. Thus, instead of using the vector wave equations for the electric field and magnetic field derived from Maxwell equations [48], it proposes the use of a single scalar wave equation, as follows:

$$\nabla^2 u - \frac{n^2}{c^2}\frac{\partial u}{\partial t^2} = 0, \tag{2.5}$$

where n is the refractive index, c is the speed of light and u is the time-dependent scalar field. The use of a single scalar differential equation is justified by making some assumptions on the surrounding media [48], [49]. In general, the medium needs to be dielectric (real permittivity, normally the case for visible light), it needs to be a linear material (a linear relationship between electric field and displacement field), to be homogeneous (constant permittivity across the region of propagation), to be isotropic (its behavior does not depend on the field polarization), to be non-dispersive (permittivity independent of frequency) and non-magnetic (vacuum permeability).

The Helmholtz equation is an equivalent version of the wave equation (2.5). If we focus on time-harmonic fields, we can assume $u = Re\{Ue^{-j\omega t}\}$ in Equation (2.5). This procedure results in the following equation:

$$(\nabla^2 + k^2)U = 0, \tag{2.6}$$

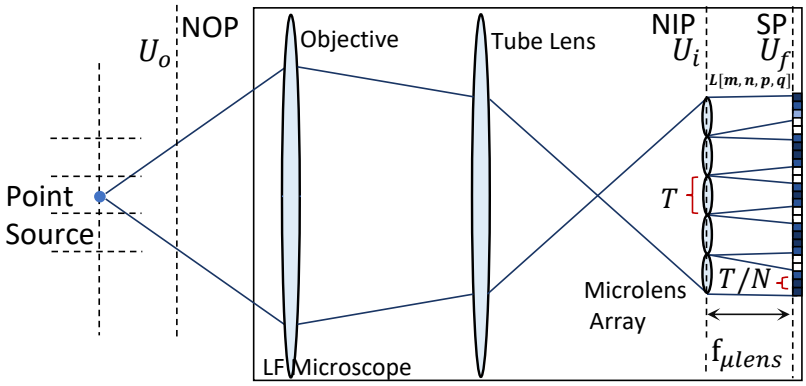


Figure 2.8: LF microscope diagram. A LF microscope captures spatial and angular information from the light by interposing a MLA between the tube lens and the camera sensor. The objective and tube lens form a 4F system that magnifies in the native image plane (NIP) the field at the native object plane (NOP). The shown LF microscope corresponds to the plenoptic 1.0 configuration, where the distance between the MLA and the sensor plane is the focal length of the microlenses. Contrary to the LF camera, the optical system is orthographic by design, and diffraction effects are dominant due to the size of the samples.

where the angular wavenumber is $k = n\omega/c$, c is the speed of light in vacuum, and n is the refractive index of the media. This equality is known as the Helmholtz equation and is a basic equation behind wave optics. Notice that, in general, the Helmholtz equation holds for any scalar field u with a valid Fourier transform U since the Helmholtz equation can be derived by taking the Fourier transform in the temporal dimension of both sides of Equation (2.5).

The Helmholtz equation is not used directly to describe optical devices. Instead, various simplifications and more practical concepts are exploited. For instance, Rayleigh-Sommerfeld solution, Fresnel-Kirchoff diffraction, paraxial approximations, or the concept of angular spectrum are used to describe propagation. Furthermore, simple optical devices such as lenses or an array of lenses can be defined conveniently with a complex function named transmittance. See Appendix (6.2.4) for a more precise description of these approximations. Altogether, these concepts allow describing complex devices such as microscopes, as explained in the following paragraphs.

LF microscope

Optical configurations designed for LF photography can be adapted to microscopy. In [45], Levoy et al. first proposed this idea by implementing the plenoptic 1.0 configuration in a standard microscope: A MLA is placed at the native imaging plane of the 4F system with a camera placed at the back focal plane of the MLA, as shown in Figure 2.8.

Wave optics accurately describes the functioning of a LF microscope. Specifically, the system behaviour can be divided into four processes: initial propagation from the source location to the native image plane (NIP), modulation due to the MLA, propagation to the image sensor plane, and discretization due to the sensor pixels. See the complete optical path in Figure 2.8.

The initial propagation from the source location to the NIP is greatly affected by diffraction. The LF just after the tube lens can be approximated from the far field of a point source (dipole) as follows [48, 50]:

$$U(P_1) = P(P_1)\frac{exp(jkf)}{f}, \tag{2.7}$$

where $P(P_1)$ is called apodization function, P_1 is a point of interest, and f is the focal length of the lens. The apodization function is introduced to model the field distribution on the spherical wavefront just after the lens. It also models the finite extent of the device, which, together with the sample sizes, greatly contributes to observing diffraction effects. Then, the field at the NIP can be found using the first Rayleigh-Sommerfeld solution (Appendix 2 explains the first and second Rayleigh-Sommerfeld solutions). The first Rayleigh-Sommerfeld solution leads directly to the desired field since the field in the boundary is known. Note that the surface causing diffraction is a lens and the field in the boundary is the field measured just after the lens, shown in Equation 2.7. The resulting integral is simplified assuming a circular lens and using the Debye approximations leading to the Debye integral, as mentioned in [50], [51]. Finally, considering that the 4f system magnifies in the NIP the field at the NOP by a factor M, the field at the native image plane at a point $\mathbf{x} = (x_1, x_2)$, due to a source located at $\mathbf{p} = (p_1, p_2, p_3)$ can be written as:

$$U_i(\mathbf{x}, \mathbf{p}) \propto \int_0^{\alpha} \exp\Big(jkp_3\cos(\theta)\Big)P(\theta)J_0\Big(kr\sin(\theta)\Big)sin(\theta)d\theta, \tag{2.8}$$

where $P(\theta) = \sqrt{\cos(\theta)}$, $r = \sqrt{(x_1/M - p_1)^2 + (x_2/M - p_2)^2}$, $k = \frac{2n\pi}{\lambda}$ and $\alpha = \sin^{-1}\left(\frac{NA}{n}\right)$. Here, the function $P(\theta)$ is the apodization, N is the numerical aperture, n is the refractive index, λ is the wavelength, and M is the microscope magnification. Notice that this is a proportional relationship because any constant factor can be safely ignored in the image formation process since we are interested in the power of the final field.

The next step is multiplying the complex field $U_i(\mathbf{x}, \mathbf{p})$ by the transmittance $\Phi(\mathbf{x})$ of the rectangular MLA and propagating the field to the sensor. The propagation is a linear shift-invariant process with transfer function $G(k_x, k_y)$ (see Appendix 6.2.4 for more details on the propagation transfer function and lens transmittance). Therefore, the field at the image sensor plane can be written as follows:

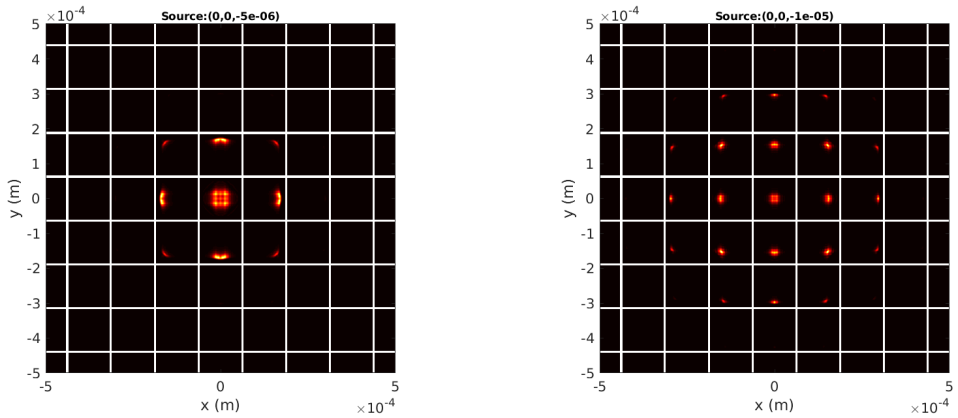
$$U_f(\mathbf{x}, \mathbf{p}) = \mathfrak{F}^{-1}\left\{\mathfrak{F}\left\{U_i(\mathbf{x}, \mathbf{p})\Phi(\mathbf{x})\right\} G(k_x, k_y)\right\}, \tag{2.9}$$

where $\mathfrak{F}$ is the 2D Fourier transform. Finally, the camera sensor captures the light intensity, which is the square magnitude of the field in Equation (2.9). This intensity is named $h(\mathbf{x}, \mathbf{p})$ and is computed as follows:

$$h(\mathbf{x}, \mathbf{p}) = |U_f(\mathbf{x}, \mathbf{p})|^2, \tag{2.10}$$

where the function $h(\mathbf{x}, \mathbf{p})$ is also the impulse response of the continuous system describing the microscope. The output of this continuous system still needs to be discretized by the pixels in the image sensor to obtain the observed LF image. Thus, the final step is modelled as a convolution with a square kernel $k(\mathbf{x})$ describing the pixel shape, followed by sampling with sampling interval $\mathbf{\Delta x}$ equal to the microlens pitch.

In this work, we consider that there is no coherent interference when light from two emitters interacts. The complex field from two different sources cannot be linearly combined. However, we can safely assume that the map from the source distribution to the light intensity measured in the sensor is linear [51]. Therefore, the LF microscope can be entirely characterized by its impulse response. In Figure 2.9, we show the simulated impulse response of the LF microscope corresponding to two different locations of the input. In a real scenario, additional factors, like occlusion or non-constant refractive indexes, could



(a) PSF at depth of 5 µm.

(b) PSF at depth of 10 µm.

Figure 2.9: Simulated LF for ideal point sources located at two different depths. The observed pattern becomes wider as the point source moves away from the native object plane. LFM exploits the diversity of the observed pattern to reconstruct points at different 3D locations from a single 2D image. The square grid shown in white represents the MLA.

make the system nonlinear. However, we found that nonlinear effects can be safely ignored for the applications discussed in this work.

2.2.3 Phase-Space-Optics Perspective

Phase-space optics allows combining wave and ray optics description of the LF into a single concept. The phase-space analysis is interesting from a theoretical point of view since it gives a justification to apply ideas from ray optics to wave phenomena. Even though the ray optics and wave optics perspectives are enough to follow the remaining of this thesis, we show these ideas for completeness.

Phase-space optics refers to the study of optical signals as a function of both the space and the spatial frequency. In optics, the spatial frequency of the signal also carries information about the direction of propagation of the wave. Thus, phase-space optics is useful for simultaneously analyzing spatial and angular information in optical phenomena [52]. This section focuses on studying LF systems using the Wigner Distribution Function (WDF).

WDF

The WDF is a transformation that allows studying a field as a function of the space and its Fourier conjugate variable simultaneously. The WDF of any complex scalar field $U_1(x, y)$

measured on an x-y plane is defined as follows:

$$W_{U_1}(x, y, f_x, f_y) = \iint U_1(x + \frac{x'}{2}, y + \frac{y'}{2}) U_1^*(x - \frac{x'}{2}, y - \frac{y'}{2}) e^{-j2\pi(f_x x' + f_y y')} dx' dy'. \quad (2.11)$$

The definition of WDF shown in Equation(2.11) requires the use of coherent fields, but the concept of WDF can be extended to partially coherent fields using a stochastic description [53]. As mentioned in Section 2.2.2, the field from two different sources is assumed to be incoherent in this work. However, the field from a single point source is coherent with itself. Therefore, the transformation shown in Equation (2.11) is still valid in LFM.

In signal processing, the short-time Fourier transform (STFT) is a well-known transformation used to study a function with respect to the original dimension and its Fourier conjugate simultaneously. Interestingly, the WDF can be related to the STFT [54]. The definition of the STFT of $U_1(x, y)$ is:

$$STFT\{U_1\}(x, y, f_x, f_y) = \iint U_1(x', y') a(x' - x, y' - y) e^{-j2\pi(f_x x' + f_y y')} dx' dy', \quad (2.12)$$

where $a(x, y)$ is a sliding window. The square magnitude of the STFT (the spectrogram) is related to the WDF by a convolution operation as follows:

$$|STFT\{U_1\}(x, y, f_x, f_y)|^2 = W_{U_1}(x, y, f_x, f_y) * W_a(x, y, f_x, f_y), \quad (2.13)$$

where $W_a(\cdot)$ is the WDF of the sliding window $a(x, y)$ in Equation(2.12).

Even though the STFT is more common in signal processing, the WDF has more appealing characteristics in optics. For instance, many typical operations in wave optics involving integrals, like propagation or multiplication by transmittances like lens modulation, can be reduced to simple shearing of the WDF [55]. Various useful WDF properties are listed in [56]. In addition, integrating the WDF along the frequency axis gives the intensity of the field at the depth the WDF was measured, while the integration along the spatial axis gives the power spectrum for that depth [56], [55]. Moreover, the WDF of a point source is a 2D plane immersed in the 4D space, like the LF of a point source [52]. Therefore, the WDF can be interpreted as an extension of the concept of LF used in ray

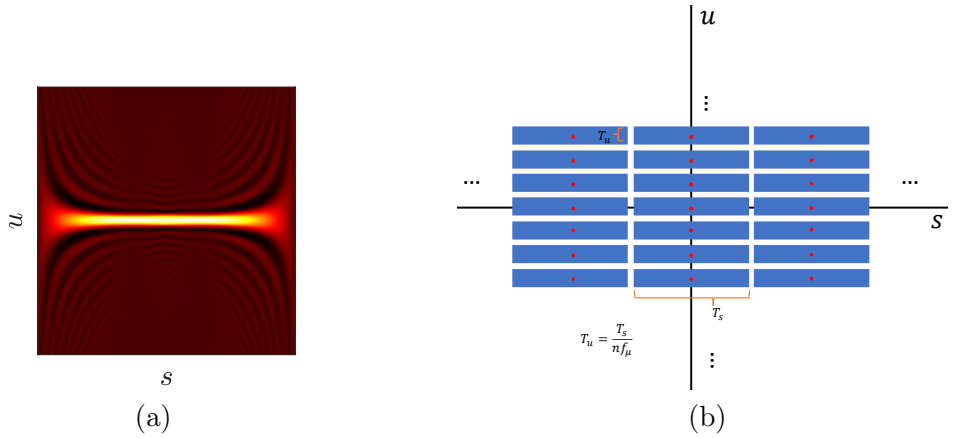


Figure 2.10: Sampling kernel predicted by phase-space optics. Part (a) shows the sampling kernel predicted for rectangular shaped-microlenses and rectangular-shaped pixels using phase-space optics. Ray optics uses a simplified model that predicts a rectangular-shaped kernel shown in blue in (b), which approximates the kernel in part (a). The sampling pattern predicted by phase-space and ray optics coincide and is shown with red dots in part (b). The shown figures correspond to a plenoptic 1.0 configuration.

optics.

Capturing the WDF

The use of the WDF as a tool to describe a LF camera was proposed by Levoy et al. in [57]. Since the WDF is a generalization of the concept of LF, it allows a simple description of a LF camera or microscope. By following an analysis similar to [57], and using the WDF properties described in [55], it is possible to show that the relationship between the WDF at the native image plane $W_{U_i}(s,t,u,v)$ and the field captured at the sensor plane $L[m,n,p,q]$ is defined as follows:

$$L[m,n,p,q] = L(mT_s, nT_t, pT_u, qT_v), \tag{2.14}$$

where $L(s,t,u,v) = W_{U_i}(s,t,u,v) \circledast K(s,t,u,v)$, $\circledast$ means convolution, the spatial sampling intervals are $T_s = T_t = T$, T is the microlens pitch, and the angular sampling intervals T_u and T_v are related to the pixel pitch as follows $T_u = T_v = \frac{T}{N\lambda f_\mu}$, where λ is the wavelength, f_μ is the microlens focal length and N is the one-dimensional number of pixels under each microlens. The 4D discrete signal, $L[m,n,p,q]$, is just a rearrangement of the pixels of the 2D image capture at the sensor, as mentioned in Section (2.2.1). For Equation 2.14 to

hold, it is necessary that images behind each microlens do not overlap, which is ensured if the microlens and objective numerical apertures are matched by design [30], [45]. Due to matched numerical apertures, it also holds that $T_u = T_v = \frac{2NA}{M\lambda N}$, where NA is the numerical aperture of the microscope and M is the magnification factor. Equation (2.14) is derived for a plenoptic 1.0 LF microscope with squared grid MLA and sensor. The function $K(s,t,u,v)$ is a kernel defined by the MLA and the sensor grid. If the shape of each microlens and pixel sensor is known, the kernel $K(s,t,u,v)$ is computed as follows:

$$K(s,t,u,v) = W_T(-s,-t,u,v) \circledast P(u,v), \tag{2.15}$$

where the function $W_T(\cdot)$ is the WDF of the pupil function $T(x,y)$, which is an indicator function that is one inside the microlens and zero elsewhere, and the function $P(u,v)$ is an indicator function for the pixels in the camera sensor. Note that the 4D kernel $K(s,t,u,v)$ does not depend on the microlens focal length but on the shape of the microlens $T(x,y)$ and pixels $P(u,v)$ due to the plenoptic 1.0 configuration. See the complete derivation in Appendix (6.2.4).

The MLA followed by propagation, and a camera sensor, can be interpreted as a sampling device of the 4D continuous WDF. For this to hold, light from adjacent microlenses must not overlap. As mentioned in Section 2.2.1, the geometry and optical parameters of the configuration determine the shape of the kernel and sampling pattern. For instance, the sampling kernel for rectangular microlens and rectangular sensor pixels is depicted in Figure 2.10 (a). Observe that the shown kernel is a more accurate description of the sampling kernel predicted using ray optics in Section 2.2.1, and again depicted in Figure 2.10 (b). Specifically, the sampling kernel shown in Figure 2.10 (a) corresponds to the rectangular kernel shown in blue in part (b). The predicted sampling grid still coincides with the one predicted by ray optics, and it is shown with red dots in Figure 2.10 (b).

The system that goes from the point source to just before the MLA is a wide-field (WF) microscope. Therefore, a LF microscope can be modelled as a standard microscope followed by a WDF sampling device. See Figure 2.11. This observation was used by Levoy et al in [45]. They propose to use the properties of the WDF for reconstruction: A single slice of a focal stack is obtained by summation of $L[m,n,p,q]$ along the frequency

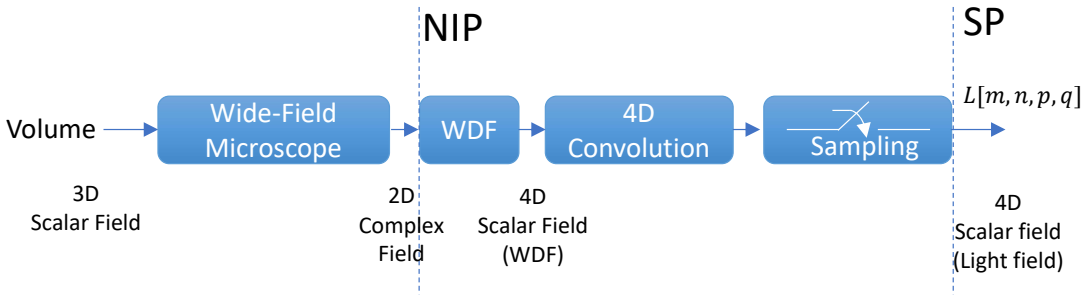


Figure 2.11: Block diagram of a LF microscope based on the WDF. The LFM system can be interpreted as a LF microscope followed by a 4D WDF sampling device. Thus, the properties of the WDF can be exploited for the reconstruction of 3D images. This observation was exploited by Levoy et al in [45].

dimensions. Then, $L[m,n,p,q]$ is sheared so that the WDF is focused at a different depth. This process is performed repeatedly (in the Fourier domain) to obtain a blurred focal stack, which is finally deconvolved using a standard 3D deconvolution approach.

Although the phase-space formulation is conceptually interesting, this description has hardly been exploited for reconstructing high-resolution 3D volumes [58]. A reconstruction based on the phase-space description would need to manipulate high-resolution 4D WDF, which implies the need for high computational resources.

2.3 Computational Methods for 3D Reconstruction from LFM Data

LFM is a technique that relies on computational approaches to perform the reconstruction of 3D volumes. Computational methods in LFM differ from those in photography because they must consider wave optics to describe the system accurately. In contrast, the ray-optics model is usually accurate enough in photography.

Reconstruction approaches for LFM are based on both deep learning and model-based strategies. Deep-learning methods for LFM are faster and perform better than classic model-based approaches if they are evaluated in controlled scenarios, e.g., huge training datasets, low background noise, non-scattering media, and transparent samples. On the other hand, model-based approaches are more robust than learning methods and helpful in adverse conditions, as shown in works studying neuronal imaging in mammalian brain tissue [59], [60].

2.3.1 Model-based Methods

Early work in LFM exploit refocusing [45] or the Richardson-Lucy (RL) algorithm for reconstruction [51]. As mentioned in Section 2.2.3, Levoy et al. first introduced an algorithm that follows two steps [45]. It refocuses the LF to obtain a raw volume or focal stack, then sharpens the stack using 3D deconvolution. Even though this method is fast, the reconstruction has limited quality, and the microlens pitch size determines the lateral resolution of the reconstructed volume. Specifically, the lateral resolution is always the same as the sampling interval of the LF function, which is restricted by the existing trade-off between spatial and angular sampling of the LF [45]. In LFM, a single camera sensor and a MLA are used to capture both the position and direction of the light rays or, equivalently, both spatial and angular information of the LF. This optical arrangement imposes physical limitations on the sampling rates of the LF function: the spatial sampling interval is proportional to the microlens pitch, whereas the angular sampling interval is inversely proportional to the number of pixels under each microlens [45]. Thus, if the LF function is sampled with a high spatial sampling density, it is necessarily sampled with a low angular sampling density and vice versa.

Later, Broxton et al. [51] proposed computing the measurement matrix of the system by using a wave-optics model. This model is used to pose an inverse problem solved using the RL algorithm [51]. RL is an algorithm with multiplicative updates that assumes Poisson noise for maximum likelihood estimation [61] [62]. This method achieves better performance than the refocusing approach in terms of quality. Furthermore, the resolution of the reconstructed volume is not limited by the microlens pitch. However, this strategy is computationally demanding and suffers from square-like artifacts at specific depths.

Currently, model-based reconstruction approaches rely mainly on a RL strategy. In [18], Prevedel et al. proposed a RL-like algorithm called Iterative Image Space Reconstruction Algorithm (ISRA) [63]. When applied to time series extraction from neuronal cells expressing fluorescent activity reporters, Nöbauer et al. [17] included a reconstruction step using total variation and sparse priors using a modified version of ISRA. More recently, Lu et al. [58] proposed a Phase-Space deconvolution method that avoids reconstruction artifacts by first upscaling the LF image and then using ISRA for each LF view.

Stefanoiu et al. [64] proposed an approach that includes a filtering step to remove artifacts after each RL iteration.

2.3.2 Learning-based Methods

Apart from model-based methods, various approaches that exploit deep learning for reconstruction have been proposed recently. In [65], Wang, et al. describe the first approach that uses an end-to-end convolutional neural network (CNN) for reconstruction. The VCD-Net network is a 2D U-Net trained using synthetic LF data and 3D images obtained with confocal microscopy as labels. The VCD-Net is tested on real LF data by imaging neuron activity in C. elegans and blood flow in the heart of zebrafish larvae.

Later, a technique that uses a mixed reconstruction approach was proposed by Li et al. in [66]. The network named deepLFM is designed to enhance the reconstruction obtained after a few RL iterations on LF images. DeepLFM is a 3D U-Net trained and tested using labels obtained by 3D imaging K562 cells with confocal fluorescence microscopy. Then, Page et al. proposed 3D reconstruction from LF images using a network based on a 2D U-net named LFMNet [43]. LFMNet is trained on real LF data and 3D stacks obtained via confocal microscopy. The training data is obtained after imaging brain slices with fluorescently labeled blood vessels.

Finally, a convolutional neural network (CNN) named HyLFM that can be retrained to refine the 3D reconstruction with the aid of an additional single plane selective-plane illumination microscopy (SPIM) image has been proposed in [67]. HyLFM is trained on real LF images and SPIM stacks as labels. HyLFM is specifically tested to image medaka heart dynamics and zebrafish neuronal activity.

Chapter 3

Shift-Invariant-Subspace Discretization and Model Simplification

THE conventional reconstruction methods used for LFM require long computation times, which clashes with the goal of LFM to analyze time series of volumetric samples or to study fast biological dynamics of specimens in-vivo. A single spatio-temporal measurement contains hundreds of light field images. Studying large brain regions implies reconstructing thousands of 3D images from LF images. Since conventional reconstruction methods may require hours for single 3D reconstruction, the computational burden becomes a bottleneck for the practical use of LFM in neuroscience. High-performance computing or powerful graphical processor units usually alleviate this issue, but it makes LFM out of reach for many scientists. Thus, time constraint limits the complexity of the reconstruction algorithm for many time-sensitive applications. This chapter studies model simplification methods that accelerates the computation of the forward model.

We propose a generalized discretization framework based on shift-invariant subspace modelling for LFM. Unlike the discretization approach used in conventional reconstruction methods, our framework allows diverse sampling densities and template functions. The canonical discretization approach is a particular case of our framework.

The LFM system is modelled using filter banks, and the corresponding forward model is simplified using singular value decomposition (SVD). Our approach allows faster

computation compared to conventional modelling. Furthermore, our model can be conveniently implemented using a linear convolutional neural network (CNN).

3.1 Forward Model Analysis

As mention in Chapter 2, a LF microscope can be described as an operator that transforms a 3D scalar field into a 2D discrete image called a “light field", as in [51]. For simplicity, we ignore one lateral dimension and assume unit magnification since extending the analysis to the original problem is straightforward. Under these assumptions, a monochromatic LF microscope can be modelled as an operator $\mathcal{H}_d$ that transforms an intensity distribution $f(x, z)$ defined for any lateral and axial coordinates x, z into a pixel value $g[k]$ for any index k as follows:

$$g[k] = \mathcal{H}_d\{f(x, z)\}. \tag{3.1}$$

The functions $f(\cdot, \cdot) \in L^2 : \mathbb{R}^2 \mapsto \mathbb{R}^+$ and $g[\cdot] \in \ell^2 : \mathbb{Z} \mapsto \mathbb{R}^+$ represent the volumetric input and the LF, respectively. To characterize the LF microscope $\mathcal{H}_d$, we divide it into a cascade of two blocks, as shown in Figure 3.1(a). The first block describes a continuous ideal microscope $\mathcal{H}$ that is usually described using wave optics. The second block represents a discretization process of the LF. This process is performed intrinsically by the pixels in the image sensor, and it can be modelled by a convolution followed by sampling.

For fluorescent microscopy imaging, it is realistic to assume that light emitted from two different sources is incoherent and that effects of occlusion or non-constant refractive indexes can be safely omitted. Based on these assumptions, the operator $\mathcal{H}$ that maps the volumetric input to the output light intensity (magnitude square of the complex field) is linear. Hence, for a volumetric input $f(\mathbf{p})$, the intensity $g(x)$ observed before sampling at the image sensor can be described with a superposition integral as follows:

$$g(x) = \int h(x, \mathbf{p}) f(\mathbf{p}) d\mathbf{p}, \tag{3.2}$$

where the function $h(x, \mathbf{p})$ is the impulse response of the system $\mathcal{H}$, or, equivalently, the light intensity at the output when the input is a point source located at $\mathbf{p} = (x_\mathbf{p}, z_\mathbf{p}) \in \mathbb{R}^2$.

In this thesis, we assume that $h(x, \mathbf{p})$ is computed using the optical analysis first

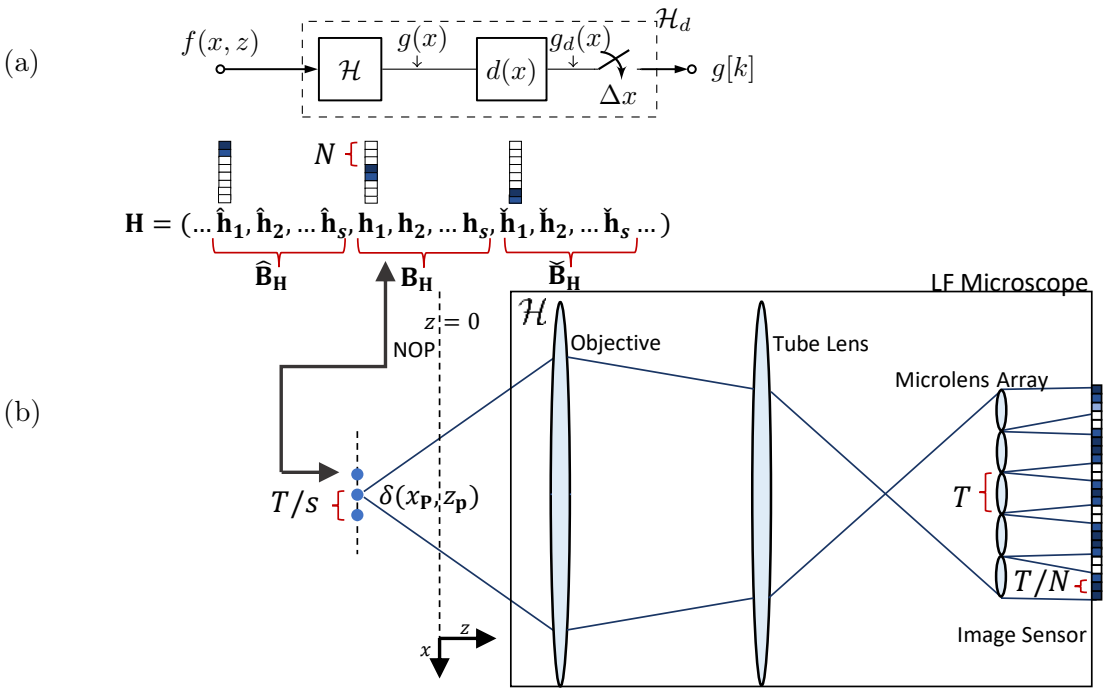


Figure 3.1: Description of a LF system. (a) Basic block diagram of a LF microscope $\mathcal{H}_d$. The system is composed of a continuous ideal microscope $\mathcal{H}$ followed by a discretization step represented by a convolution with kernel $d(x)$ followed by sampling. (b) Conventional discretization approach and optical diagram. The conventional approach computes the impulse response for different input locations, which leads to a block-circulant-like matrix for each z-location. The upsampling factor s and the number of pixels under each microlens N define the block sizes of H.

proposed by Broxton et al. [51]. Thus, the impulse response $h(x, \mathbf{p})$ is described by studying three parts of the system: (a) propagation from the source to the microlens array, (b) modulation due to the microlenses, and (c) propagation from the microlens array to the image sensor plane. As mentioned in Chapter 2, an additional process must model the LF discretization performed by the image sensor. This process is a convolution with a kernel $d(x)$ followed by sampling. The final output of the system is a discrete image $g[k]$:

$$g[k] = g_d(k\Delta x), \tag{3.3}$$

where $g_d(x) = g(x) * d(x)$, and Δx is the sampling interval. In many practical cases, it holds that $\Delta x = T/N$, where T is the microlens pitch and N is the number of pixels under each microlens. See Figure 3.1 (b).

As explained in [51], the system $\mathcal{H}$ is not shift-invariant. However, it satisfies an important property: periodic shift-invariance. Namely, ignoring the finite extent of the

device, for any $x, z \in \mathbb{R}$, if the input is shifted by multiples of T along the lateral axis, then the output is also shifted by multiples of T. Hence, the following relationship holds:

$$g(x - nT) = \mathcal{H}\{f(x - nT, z)\}, \forall n. \tag{3.4}$$

This means that for any depth z, this equation describes a periodically shift-invariant behaviour. An equivalent form to characterize this property is to state that the impulse response $h(x, x_{\mathbf{p}}, z_{\mathbf{p}})$ of the system $\mathcal{H}$ is periodic:

$$h(x, x_{\mathbf{p}}, z_{\mathbf{p}}) = h(x - T, x_{\mathbf{p}} - T, z_{\mathbf{p}}), \tag{3.5}$$

where $x_{\mathbf{p}}$ and $z_{\mathbf{p}}$ are the coordinates of the input point source. In the traditional approach, the entire system is discretized for numerical evaluation [51], as follows: first, a Dirac delta is shifted laterally and axially; then, each shifted Dirac is used individually as an input to compute the corresponding impulse response, which is then stored in a matrix. The shifts of the Dirac delta are defined by an axial sampling interval Δz and a lateral sampling interval $\frac{T}{s}$, where the integer s is the upsampling factor, as shown in Figure 3.1(b). Since the system is not shift-invariant, multiple impulse responses for different positions must be stored. The final discrete system can be described by:

$$\mathbf{g} = \mathbf{H}\mathbf{f}, \tag{3.6}$$

where matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ maps a vectorized volume $\mathbf{f} \in \mathbb{R}^n$ into a LF $\mathbf{g} \in \mathbb{R}^m$. The number n of voxels of the volume is usually much larger than the number m of pixels of the LF.

In [51], the inverse problem derived from Equation (3.6) is solved using RL by assuming LF data corrupted by Poisson noise. Since property (3.4) leads to matrix $\mathbf{H}$ that is block-circulant for each z position, it is feasible to solve the problem by using RL [51]. However, this can be highly computationally demanding. Furthermore, the recovered volume suffers from artifacts near the native object plane (NOP): the plane where a point source is in focus at the microlens array, as described in [51] (conventionally, the plane $z = 0$).

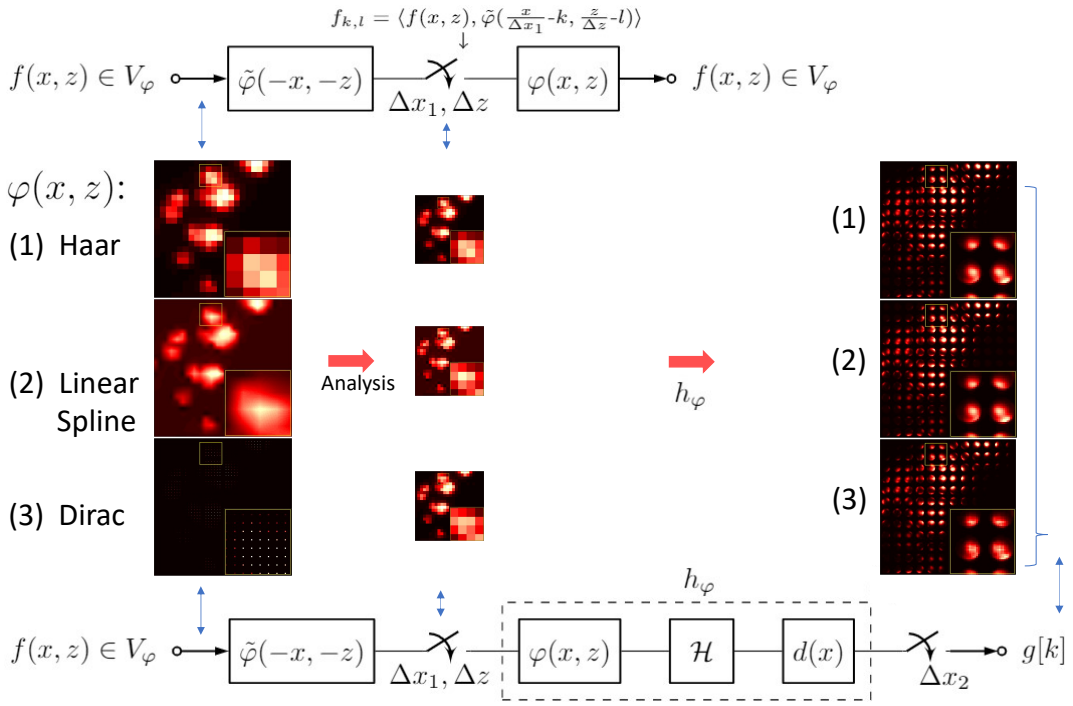


Figure 3.2: Discretization of the LF system using shift-invariant subspaces. We show the discretization of a volume f. If f belongs to the shift-invariant subspace defined by template function $\varphi(x,z)$, the discretization can be inverted after filtering with $\varphi(x,z)$, as shown in the block diagram at the top of the figure. This allows performing a non-standard discretization that generalizes the conventional discretization method in LFM, as indicated at the bottom of the figure. Notice that the input discretization depends on the sampling density and template function $\varphi(x,z)$, while the output is discretized intrinsically by the device depending on the physics of the pixel sensors. In our example we show three possible choices for the input filters (Haar, Linear Spline and Dirac delta).

3.1.1 A General Discretization Framework

As mentioned previously, the standard approach presented in [51] discretizes the system by computing the impulse response for different input positions. However, a more general strategy is to discretize both the input and output using the assumption that the volume $f(x,z)$ belongs to a shift-invariant subspace (SIS) defined by a template function $\varphi(x,z)$, which was proposed in [68] for convolution operators in digital holography. If a function $f(x,z)$ belongs to a SIS V_φ generated by $\varphi(x,z)$, it can be written as follows:

$$f(x,z) = \sum_k \sum_l f_{k,l} \varphi\left(\frac{x}{\Delta x_1} - k, \frac{z}{\Delta z} - l\right), \tag{3.7}$$

where $f_{k,l} = \langle f(x,z), \tilde{\varphi}(\frac{x}{\Delta x_1} - k, \frac{z}{\Delta z} - l)\rangle$ and $\tilde{\varphi}(x,z)$ is the dual basis of $\varphi(x,z)$. As shown in Figure 3.2 (top), the relationship shown in Equation (3.7) can be represented with a simple pipeline: convolution, sampling, convolution. Note that any input $f(x,z)$ belonging to the subspace V_φ can be perfectly reconstructed from its samples $f_{k,l}$. Furthermore, the samples $f_{k,l}$ can be computed by filtering $f(x,z)$ with $\tilde{\varphi}(x,z)$ and then sampling with sampling interval Δx_1, Δz. This idea is exemplified in Figure 3.2 (top) by using three types of filters: a Haar, a linear spline and a Dirac delta.

If $f(x,z)$ is the input to the LF microscope, the model in Equation (3.7) suggests an alternative discretization process illustrated in Figure 3.2 (bottom). The input is sampled using the template function $\varphi(x,z)$ with sampling intervals Δx_1, Δz. The output is sampled using the template function $d(x)$ with a sampling interval Δx_2. Notice that the sampling of the output is performed intrinsically by the LF microscope, whereas the sampling of the input depends on how $f(x,z)$ is modelled. For example, if we assume $f(x,z)$ is a bandlimited signal, $\varphi(x,z)$ is the sinc function. If we assume $f(x,z)$ is a uniform spline, then $\varphi(x,z)$ is a spline. In contrast, $d(x)$ is usually a fixed box function (Haar filter) that models the effect of each pixel in the camera sensor [63]. Since the filter $d(x)$ should have support equal to the pixel size to simulate the effect of the pixel, a box function is usually enough to model the system properly. See Figure 3.2.

The discretization process leads to a new system defined by the impulse response $h_\varphi(x, x_\mathbf{p}, z_\mathbf{p})$. This system is equivalent to a cascade of three systems, (a) a 2D convolution with kernel $\varphi(x,z)$, (b) the system $\mathcal{H}$, and (c) a 1D convolution with kernel $d(x)$. See Figure 3.2 (bottom) for clarification. Therefore, by construction, periodicity still holds for the impulse response $h_\varphi(x, x_\mathbf{p}, z_\mathbf{p})$:

$$h_\varphi(x, x_\mathbf{p}, z_\mathbf{p}) = h_\varphi(x - T, x_\mathbf{p} - T, z_\mathbf{p}). \tag{3.8}$$

The corresponding discretized impulse response can be found from the impulse response $h_\varphi(x, x_\mathbf{p}, z_\mathbf{p})$ by sampling it along each dimension. Therefore, the discrete system is defined by:

$$h_\varphi[k, k_\mathbf{p}, l_\mathbf{p}] = h_\varphi(k\Delta x_2, k_\mathbf{p}\Delta x_1, l_\mathbf{p}\Delta z). \tag{3.9}$$

Furthermore, using Equation (3.8) and Equation (3.9) it is possible to show that a periodicity property also holds for the discrete impulse response $h_\varphi[\cdot]$:

$$h_\varphi[k, k_\mathbf{p}, l_\mathbf{p}] = h_\varphi[k - rq, k_\mathbf{p} - ts, l_\mathbf{p}], \tag{3.10}$$

where we assumed that integers q, r, s and t exist such that $\frac{\Delta x_1}{T} = \frac{q}{s}$ and $\frac{\Delta x_2}{T} = \frac{t}{r}$ and where both fractions are irreducible. To see this, note that from Equation (3.8), it always holds that:

$$h_\varphi(x, x_\mathbf{p}, z_{[]p}) = h_\varphi(x - aT, x_\mathbf{p} - aT, z_\mathbf{p}), \tag{3.11}$$

where a can be any integer. Then, using Equation (3.9), we can write the following equivalence:

$$h_\varphi[k, k_\mathbf{p}, l_\mathbf{p}] = h_\varphi(k\Delta x_2, k_\mathbf{p}\Delta x_1, l_\mathbf{p}\Delta z) = h_\varphi((k - a\frac{T}{\Delta x_2})\Delta x_2, (k_\mathbf{p} - a\frac{T}{\Delta x_1})\Delta x_1, l_\mathbf{p}\Delta z). \tag{3.12}$$

Finally, Equation (3.10) is found by choosing $a = tq$, and using the assumption that $\frac{\Delta x_1}{T} = \frac{q}{s}$ and $\frac{\Delta x_2}{T} = \frac{t}{r}$.

The conventional discretization procedure can be understood as a particular case of this more general framework. Note that if we allow $\varphi(x, z)$ to be a Dirac delta, $q = 1$ and $t = 1$, our framework reduces to the standard discretization approach. In this case, s is the upsampling factor, r is the number of pixels under each microlens N, and $h_\varphi[k, k_\mathbf{p}, l_\mathbf{p}]$ is the standard discrete impulse response, as described in [51].

3.1.2 Filter-Bank Description and Model Simplification

Equation (3.10) indicates that periodically-shift invariance holds for each depth of the discrete system. Therefore, the measurement matrix $\mathbf{H}$ derived from the discretized impulse response $h_\varphi[k, k_\mathbf{p}, l_\mathbf{p}]$ can be written as

$$\mathbf{H} = \sum_{i=1}^{D} \mathbf{H}_i \mathbf{S}_i, \tag{3.13}$$

where $\mathbf{S}_i$ is an auxiliary matrix that selects the depth $l_\mathbf{p} = i$ from the discrete volumetric input, D is the number of depths, and each matrix $\mathbf{H}_i$ follows a block-circulant structure due to Equation (3.10). Furthermore, as depicted in Figure 3.3, the periodic shift-invariant

property allows each $\mathbf{H}_i$ to be represented as a filter bank that performs a set of convolutions with the input. This accelerates the computation of the forward model. Assuming that each $\mathbf{H}_i$ is full rank, $2 \times ts \times D$ convolutions are needed to describe the whole $\mathbf{H}$. This is because the number of branches needed to represent each $\mathbf{H}_i$ equals ts, and 2 convolutions are required for each branch, as shown in Figure 3.3 (b). However, these computational requirements may be excessive in practice since the number ts of branches can be very high.

To mitigate this issue, we propose a simplification of the forward model. The computation can be accelerated by reducing the number of branches of each filter bank to F by optimally choosing the corresponding filters. First, we analyze the case $F = 1$, and then we generalize for arbitrary F. Thus, to approximate a given $\mathbf{H}_i$ using a filter bank with a single branch, we state the following optimization problem:

$$\min_{\mathbf{c},\mathbf{v}} \quad \|\mathbf{H}_i - \mathbf{C}\mathbf{U}_{rq}\mathbf{D}_{ts}\mathbf{V}^{\mathsf{T}}\|_F^2, \tag{3.14}$$

where the matrix $\mathbf{D}_{ts}$ is a downsampling matrix of factor ts and the matrix $\mathbf{U}_{rq}$ models upsampling by rq. The matrix $\mathbf{V}$ is circulant, and it is completely defined by one column $\mathbf{v}$, which corresponds to the input filter of the filter bank. Similarly, the matrix $\mathbf{C}$ is circulant, and one column $\mathbf{c}$ corresponds to the output filter of the filter bank. The proper selection of $\mathbf{c}$ and $\mathbf{v}$ will allow the best approximation of $\mathbf{H}_i$ in the least square sense. The norm used here is the Frobenius norm.

To solve this problem, we limit the analysis to $\mathbf{v}$ with compact support ts. Since both $\mathbf{H}_i$ and $\mathbf{C}\mathbf{U}_{rq}\mathbf{D}_{ts}\mathbf{V}^{\mathsf{T}}$ have a similar structure, they can be expressed as follows:

$$\mathbf{H}_i = (...\hat{\mathbf{B}}_{\mathrm{H}}, \mathbf{B}_{\mathrm{H}}, \check{\mathbf{B}}_{\mathrm{H}}...), \tag{3.15}$$

and

$$\mathbf{C}\mathbf{U}_{rq}\mathbf{D}_{ts}\mathbf{V}^{\mathsf{T}} = (...\hat{\mathbf{B}}_{\mathrm{CV}}, \mathbf{B}_{\mathrm{CV}}, \check{\mathbf{B}}_{\mathrm{CV}}...). \tag{3.16}$$

The notation $\hat{\mathbf{B}}_{\mathrm{H}}$ means that each column of $\mathbf{B}_{\mathrm{H}}$ is shifted upward by rq elements. Similarly, $\check{\mathbf{B}}_{\mathrm{H}}$ indicates a downward shift by rq elements. Similar notations apply to $\mathbf{B}_{\mathrm{CV}}$. See Figure 3.1 (b) and Figure 3.3 (a) for clarification. Hence, using Equation (3.15) and (3.16),

(a)

(b)

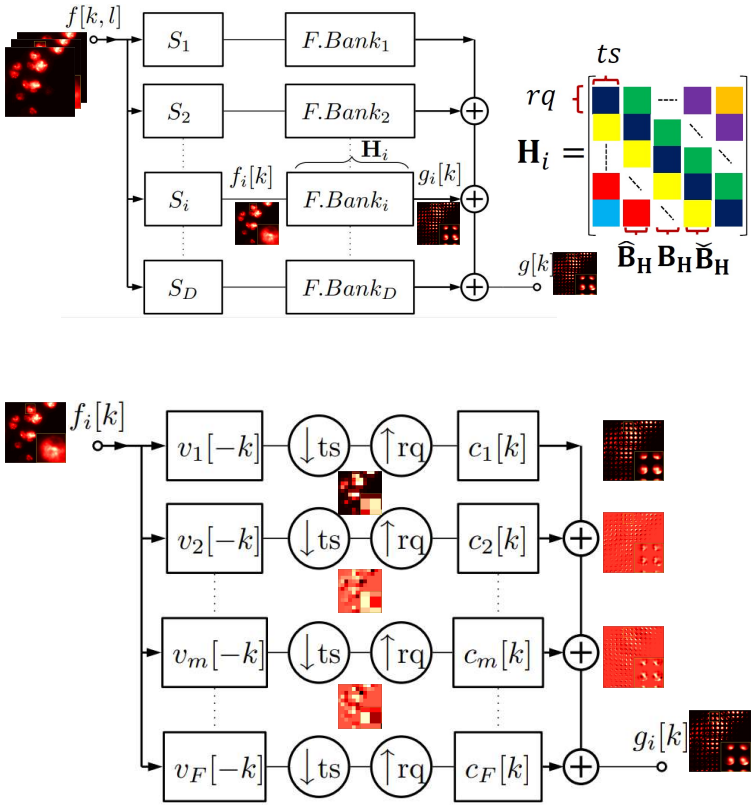


Figure 3.3: Filter bank representation. In (a), a block diagram of LF microscope is depicted. For each depth i, the linear system S_i performs a slicing operation, which chooses the respective i-th slice of the volume $f_i[k] = f[k, i]$. Then, each slice is passed through a filter bank to obtain a LF $g_i[k]$ per depth i. Finally, the LF $g[k]$ is the summation of all $g_i[k]$. (b) For each depth, the filter bank can be approximated by reducing the number of branches of the filter bank from ts to F. The sampling filter $v_m[k]$ defines the circulant matrix $\mathbf{V_m}$ and the filter $c_m[k] = \sigma_m u_m[k]$ defines the circulant matrix $\mathbf{C_m}$. Filters $v_m[k]$, $u_m[k]$ are the left and right singular vectors and coefficients σ_m are the singular values coming from a singular value decomposition. Notice that when $F = ts$, the approximation becomes equality.

expression (3.14) reduces to the minimization of $\|\mathbf{B_H} - \mathbf{B_{CV}}\|_F^2$. Based on the compact support assumption for $\mathbf{v}$, it holds that $\mathbf{B}_{\mathrm{CV}} = \mathbf{c}\mathbf{v}^\mathsf{T}$. Therefore, $\mathbf{c}$ and $\mathbf{v}$ should be chosen such that:

$$\min_{\mathbf{c},\mathbf{v}} \quad \|\mathbf{B_H} - \mathbf{c}\mathbf{v}^\mathsf{T}\|_F^2. \tag{3.17}$$

This minimization is a rank-1 approximation easily solved by using the SVD of $\mathbf{B_H}$. More importantly, computing the approximated forward model $\mathbf{C}\mathbf{U}_{rq}\mathbf{D}_{ts}\mathbf{V}^\mathsf{T}$ involves only two convolutions, implying much less computational complexity than computing the whole model.

As a generalization to the above rank-1 (very coarse) approximation of $\mathbf{H}_i$ by solving problem (3.14), the following rank-F approximation that keeps the largest F singular values of $\mathbf{B_H}$ is proposed:

$$\mathbf{H}_i \approx \sum_{m=1}^{F} \mathbf{C}_m \mathbf{U}_{rq} \mathbf{D}_{ts} \mathbf{V}_m^{\mathsf{T}}. \tag{3.18}$$

For each matrix $\mathbf{V}_m$, the respective $\mathbf{v}_m$ is a right singular vector of $\mathbf{B_H}$ and the $\mathbf{c}_m$ of $\mathbf{C}_m$ is a left singular vector $\mathbf{u}_m$ multiplied by the respective singular value σ_m: $\mathbf{c}_m = \sigma_m \mathbf{u}_m$. Notice that, since this approximation comes from an SVD, the summation progressively improves the approximation of the given matrix $\mathbf{H}_i$. In particular, when F equals ts, the approximation turns into an equality. We also note that Equation (3.18) can be interpreted as a filter bank with F branches. Thus, for each depth, the approximation of $\mathbf{H}_i$ with F terms reduces the number of branches of the filter bank from a maximum of ts to F (see Figure 3.3). Finally, we highlight that given a volumetric input with an arbitrary number of depths D, the approximation in Equation (3.18) can be applied to each matrix $\mathbf{H}_i$ independently for every depth $i = 1...D$ using a depth-dependent number of branches F_i.

3.1.3 Experiments and Results

For these experiments, the input sampling intervals are set as $\Delta x_1 = T/N_1$ for both lateral axes and $\Delta z = T$, while the output sampling intervals are set to $\Delta x_2 = T/N_2$ for both lateral axes, where $N_1 = 8$, $N_2 = 19$, $T = 125\ \mu m$, and the microscope magnifies the volume by a factor of $M = 25$. In part (a), we set $\varphi(x, z)$ to be a Dirac delta and a linear spline in part (b). Figure 3.4 shows $F = 8$ optimal filters computed using our SVD approximation for two depths. The shown filters are the equivalent 2D versions of the filters $\{v_m\}_{m=1}^{F}$ of the filter bank shown in Figure 3.3. One can observe that the shape of each filter changes with the depth. In part (a), for $z = 0\ \mu m$, the filter for singular value σ_1 is close to a square, which means that the system averages a region of the size of the microlens area, and projects this region into the LF space. Furthermore, the singular value σ_2 is almost 0.09 times the singular value σ_1, which shows the rank deficiency of the matrix $\mathbf{B_H}$ for this depth. However, changing the template function to a linear spline modifies the measurement matrix of the system, as shown in Equation (3.9). We observe in part (b)

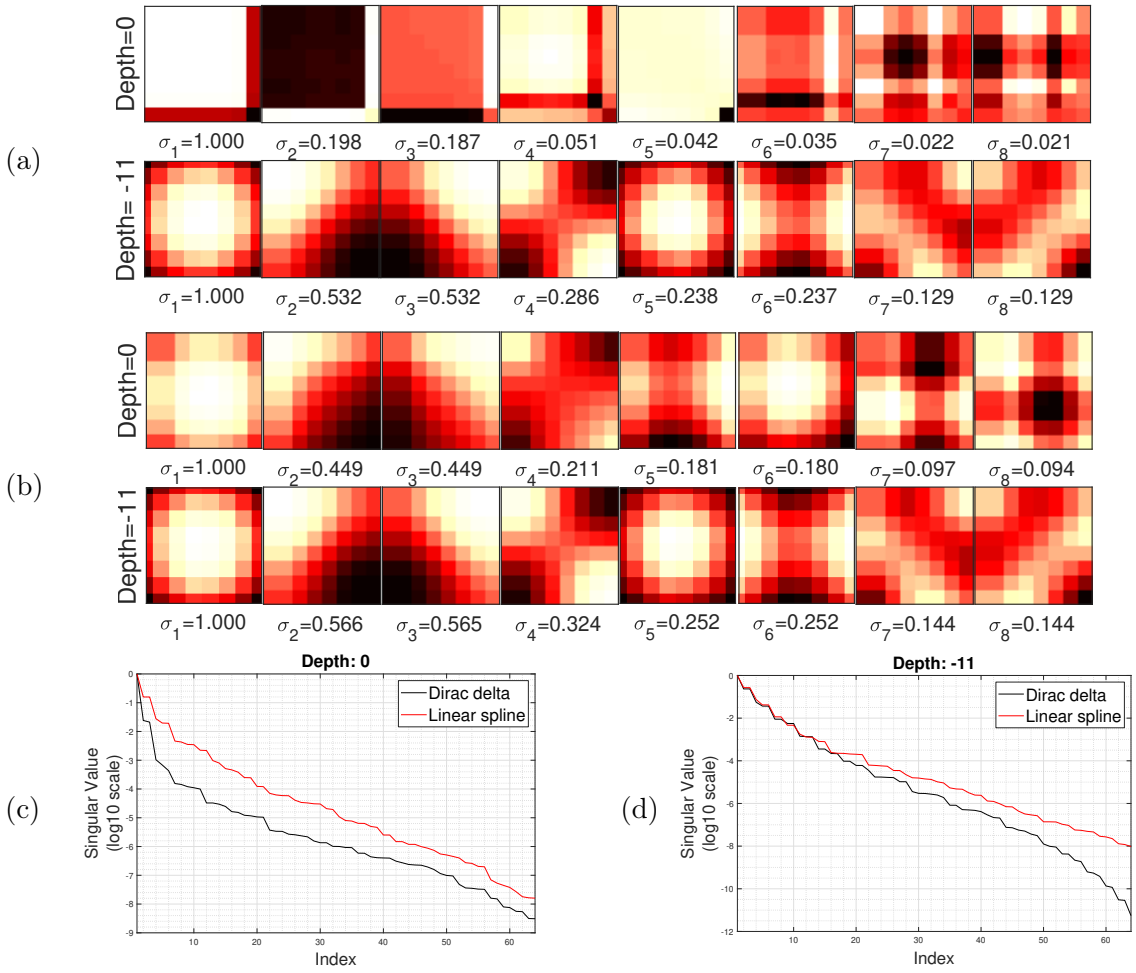


Figure 3.4: Optimal analysis filters. This figure shows $F = 8$ optimal filters found for two different depths ($z = -11\ \mu m$, $z = 0\ \mu m$) from a maximum of $ts = 8 \times 8$ filters. For each depth, these filters are the 2-D version of the 1-D filters named $\{v_m\}_{m=1}^{F}$. In (a) we use a Dirac Delta as the template function φ and in (b) we use a spline. The shown singular values σ_m are normalized to be between 0 and 1. The use of a spline changes the singular value distribution with respect the original Dirac delta template. In (c), we show the distribution of the singular values for depth $z = 0\mu m$ and the distribution for $z = -11\mu m$ in (d).

that the optimal filter corresponding to σ_1 for $z = 0\ \mu m$ is not a square anymore, which shows how the rank deficiency of the matrix $\mathbf{B_H}$ was reduced due to the integration along the axial dimension that occurs when a spline is selected as a template function.The use of linear splines as template function changes the singular value distribution with respect the original Dirac delta, as shown in Figure 3.3 (c) and (d). Note that this behaviour is more evident for $z = 0\mu m$ (Figure 3.3 (c)) since this depth inherently shows a low-rank behaviour. Classic Dirac delta approach does not exploit correlation between adjacent points. Using a template function that allows integration along x, y and z axes approximate

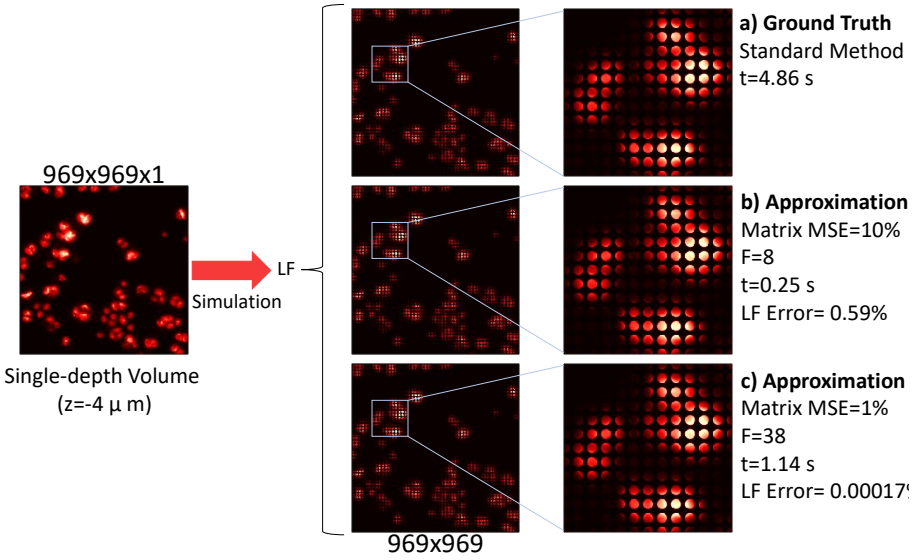


Figure 3.5: Model simplification example. A synthetic LF was computed from a single depth volume. In (a) we show the simulated LF image using the standard model without any approximation. Furthermore, we show results using our approximation method by setting $F = 8$ and $F = 38$, and this leads to a matrix error (mean square error) of 10% in (b) and 1% in (c). The rightmost column shows the average computation time and the error between the approximated LF image and the ground truth. The single-depth volume is a sample of Drosophila melanogaster Kc167 cells taken from a publicly available library [69]. All the methods were tested in a CPU (Intel Core i7-6700, 16 GB RAM) using MATLAB R2018b.

better the impulse response of the system without effectively increasing the input sampling rate and the data dimensionality. In this work, we study splines for discretization but any other template function could be used. However, linear splines are attractive due to their simplicity.

As explained previously, truncating the number of branches of each filter bank of the model accelerates its computation. To show the utility of this approximation, we take a single-depth volume ($z = -4\ \mu m$) and simulate the forward model using both the standard method and our approximation. We use the same downsampling and upsampling factor $ts = rq = 19$ for both lateral axes. Figure 3.5 shows visual results for $F = 8$ and $F = 38$, which correspond to a matrix approximation error of approximately 10% and 1%, respectively. Notice that the number of filters F is much lower than a maximum $ts = 19 \times 19$. Furthermore, for both cases, the approximated LF is visually indistinguishable from the ground truth, the error measured from the LF is less than 6% for $F = 8$, and it is almost zero when $F = 38$. The latter is because the system matrix is inherently low

rank. Furthermore, in this experiment, our simplified model is almost 19 times faster to compute when $F = 8$ and four times faster when $F = 38$ than the standard approach.

3.2 Forward Model as a Linear CNN

In this section, we propose a novel description of the LF system by using convolutional layers. The previous section shows that the system can be described and approximated using filter banks. This section converts the filter bank description into a linear Convolutional Neural Network (CNN). The linear CNN allows faster computation by exploiting deep learning software optimized to compute convolutional layers.

3.2.1 4D Representation of LF

As mentioned in Chapter 2, the LF can be efficiently represented as a 4D function in photography-related applications. Specifically, the 2D image captured with a LF camera is reordered into a 4D array that is a sampled version of the continuous LF function [30]. Two dimensions of the array represent horizontal and vertical spatial coordinates, while the other two represent horizontal and vertical spatial frequencies. The 4D LF can be interpreted as a collection of sub-aperture images or views, which are 2D images obtained when the spatial frequencies are fixed [30]. See Figure 3.6 (a) for clarification. In microscopy, the idea of capturing a 4D LF is still valid if the array is interpreted as a sampled version of a 4D Wigner distribution function, a generalization of the concept of LF that considers the effects of diffraction and briefly discussed in Section 2.2.3 [57].

3.2.2 Linear CNN

The representation of the LF as a group of views has a convenient property. Unlike the 2D LF image, each view is not an abstract pattern. Each view preserves the structure of the original scene since it only carries spatial information. Furthermore, this multi-view representation is attractive for 3D reconstruction in LFM since it is more suitable to fit conventional CNN architectures, as also proposed in [70].

Since the 4D LF can be obtained by just rearranging pixels of the 2D LF, the filterbank description of the LF system described in Section 3.1.2 can be adjusted to explain the formation of a group of sub-aperture images. Specifically, reordering the synthesis

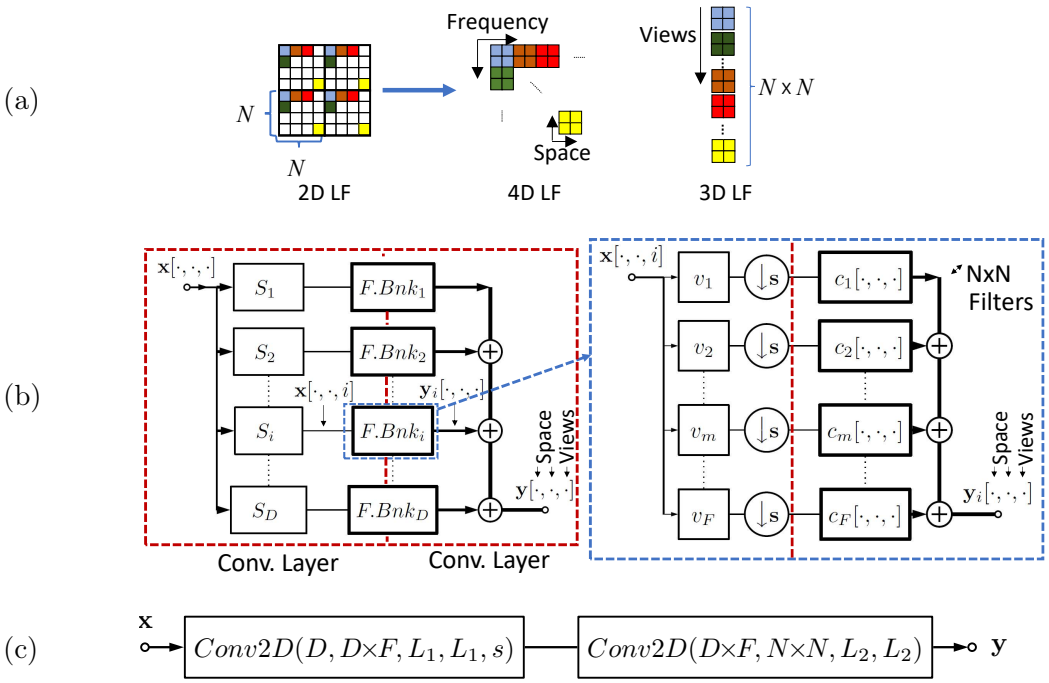


Figure 3.6: Linear forward model. In (a) we show how a 2D LF can be transformed into a set of $N \times N$ sub-aperture images, where $N \times N$ is the number of pixels under each microlens. In (b), we show how to describe the image formation of a 3D LF image using filter banks. Observe to the right that the model does not contain upsampling blocks since they were absorbed due to the reordering; furthermore, the synthesis filters were replaced by a group of $N \times N$ filters per branch that introduce the additional dimension representing the view index. In (c) we show the representation of the LFM system as a forward CNN $f(\cdot)$. The previous structure (b) is a particular case of the CNN $f(\cdot)$. The notation $Conv2D(\cdot, \cdot, \cdot, \cdot, \cdot)$ means a 2D convolutional layer with input parameters ordered as follows: number of input channels, number of output channels, height of the filter, width of the filter, and stride. If the stride is omitted, it means unit stride.

filter of each branch allows simple computation of the sub-aperture images, as shown in Figure 3.6 (b). Note that this new representation follows the basic structure of the original filter bank in Figure 3.3 (a). However, there is no upsampling block since the original synthesis filter bank is replaced by a group of $N \times N$ filters that leads to multiple outputs forming a set of views or sub-aperture images.

The multiple-view filterbank description of the system can be conveniently implemented by using convolutional layers. The LF system can be written as a cascade of two 2D convolutional layers without bias term, where the first layer has a stride given by the downsampling factor s and the second layer has a unit stride. We represent the CNN by $f(\cdot)$ and call it forward CNN (see Figure 3.6 (c) for clarification). Since the forward CNN

Table 3.1: Forward model computation time.

	Traditional [17]	Linear CNN
Time (s)	12.36	0.13

is derived from the filter bank representation, the parameters of $f(\cdot)$ have a connection with the parameters of the microscope and the physics of the system: the number $N \times N$ of output channels is defined by the number of pixels under each microlens, the number of input channels D is the number of depths, the filter size L_1 is equal to the dowsampling factor s, as explained in Section 3.1.2, the filter size L_2 is instead given by the support of the PSF, specifically if the support of the PSF related to the largest depth is M, then $L_2 = M/N$. Finally, the parameter F is related to the upsampling factor. If F is set to $s \times s$, it resembles the theoretical model exactly, while if it is set to a smaller value, it performs an approximation, as explained in Section 3.1.2.

The CNN representation offers computational advantages compared to the traditional method. The restriction on the filter size L_1 needed for our SVD method could be removed, and the weights can be found using any optimization framework designed for CNN training. Moreover, the adjoint operator (transpose) can be computed easily by permutations and reflections of the weights of the CNN, which is a task that is demanding in typical approaches [71]. Furthermore, the CNN allows parameters to be computed directly from the theoretical model or they can be learned if a labelled dataset is available. Finally, in Table 3.1, we compare the computation time required to compute the forward model using the linear CNN and the traditional method [17]. For this experiment, we use a 3D input of size $321 \times 321 \times 53$ voxels and the output is of size 2033×2033. The proposed CNN architecture of the forward model exploits existing software for deep learning and achieves faster computation. Both methods were evaluated on a GeForce GTX 1080 Ti. These advantages are exploited for reconstruction in Chapter 5.

3.3 Summary

We have proposed a novel discretization method that uses more diverse sampling densities and shift-invariant subspaces to sample the impulse response of the LF system. We have shown that the periodic-shift invariance still holds in the discrete domain if the ratio

between the input and output sampling interval is a rational number.

Moreover, we have proposed to model the system using filter banks. Both analysis and synthesis filters of the filter bank can be chosen using the SVD. Furthermore, the SVD allows a method that accelerates the computation of the forward model. We can approximate the forward model progressively by using a truncated SVD.

Finally, we have proposed a method to reshape the forward model into a linear CNN. This novel system description allows taking advantage of existing software designed for deep learning for faster computation and optimization.

Chapter 4

Model-Based Reconstruction for LFM

LIGHT-FIELD microscopy aims to recover high-quality 3D images from single LF measurements. However, using a 2D image to encode 3D spatial information limits the quality of the reconstruction. Therefore, typical model-based reconstruction approaches suffer from low resolution and reconstruction artifacts.

This chapter analyzes the model-based reconstruction of 3D volumes from a single light field image. We first focus on aliasing-free reconstruction under ideal settings by exploiting the shift-invariant subspace assumption. Then, instead of using conventional methods for inversion based on the Richardson Lucy (RL) algorithm, we propose a general method for reconstruction under real acquisition conditions based on ADMM. Our approach alleviates typical reconstruction artifact in RL-based methods. Furthermore, it is faster than typical methods by exploiting simplification techniques explained in Chapter 3.

4.1 Ideal Reconstruction Under Shift-invariant-subspace Assumption

In Section 3.1.1, we discussed an alternative discretization process based on the assumption that the volumetric signal $f(x, z)$ lies in a shift-invariant subspace generated by $\varphi(x, z)$. This is again depicted in Figure 4.1 (a). Under this assumption, the image formation process is described by the discrete filter $h_\varphi[k, k_\mathbf{p}, l_\mathbf{p}]$ and, if this filter is invertible, reconstruction is achieved by first filtering the discrete light field image $g[k]$ with $h_\varphi^{-1}[k, k_\mathbf{p}, l_\mathbf{p}]$

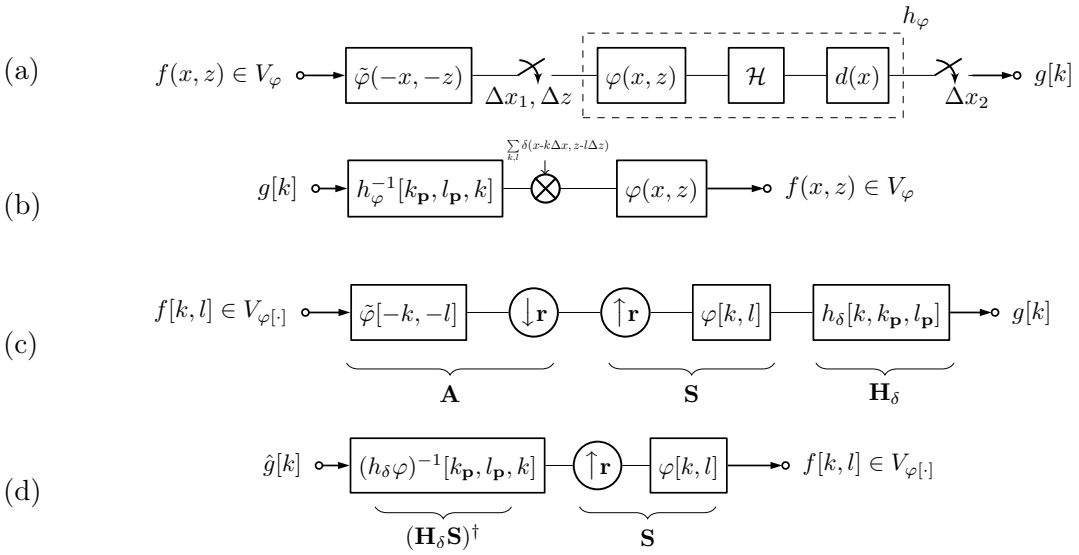


Figure 4.1: Shift-invariant subspace assumption. In (a), we show a method to discretize the system using a shift-invariant subspace (SIS) V_φ. An arbitrary volumetric signal $f(x,z) \notin V_\varphi$ is discretized with a template function $\varphi(x,z)$ at an arbitrary sampling density. In contrast, the output is discretized intrinsically by the microscope with the kernel $d(x)$ at a sampling density defined by the pixel sensor to obtain a discrete output $g[k]$. In (b) we show how to perform perfect reconstruction from $g[k]$ under the assumption that the volume $f(x,z)$ belongs to the assumed SIS V_φ and the discrete system $h_\varphi[k, k_\mathbf{p}, l_\mathbf{p}]$ is invertible. In (c), we show the discrete version of the shift-invariant subspace discretization shown in (a). In (d), we show how to perform perfect reconstruction in the discrete space, under the assumption that $\mathbf{H}_\delta\mathbf{S}$ is full rank and the discrete volume $f[x,z]$ belongs to the assumed discrete SIS $V_{\varphi[\cdot]}$.

and then by resynthesizing the original volume using a continuous filter $\varphi(x,z)$, see Figure 4.1 (b). In other words, perfect reconstruction (aliasing-free) of the continuous signal $f(x,z)$ is always achievable if two conditions hold: (a) $h_\varphi[\cdot]$ is invertible and (b) the volume being imaged genuinely belongs to the SIS generated by $\varphi(x,z)$ e.g. Linear spline, box function, sinc function (which models band-limited functions). As an analogy, in the problem of reconstructing a signal from its discrete samples, the named conditions are analog to the requirement for perfect reconstruction of the input. In Nyquist theorem, when the input signal genuinely belongs to the space of band-limited functions when the cutoff frequency equal to half the sampling frequency, the signal can be perfectly reconstructed from its samples. This observation suggests a way to find volumes that can be reconstructed from the light field image.

In practice, it is difficult to find a closed-form expression for h_φ; we therefore first discretize both $h(x, x_\mathbf{p}, z_\mathbf{p})$ and $f(x, x_\mathbf{p}, z)$ as in [51], but at a much finer resolution than

the native lateral resolution of the LFM and then apply the shift-invariant model to the discretized f, which we denote as $f[k,l]$. This leads to the alternative model shown in Figure 4.1 (c), which can be described in matrix vector form as follows:

$$\mathbf{g} = \mathbf{H}_\delta \mathbf{S}_\varphi \mathbf{A}_\varphi \mathbf{f}. \tag{4.1}$$

Here, vector $\mathbf{f} \in \mathbb{R}^n$ is the vectorized volumetric input, vector $\mathbf{g} \in \mathbb{R}^m$ is the light field image, matrix $\mathbf{A}_\varphi \in \mathbb{R}^{\frac{n}{\mathrm{r}_1\mathrm{r}_2}\times n}$ represents the convolution with the filter $\tilde{\varphi}(x,z)$ followed by downsampling, matrix $\mathbf{S}_\varphi \in \mathbb{R}^{n\times\frac{n}{\mathrm{r}_1\mathrm{r}_2}}$ models upsampling followed by convolution with the filter $\varphi(x,z)$, the downsampling and upsampling are defined by the factor $\mathbf{r} = (\mathrm{r}_1, \mathrm{r}_2)$, and $\mathbf{H}_\delta \in \mathbb{R}^{m\times n}$ is the forward model discretized using [51] (see Figure 4.1 (c)).

To recover the original volume $\mathbf{f}$, one can choose the downsampling factor $\mathbf{r}$ and filter $\varphi[\cdot]$ in Figure 4.1 (c) in a way that $\mathbf{H}_\delta\mathbf{S}_\varphi$ is full column rank and then the pseudoinverse can be used as follows:

$$\mathbf{f} = \mathbf{S}_\varphi(\mathbf{H}_\delta\mathbf{S}_\varphi)^\dagger\mathbf{g}. \tag{4.2}$$

Notice that this approach allows perfect reconstruction if matrix $\mathbf{H}_\delta\mathbf{S}_\varphi$ is full rank and the input belongs to the assumed SIS. This process is depicted in Figure 4.1 (d).

However, we highlight that there are limitations that make $\mathbf{H}_\delta$ rank deficient that cannot be overcome by the selection of a proper SIS. These limitations arise from the inherent lack of diversity in the impulse response for different input positions, e.g., the impulse response for in-focus points. Thus, the rank of $\mathbf{H}_\delta$ imposes a limit on the maximum number of samples that can be recovered. In particular, if the lateral sampling interval Δx_1 is decreased, the axial sampling interval Δz needs to be increased. This implies that it is not feasible to achieve high resolution simultaneously in both axes. Finally, one needs to choose the template function $\varphi[\cdot]$ and downsampling factor $\mathbf{r}$ such that the inversion of $\mathbf{H}_\delta\mathbf{S}_\varphi$ is as stable as possible, which gives a degree of freedom to investigate in practice.

4.1.1 Experiments

In this section, we demonstrate this framework by modelling the SIS using separable 3D filters formed by one-dimensional linear splines. For the subspace used in the first scenario (scenario 1), we set a downsampling factor $\mathbf{r} = (4, 4, 16)$ which means a factor of 4 for both

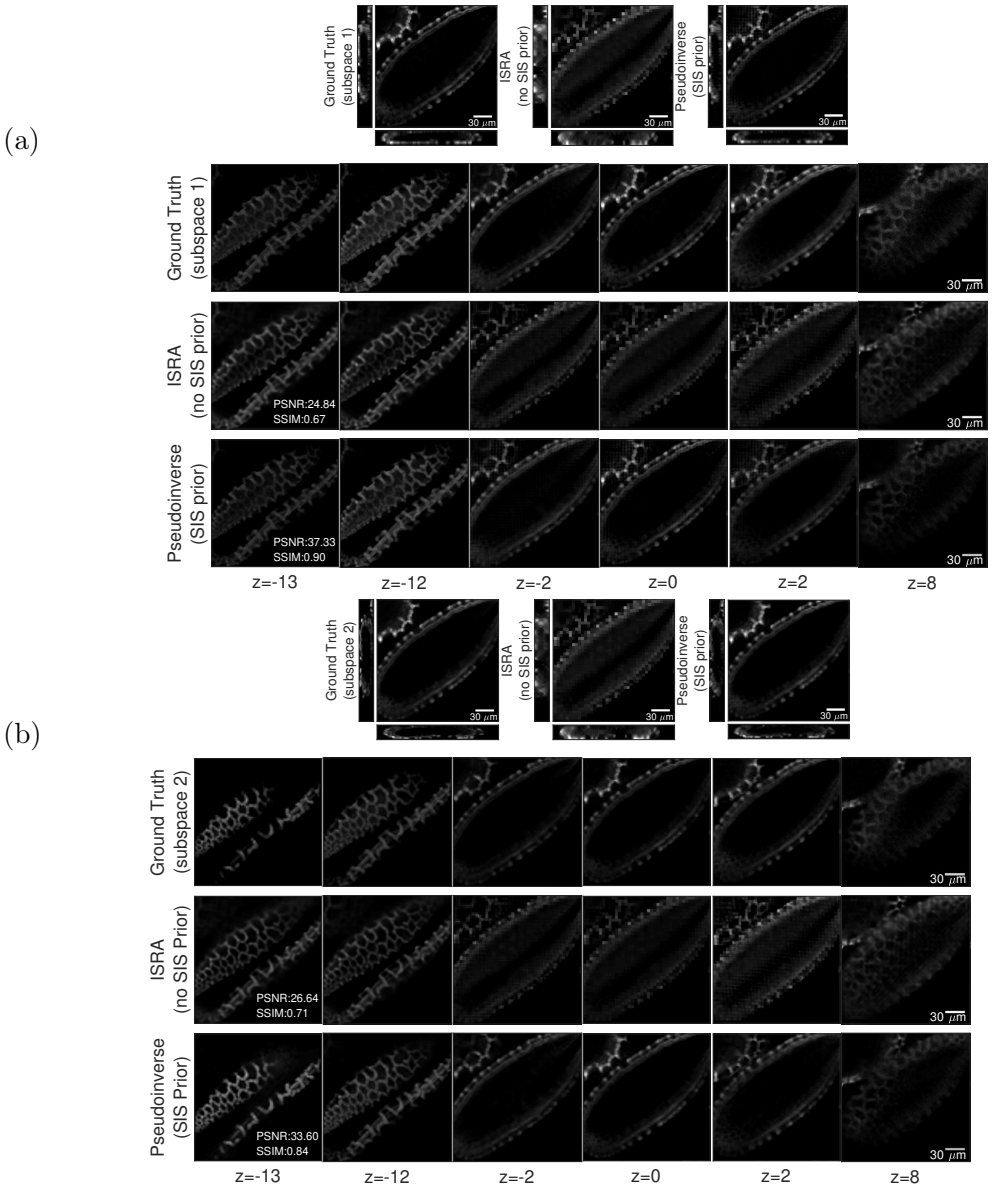


Figure 4.2: Ideal reconstruction using the pseudoinverse on synthetic data. In this experiment, the ground truth volume satisfies the SIS assumption. We show two volumes lying into 2 different subspaces as ground truth, subspace 1 (a) and subspace 2 (b). Then, we show the respective reconstruction without any prior assumption (using ISRA) and the reconstruction using the pseudoinverse with the prior assumption. Top, we show the in-focus plane, one xz, and one yz plane. Below, we show additional slices for different depths. All the distances are measured in μm. The volumetric data was taken from Lilium Longiflorum Pollen [69], while the light field images were simulated from the 3D pollen volumes.

lateral dimensions and 16 for the axial dimension. In the second case, scenario 2, we use the same filter type but a downsampling factor $\mathbf{r} = (8, 8, 8)$. For this experiment, we use publicly available 3D data: the Lilium Longiflorum Pollen [69]. This volume is projected

into the named subspaces to ensure the assumption holds. Figure 4.2 (a) shows that the ground truth volume for scenario 1 has a visually slightly higher lateral resolution than for scenario 2 due to the different downsampling factors $\mathbf{r}$ used for each subspace.

To compute $\mathbf{H}_\delta$ we use lateral and axial sampling intervals, Δx_1 and Δz, equal to $T/16$, and the output sampling interval Δx_2 defined by the pixel size of the sensor is equal to $T/19$, where T is the microlens pitch. Notice that our framework allows diverse sampling densities for both input and output, as opposed to [51]. Then, we compute the light field image using matrix $\mathbf{H}_\delta$ and attempt to recover the original volume using Equation (4.1). We clarify that the pseudoinverse was approximated iteratively since the closed-form expression of the pseudoinverse is not accessible.

Figure 4.2 shows the comparison between the proposed approach and the conventional reconstruction using ISRA. ISRA is a RL-like algorithm proposed in [63] for reconstruction in LFM. We show both lateral and axial slices of the volume. The reconstructed volume using the pseudoinverse and the SIS assumption matches the ground-truth volume very accurately, which is verified by the high Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Moreover, the reconstruction does not suffer from heavy square-like artifacts near $z = 0$ as in the standard reconstruction approach (ISRA). Note that the volume was reconstructed at the maximum sampling density $16/T$ for both scenarios. However, for scenario 1, the volumetric signal lies in a SIS with a shift equal to 0.25 times the microlens pitch, and for scenario 2, the volume lies in a SIS with a shift equal to 0.5 times the microlens pitch. This experiment shows that, for the proposed ideal scenarios, exploiting the underlying low-dimensionality of the volume allows artifact-free reconstruction with much higher accuracy than ISRA in terms of two metrics, SSIM and PSNR.

4.2 General Scenario and Additional priors

As mentioned previously, when no prior assumption is made about the input, the reconstruction suffers from square-like artifacts near the plane $z = 0$. This is explained in [51] and also verified in our experiments in the previous section. The reason is that a solution found using the pseudoinverse of $\mathbf{H}_\delta$ lies in the row space of matrix $\mathbf{H}_\delta$, which only contains

volumes with artifacts, due to the rank deficiency and the particular matrix structure for this depth.

Furthermore, during the volumetric reconstruction, commonly used algorithms such as ISRA and RL only impose mild priors (e.g., non-negativity and noise distribution), which are insufficient to regularize the solution. Therefore, we suggest incorporating more advanced priors as additional regularization to enforce the solution to live in a richer space far away from the row space of $\mathbf{H}_\delta$.

In particular, we propose two techniques for volume reconstruction. First, we propose to solve the following optimization problem:

$$\begin{aligned} \min_{\mathbf{f}} \quad & \|\mathbf{H}_\delta \mathbf{f} - \mathbf{g}\|_2^2 + \|\mathbf{D}_x^n \mathbf{f}\|_1 + \|\mathbf{D}_y^n \mathbf{f}\|_1 + \|\mathbf{D}_z^n \mathbf{f}\|_1 \\ \text{s.t.} \quad & \mathbf{f} \geq \mathbf{0}, \end{aligned} \tag{4.3}$$

where $\mathbf{f}$ is the desired volume, $\mathbf{g}$ is the light field image, and $\mathbf{D}_x^n$, $\mathbf{D}_y^n$ and $\mathbf{D}_z^n$ are the n-th order derivatives along each axis. The parameter n can be adjusted experimentally since the derivative order does not imply too much additional computation. Notice that if $n = 1$, the regularizer is an anisotropic total variation. However, this whole optimization is computationally demanding due to the high dimension of matrix $\mathbf{H}_\delta$. To alleviate this issue, we propose to use our approximation method based on SVD to compute the forward model.

As shown in the previous section, a SIS spanned by linear splines is able to reproduce typical structures that appear in the LFM image of a pollen volume. Considering this, we state the following optimization problem that exploits the SIS assumption to impose an additional constraint on the reconstructed volume:

$$\begin{aligned} \min_{\mathbf{f}} \quad & \|\mathbf{H}_\delta \mathbf{S}_\varphi \mathbf{A}_\varphi \mathbf{f} - \mathbf{g}\|_2^2 + \|\mathbf{D}_x^n \mathbf{f}\|_1 + \|\mathbf{D}_y^n \mathbf{f}\|_1 + \|\mathbf{D}_z^n \mathbf{f}\|_1 \\ \text{s.t.} \quad & \mathbf{f} \geq \mathbf{0}, \end{aligned} \tag{4.4}$$

where matrix $\mathbf{A}_\varphi$ and matrix $\mathbf{S}_\varphi$ have been defined before. Note that if the projection $\mathbf{S}_\varphi \mathbf{A}_\varphi$ is orthogonal, the transpose of the matrix $\mathbf{H}_\delta \mathbf{S}_\varphi \mathbf{A}_\varphi$ satisfies:

$$(\mathbf{H}_\delta \mathbf{S}_\varphi \mathbf{A}_\varphi)^\mathsf{T} = (\mathbf{S}_\varphi \mathbf{A}_\varphi)^\mathsf{T} \mathbf{H}_\delta{}^\mathsf{T} = \mathbf{S}_\varphi \mathbf{A}_\varphi \mathbf{H}_\delta{}^\mathsf{T}. \tag{4.5}$$

Therefore, the rows of the new measurement matrix span volumes that always lie

in the assumed subspace. This characteristic, together with the sparsity imposed into the n-th order derivatives, helps to reduce typical reconstructions artifacts. Furthermore, the dimension of the matrix $\mathbf{H}_\delta\mathbf{S}_\varphi$ is smaller than that of $\mathbf{H}_\delta$. This allows solving the problem by exploiting the underlying low-dimensional 3D space where $\mathbf{f}$ lives, thereby contributing to efficient computation.

Finally, to solve optimization problems (4.3) and (4.4), we require an algorithm algorithm that avoids strong artifacts from the computation of the transpose. To this end, we propose using ADMM since it allows splitting the original optimization problem into simpler sub-problems. Furthermore, ADMM allows detaching physical constraints to any prior imposed. However, conventional ADMM implementations also require applying the inverse of the system. For huge measurement matrixes, this step is usually approximated with iterative methods such as gradient descent or conjugate gradient [72]. To avoid time-consuming inner loops, we propose to adapt the inner-loop-free ADMM proposed in [72] for cryogenic electron microscopy to the light field problem, as shown in Algorithm 1. Note that Algorithm 1 can be implemented efficiently since it only needs the computation of the forward model, the adjoint (transpose), additional convolutions, and point-wise non-linear operators (soft-thresholding and rectifier linear units). Furthermore, Equation (4.6) is straightforward to solve in the Fourier domain.

Our reconstruction algorithm requires 6 input parameters. The learning rate α controls the gradient term in Equation (4.6), the importance values $\{\rho_x, \rho_y, \rho_z\}$ and ρ_+ control the weight of the corresponding term in Equation (4.6) and λ is defined as the threshold parameter for soft-thresholding, which is fixed for all the x, y and z axes.

4.3 Experiments and Results

In this section, we show experimental results using synthetic and real data. For synthetic data, we measure reconstruction performance using the peak signal to noise ratio (PSNR) and the structural similarity index measure (SSIM) since the ground truth is available in this case. For real data, we only show reconstructed results for visual comparison.

If not stated otherwise, the settings of the LF microscope used for both the simulated and the real data are set as: numerical aperture = 1, refractive index = 1.33, wavelength =

Algorithm 1 Volume reconstruction

Input : LF image $\mathbf{g}$, matrix block $\mathbf{B_H}$ for every depth, learning rate α, importance of derivatives $\{\rho_x, \rho_y, \rho_z\}$, non-negativity importance ρ_+, soft-threshold parameter λ, number of iterations K.

Output: Reconstructed Volume $\mathbf{f}^K$

$k = 0$

Initialize $\mathbf{f}^k$ and auxiliar variables $\{\mathbf{u}_i^k\}_{i=1}^4$, $\{\mathbf{z}_i^k\}_{i=1}^4$ to zero.

for $k < K$ **do**

$\mathbf{z}_1^{k+1} = prox_+(\mathbf{f}^k + \mathbf{u}_1^k)$

$\mathbf{z}_3^{k+1} = prox_{|\cdot|}(\mathbf{D}_y^n \mathbf{f}^k + \mathbf{u}_3^k; \lambda)$

$\mathbf{z}_2^{k+1} = prox_{|\cdot|}(\mathbf{D}_x^n \mathbf{f}^k + \mathbf{u}_2^k; \lambda)$

$\mathbf{z}_4^{k+1} = prox_{|\cdot|}(\mathbf{D}_z^n \mathbf{f}^k + \mathbf{u}_4^k; \lambda)$

$\mathbf{f}^{k+1} \leftarrow$ solve Equation (4.6) in the Fourier domain ($\mathbf{Hf}^{k+1}$ is computed from the $\mathbf{B_H}$ for every depth, possibly using our fast approximation)

$$\begin{aligned}(\mathbf{I} + \rho_+\mathbf{I} + \rho_x \mathbf{D}_x^{n\mathsf{T}}\mathbf{D}_x^n + \rho_y \mathbf{D}_y^{n\mathsf{T}}\mathbf{D}_y^n + \rho_z \mathbf{D}_z^{n\mathsf{T}}\mathbf{D}_z^n)\mathbf{f}^{k+1} \\ = \mathbf{f}^k - \alpha \mathbf{H}^\mathsf{T}(\mathbf{Hf}^k - \mathbf{g}) + \rho_+(\mathbf{z}_1^k - \mathbf{u}_1^k) \\ + \rho_x \mathbf{D}_x^{n\mathsf{T}}(\mathbf{z}_2^k - \mathbf{u}_2^k) + \rho_y \mathbf{D}_y^{n\mathsf{T}}(\mathbf{z}_3^k - \mathbf{u}_3^k) \\ + \rho_z \mathbf{D}_z^{n\mathsf{T}}(\mathbf{z}_4^k - \mathbf{u}_4^k) \quad (4.6)\end{aligned}$$

$\mathbf{u}_1^{k+1} = \mathbf{u}_1^k + \mathbf{f}^{k+1} - \mathbf{z}_1^{k+1}$

$\mathbf{u}_2^{k+1} = \mathbf{u}_2^k + \mathbf{D}_x^n \mathbf{f}^{k+1} - \mathbf{z}_2^{k+1}$

$\mathbf{u}_3^{k+1} = \mathbf{u}_3^k + \mathbf{D}_y^n \mathbf{f}^{k+1} - \mathbf{z}_3^{k+1}$

$\mathbf{u}_4^{k+1} = \mathbf{u}_4^k + \mathbf{D}_z^n \mathbf{f}^{k+1} - \mathbf{z}_4^{k+1}$

$k = k + 1$

end

490 nm, magnification (M) = 25, microlens pitch (T) = 125 μm, microlens focal length = 1250 μm, tube lens focal length = 0.18 m, pixels under each microlens =19. Furthermore, to compute $\mathbf{H}_\delta$ we set the lateral and axial sampling intervals, Δx_1 and Δz to $T/16$, and set the output sampling interval Δx_2 defined by the pixel size of the sensor to $T/19$, where T is the microlens pitch.

4.3.1 Synthetic Data

To validate our analysis, we perform two experiments using two types of 3D data. For the first experiment, we use the pollen structure [69] to simulate a LF image using matrix $\mathbf{H}_\delta$. The size of the synthetic LF is 779×779 pixels, while the original and reconstructed volumes are of size $656 \times 656 \times 96$ voxels (notice that $659 = 779 \times 16/19$). Then, from the LF image, we try to reconstruct the volume using five methods: (a) conventional ISRA [17], (b) ISRA with total variation, (c) artifact-free ISRA [64], (d) ADMM with sparsity regularization

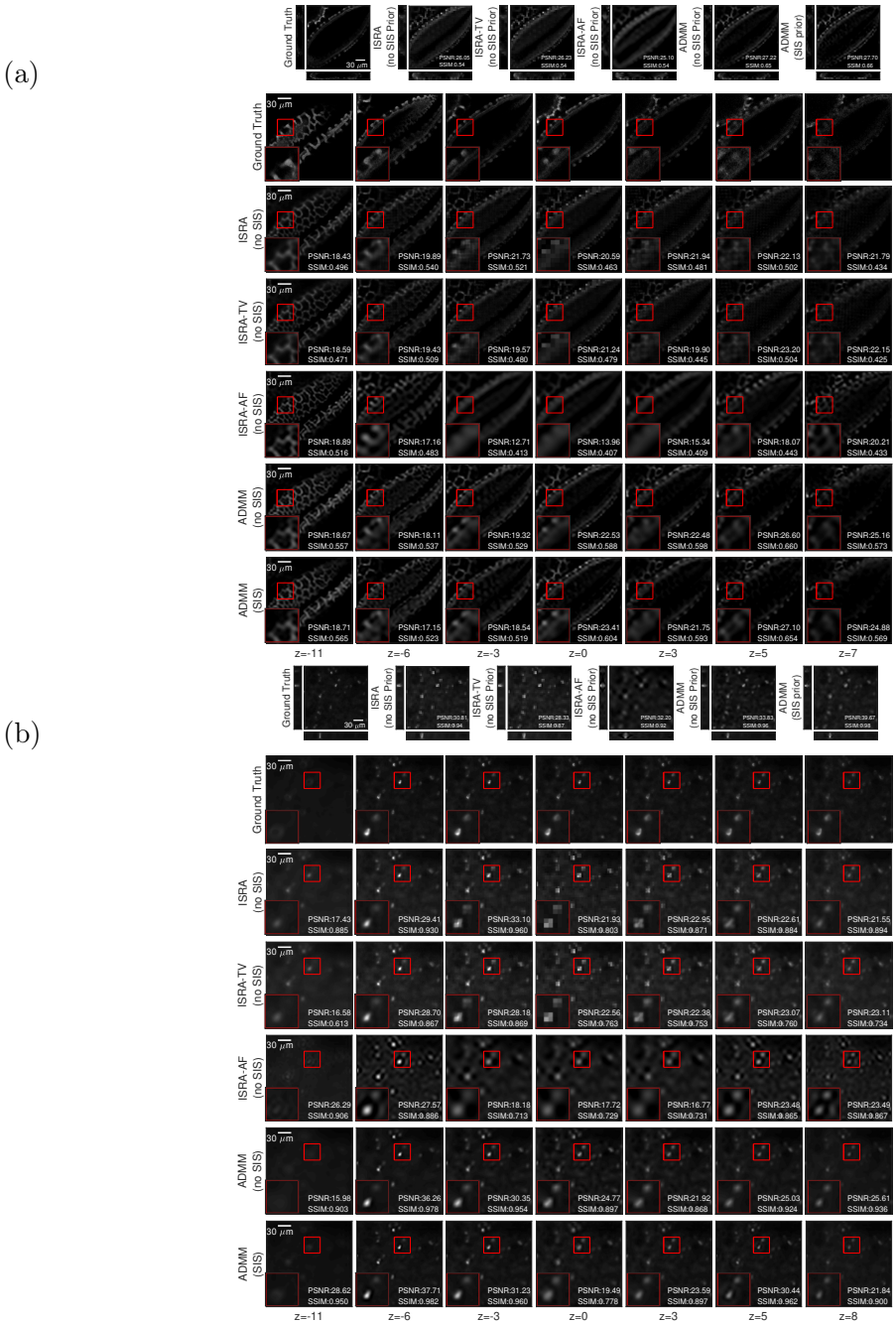


Figure 4.3: Reconstruction using synthetic LF data. Top, we show the in-focus plane, one xz, and one yz slice of the ground truth volume, the reconstruction using ISRA, ISRA with total variation [17], artifact-free (AF) ISRA [64], ADMM without SIS assumption, and ADMM with SIS assumption. The shown PSNR and SSIM correspond to the whole volume. Furthermore, below, we show additional slices for different depths. The shown PSNR and SSIM correspond to each slice. In this case, the ground truth volume does not satisfy the SIS assumption. We use the Lilium Longiflorum Pollen [69] dataset in (a), and a brain slice taken from an EGFP tagged triple transgenic mouse line in (b). The LF images were created synthetically from wide field volumes. All distances are measured in μm.

on the derivatives, as shown in Equation (4.3), and (e) ADMM with SIS assumption as well as sparsity regularization on the derivatives, as shown in Equation (4.4). For the latter case, we choose splines as template functions. Specifically, we choose the settings named 'scenario 1' in the previous section.The order of the derivative n in Equation (4.3) and (4.4) is set to 3. For all the three methods, the number of iterations is set to 24. In Appendix A.4 we discuss the use of a different stopping criteria.

Table 4.1: Performance of model-based methods on synthetic data.

	PSNR	SSIM	Time/Iteration (s)
	Pollen		
ISRA [18]	26.05	0.54	470
ISRA (Total Variation) [17]	26.18	0.53	520
ISRA (Artifact-free) [64]	25.10	0.54	485
ADMM	27.22	0.65	79
ADMM (SIS)	**27.70**	**0.66**	**32**
	Neurons		
ISRA [18]	30.81	0.94	471
ISRA (Total Variation) [17]	28.38	0.91	530
ISRA (Artifact-free) [64]	32.20	0.92	481
ADMM	33.83	0.96	68
ADMM (SIS)	**39.67**	**0.98**	**26**

We clarify that we use the same algorithm to solve both optimization problems (4.3) and (4.4). However, for the first case, our approximation method (3.18) is used to make the optimization computationally tractable, while for the second case, the approximation is not needed since $\mathbf{H}_\delta \mathbf{S}_\varphi$ has a smaller size than $\mathbf{H}_\delta$. In these experiments, we choose F in Equation (3.18) to achieve a matrix approximation error of 10%. Furthermore, note that in Algorithm 1, the additional computation performed includes convolutions and the adjoint operator of the system. The convolutions can be computed efficiently, and the adjoint operator can be found easily from the original measurement matrix using reflections and permutations. Therefore, Algorithm 1 can be implemented efficiently. Finally, the parameters α, ρ_x, ρ_y, ρ_z, ρ_+ and λ of Algorithm 1 can be found manually. In this section, we evaluate the effect of different choices of values on the performance of the method.

Figure 4.3 (a) shows the 3D structures recovered from a simulated LF image using the ISRA approaches and our method. Both the standard ISRA and the ISRA with total variation [17] introduce artifacts at planes near $z = 0$. In contrast, our ADMM

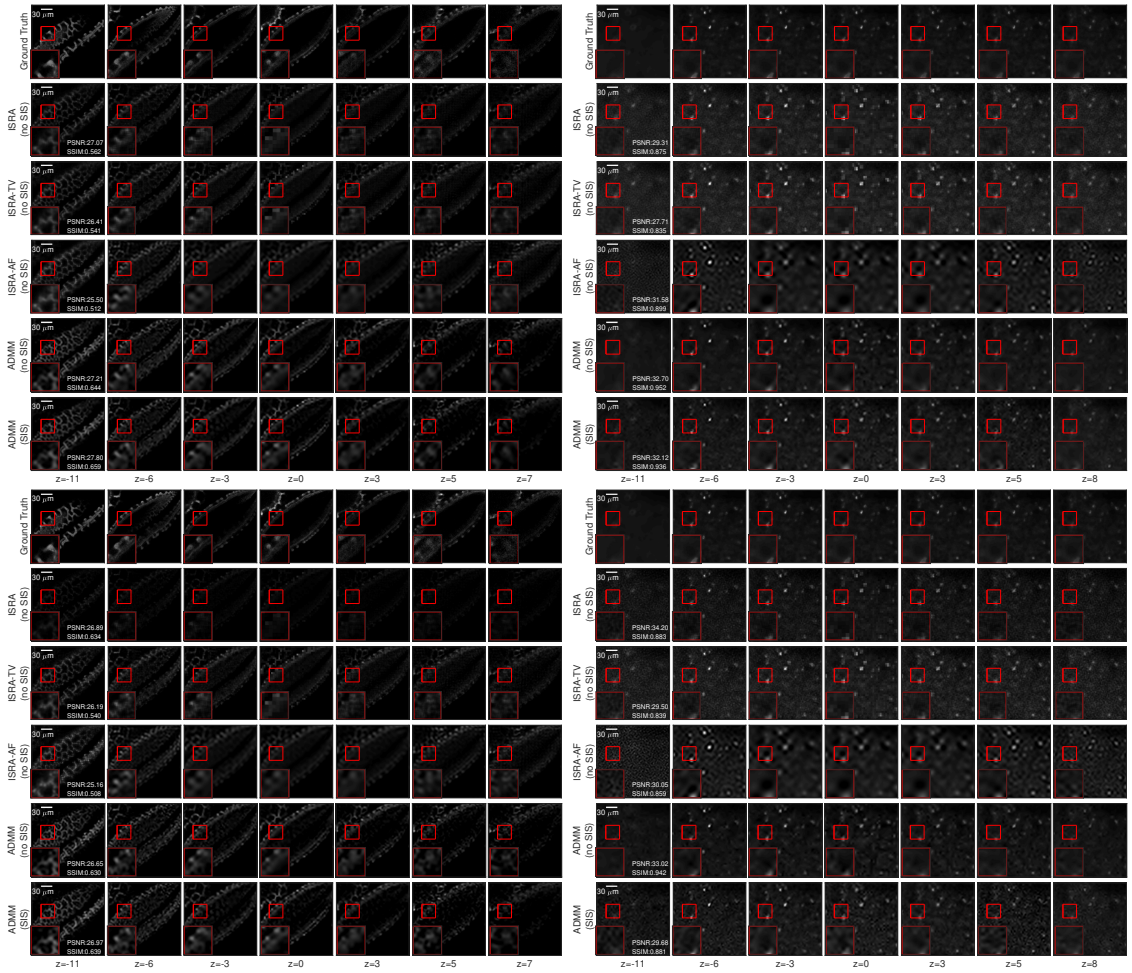
methods achieve artifact-free reconstruction. Even though the ISRA version proposed in [64] also achieves artifact-free reconstruction, the solution is over-smoothed near $z = 0$. Furthermore, our ADMM approaches give higher PSNR and SSIM than all the ISRA approaches. The ADMM with the SIS prior gives slightly higher PSNR and SSIM than the ADMM with no SIS assumption, as shown in Table 5.1.

Table 4.2: Performance of iterative methods with Poisson noise.

α_0	48	40	32	24	16	8	48	40	32	24	16	8
	Pollen											
	PSNR						SSIM					
ISRA [18]	27.17	26.89	26.35	27.16	27.07	26.73	0.62	0.63	0.54	0.57	0.56	0.55
ISRA (Total Variation) [17]	26.96	26.19	26.48	26.28	26.41	25.84	0.56	0.54	0.54	0.54	0.54	0.53
ISRA (Artifact free) [64]	26.77	25.16	25.92	26.03	25.50	25.71	0.55	0.51	0.52	0.53	0.51	0.52
ADMM	26.83	26.65	**27.11**	27.04	27.21	26.79	0.63	0.63	0.64	0.64	0.64	0.64
ADMM (SIS)	**27.82**	**26.97**	26.61	**27.27**	**27.80**	**27.67**	**0.64**	**0.64**	**0.66**	**0.65**	**0.66**	**0.66**
	Neurons											
	PSNR						SSIM					
ISRA [18]	**34.20**	27.93	32.51	32.61	29.31	33.79	0.88	0.80	0.88	0.90	0.88	0.94
ISRA (Total Variation) [17]	29.50	25.92	28.13	29.77	27.71	28.60	0.84	0.77	0.83	0.86	0.84	0.86
ISRA (Artifact-free) [64]	30.05	30.82	31.34	32.15	31.58	31.88	0.86	0.87	0.88	0.90	0.90	0.91
ADMM	33.02	**31.50**	**34.25**	**34.69**	**32.70**	34.36	**0.94**	**0.93**	**0.96**	**0.96**	**0.95**	**0.96**
ADMM (SIS)	31.96	29.68	32.88	34.60	32.12	**34.37**	0.90	0.88	0.92	0.95	0.94	**0.96**

For the second experiment, we imaged a 50-um-thick brain slice from a mouse line Ai90 [73] expressing an Enhanced Green Fluorescent Protein (EGFP) with a wide-field microscope (Figure 4.3 (b)). Then, we simulated LF images using the captured volume as ground truth data. For this type of data, we obtain similar results as in the previous case. Both ADMM methods achieve artifact-free reconstruction, while the standard ISRA and the ISRA with total variation [17] reconstructs squares near the plane $z = 0$, as shown in Figures 4.3 (b). The artifact-free ISRA [64] removes squares, but in this case the solution is distorted. Quantitatively, as shown in Table 5.1, the resulting PSNR and SSIM using the SIS assumption are slightly higher than those without the SIS assumption. Both ADMM methods outperform ISRA approaches in terms of PSNR and SSIM.

Furthermore, for the given settings, the ADMM methods are faster than the conventional methods. Each iteration of all the ISRA methods takes more than 470 seconds on average, whereas the ADMM without SIS prior takes 74 seconds, and ADMM with the SIS prior takes 29 seconds on average, as shown in Table 5.1. All the methods were tested in a CPU (Intel Core i7-6700, 16 GB RAM) using MATLAB R2018b. Further results



(a) (b)

Figure 4.4: Reconstruction using noisy synthetic LF data. Top we show reconstruction when the Poisson model in Equation (23) noise uses $\alpha_0 = 16$, and bottom we show reconstruction for $\alpha_0 = 40$. This figure shows slices for different depths of the ground truth volume, the reconstruction using ISRA, ISRA with total variation [17], artifact-free (AF) ISRA [64], ADMM without SIS assumption, and ADMM with SIS assumption. All distances are measured in μm. The shown PSNR and SSIM correspond to the whole volume. In this case, the ground truth volume does not satisfy the SIS assumption. We use the Lilium Longiflorum Pollen [69] dataset in (a), and a brain slice taken from an EGFP tagged triple transgenic mouse line in (b). The LF images were created synthetically from wide field volumes.

with synthetic LF created from mouse blood vessels [43] and mouse neurons co-expressing the functional calcium indicator GCaMP8f, and the structural marker tdTomato [74] are shown in Section 4.3.2.

Performance under noise. To further analyze the performance of our method we in-

cluded Poisson noise in the system as follows:

$$\mathbf{y} \sim \alpha_0 \text{Pois}(\frac{\mathbf{H}_\delta \mathbf{x}}{\alpha_0}), \tag{4.7}$$

which means that $\mathbf{y}$ is drawn from a Poisson distribution with mean $\frac{\mathbf{H}_\delta \mathbf{x}}{\alpha_0}$. The scalar α_0 controls the amount of noise (larger α_0 means more noise). Poisson noise is usually a source of noise related to the independence of photon detection and the quantized nature of light [75]. However, we experimentally found that light scattering is dominant when imaging mammalian brain tissue. Thus, we also evaluated our method under real acquisition conditions when imaging mammalian brain tissue, as shown in Section 4.3.3.

We use the same datasets used in the previous section for this experiment. We synthetically created LF images with different noise levels by varying the scalar α_0 from 8 to 48, as shown in Table 4.2. We found that both ADMM methods can manage adequately low to moderate noise conditions. Furthermore, the ADMM methods achieve the highest PSNR and SSIM for most cases. Specifically, for the pollen data, the ADMM with SIS prior consistently gives the best performance, as shown in Table 4.2. Visual results for the pollen dataset are shown in Figure 4.4(a). We show results for reconstruction that corresponds to $\alpha_0 = 16$ and $\alpha_0 = 40$. For both levels of noise, our ADMM with SIS assumption achieves the best performance. For this dataset, it is found that quantitatively, the original ISRA is more robust than the artifact-free and TV versions even though it introduces strong artifacts near the in-focus plane.

Similarly, for the neuron data, the ADMM without the SIS assumption achieves the best performance in most cases. Visual results for the neuron dataset are shown in Figure 4.4(b). Moreover, for this dataset, the SIS assumption only provides the best PSNR for minimum noise levels ($\alpha_0 = 8$). Furthermore we found that the artifact-free ISRA achieves the best quantitative performance among the ISRA methods for this dataset. However, qualitatively Figure 4.4 (b) shows that both the artifact-free ISRA and ISRA-TV are severely affected by noise. In particular, the noise is highly detrimental at depths $z = -11\mu m$, $z = 5\mu m$, and $z = 8\mu m$. In contrast our ADMM method without the SIS assumption can visibly tolerate noise for both $\alpha_0 = 16$ and $\alpha_0 = 40$. Furthermore, Section 4.3.3 shows that our methods also exhibit satisfactory results for real noisy acquisition

conditions.

Ablation Study. To evaluate the robustness of our algorithm, we evaluated the PSNR achieved with our ADMM methods when varying each hyper-parameter in Algorithm 1 independently, while the rest of the parameters were kept fixed using the values shown in Table 4.3. As mentioned previously, these hyper-parameters were found manually. We use the Pollen dataset in these experiments. Note that the last column in Table 4.3 shows the PSNR achieved by the corresponding configuration.

Table 4.3: ADMM hyper-parameters.

	α	ρ_+	ρ_x, ρ_y	ρ_z	λ	PSNR
(No SIS)	2×10^{-6}	1.21	1000	100	5×10^{-4}	27.22
(SIS)	2.5×10^{-6}	1.21	2000	0	5×10^{-4}	27.70

Table 4.4: Hyper-parameter study (PSNR).

		α	ρ_+	ρ_x, ρ_y	ρ_z	λ
		$[0.3, 3] \times 10^{-6}$	$[1.2, 12]$	$[250, 2500]$	$[0, 100]$	$[0.15, 15] \times 10^{-4}$
ADMM	Min.	24.91	24.32	27.32	25.84	27.31
	Max.	27.48	27.47	27.47	27.47	27.48
ADMM(SIS)	Min.	25.67	25.98	27.47	27.50	27.62
	Max.	27.82	27.86	27.71	27.71	27.71

We tested the 6 parameters of our algorithm: learning rate α, importance of derivatives $\{\rho_x, \rho_y, \rho_z\}$, non-negativity importance ρ_+ and the threshold λ for soft-thresholding. Note that we fixed ρ_x and ρ_y to be the same.

In our experiments, we found that the maximum PSNR achieved by the ADMM with the SIS prior is always larger than the maximum PSNR achieved by the ADMM without the SIS prior, as shown in Table 4.4. These results coincide with the results in Table 5.1. Furthermore, the PSNR shows no significant variations when the hyperparameters are modified within the ranges in Table 4.4. We also note that, the non-negativity importance ρ_+ and the learning rate α have a greater impact on the reconstruction.

4.3.2 Additional Datasets

In this section, we perform volume reconstruction using different settings for the LF microscope. Furthermore, two additional datasets are used for reconstruction: (1) Mouse

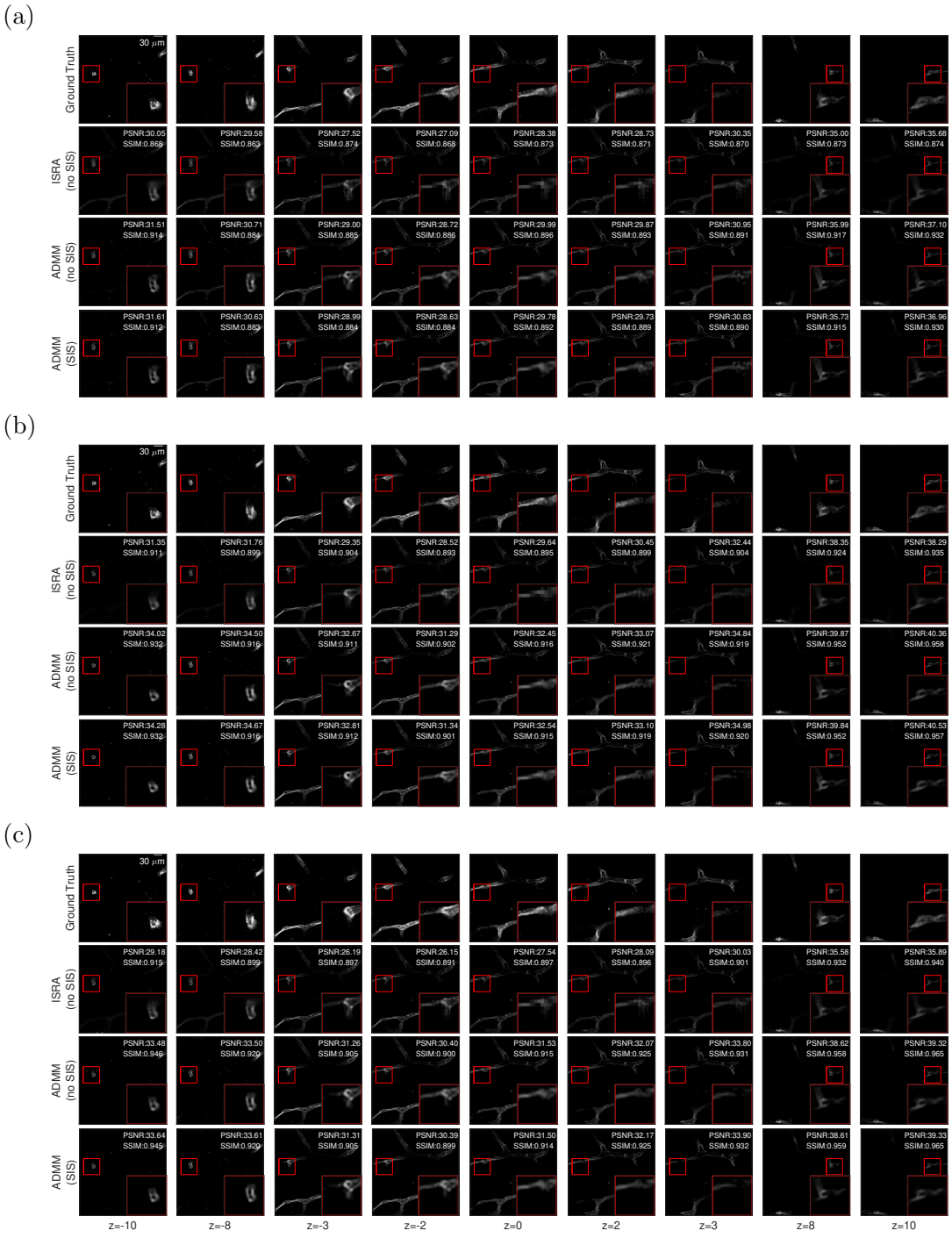


Figure 4.5: Reconstruction using synthetic LF images computed from labelled blood vessels captured using confocal microscopy from [43]. We show slices for different depths using ISRA, ADMM without SIS assumption, and ADMM with SIS assumption. The shown PSNR and SSIM are computed for each slice. Labels (a), (b) and (c) indicate the configuration used for the microscope, as explained in Section 4.3.2

.

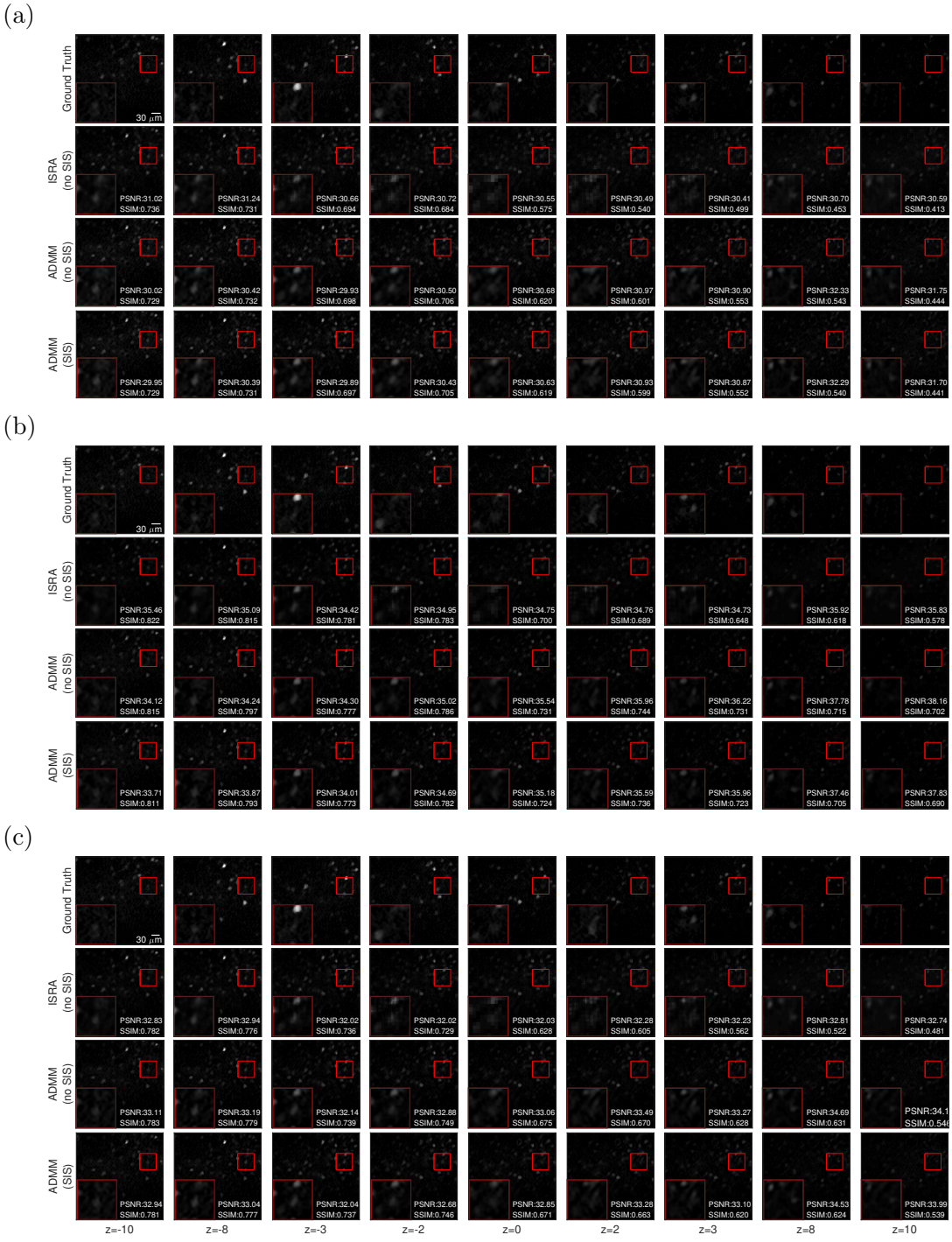


Figure 4.6: Reconstruction using synthetic LF images computed from mouse brain slices captured using our two-photon microscope. We show slices for different depths using ISRA, ADMM without SIS assumption, and ADMM with SIS assumption. The shown PSNR and SSIM are computed for each slice. Labels (a), (b) and (c) indicate the configuration used for the microscope, as explained in Section 4.3.2

brain slices with fluorescently labelled blood vessels captured using confocal microscopy from [43] and (2) Neurons from mouse brain slices expressing calcium indicator GCaMP8f taken with our two-photon microscope.

From the 3D volumes we created synthetic LF images under 3 different configurations as follows:

(a) Numerical aperture = 1, refractive index = 1.33, wavelength = 490 nm, magnification = 25, microlens pitch = 125 μm, microlens focal length = 1250 μm, tube lens focal length = 0.18 m, pixels under each microlens =19.

(b) The wavelength is changed to 660 nm while the rest of the parameters are the same as in (a).

(c) The wavelength is set to 488 nm, the numerical aperture is set to 0.5, the magnification is 22.5, the microlens focal length is 3125 μm, and the tube lens focal length is 0.3 m, while the other parameters are the same as in (a).

Table 4.5: Performance of iterative methods with different microscope settings.

$LFMSettings$	(a)	(b)	(c)	(a)	(b)	(c)
	Blood Vessel					
	PSNR			SSIM		
ISRA [18]	30.74	33.95	30.37	0.867	0.915	0.915
ADMM	**31.91**	35.43	34.30	**0.905**	**0.938**	0.939
ADMM (SIS)	31.76	**35.52**	**34.32**	0.904	0.937	**0.940**
	Two-Photon Neurons					
	PSNR			SSIM		
ISRA [18]	30.40	35.01	32.33	0.570	0.706	0.631
ADMM	**30.94**	**36.10**	**33.44**	**0.635**	**0.789**	**0.706**
ADMM (SIS)	30.90	35.77	33.29	0.634	0.782	0.702

For this experiment, we simulate a LF image using the measurement matrix named $\mathbf{H}_\delta$. The size of the reconstructed volume is 1072x1072x92, while the LF image is of size 1273x1273. We reconstruct the volume using the classic ISRA and our ADMM methods. For the ADMM with the SIS assumption, we choose splines as template functions. We choose the settings named 'scenario 1' in Section 4.1. For all the three methods, the number of iterations is set to 24.

In Figure 4.5, we show the results for the mouse blood vessels. Notice that the classic reconstruction introduces the typical square-like artifacts near the native object plane while

our ADMM methods remove artifacts. Similar artifact reduction is shown for the mouse neurons in Figure 4.6. In addition, the PSNR and SSIM for our ADMM methods are higher than those from classic ISRA, as shown in Table 4.5. In Figure 4.5 and Figure 4.6, we compute the PSNR and SSIM per depth. Notice that our ADMM methods achieve better PSNR and SSIM for most depths. In particular, there is a significant improvement close to the in-focus plane. Furthermore, our experiments found that the microscope configuration can affect the obtained PSNR and SSIM (in all the reconstruction methods). In particular, the configuration named (b) allows higher PSNR and SSIM than other configurations. For the three proposed configurations, our approaches perform better than classic ISRA.

4.3.3 Real Data

In this section, we evaluate the proposed approaches using real LF data. The experimental settings we use for the ADMM algorithms are the same as those in the previous section. The size of the real input LF image is 2033×2033 pixels. From this single image, we reconstruct a volume of size $1712 \times 1712 \times 96$ voxels covering a depth range $-15\mu m < z < 15\mu m$.

The LF image is captured from the same 50-μm-thick mouse brain slice of the previous section. We show the LF image in Figure 4.7 (a). We also capture the corresponding volume with the microscope in wide-field modality by scanning it along the z-axis without the microlens array. This wide-field image stack provides an approximation of the target volume we aim to reconstruct. The in-focus plane, together with two axial slices, are shown in Figure 4.7 (a). Furthermore, different lateral slices corresponding to seven depths are shown in Figure 4.7 (b). Notice that the non-ideal optical system and the specimen type cause increased noise in the images. To avoid noise amplification, the number of iterations used for reconstruction must be chosen properly. As proposed in previous works [17], [64], [18], a typical empirical number of iterations used for ISRA is between 8 and 10. We fixed this value to 10 for ISRA, and keep 24 iterations for our ADMM methods.

We compare the conventional method ISRA with the ADMM methods with and without SIS prior. In this case, we only present visual results since there is no ground truth available. As shown in Figure 4.7, both ADMM methods give a better reconstruction performance than ISRA in terms of artifacts reduction. In particular, these artifacts are strongly present at $z = 0\ \mu m$, but also at other slices close to the in-focus plane, such as

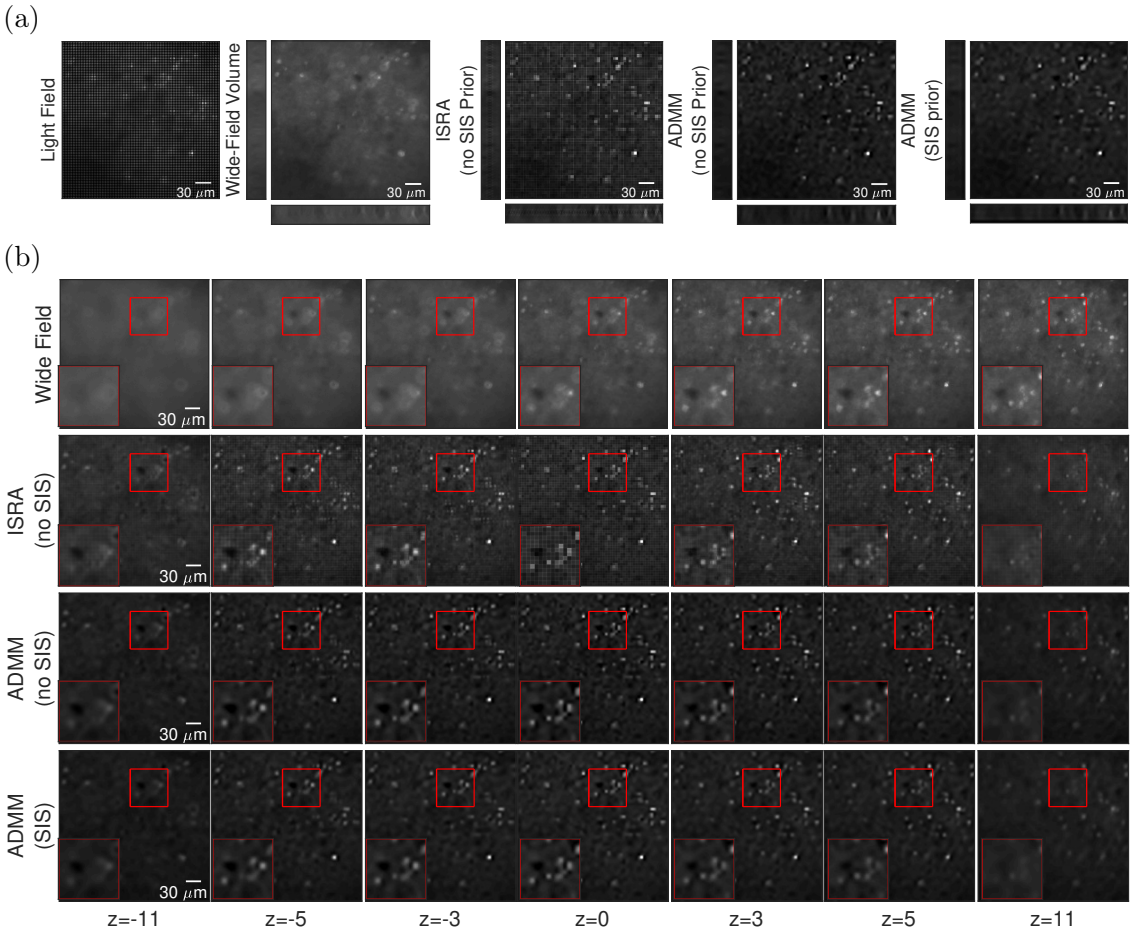


Figure 4.7: Reconstruction using real LF data from a brain slice (50 μm thick) imaged from an EGFP tagged triple transgenic mouse line. In part (a), from left to right, we show the original LF image; then, for comparison, we show a wide-field image taken without the microlens array; finally, the reconstruction using ISRA, our ADMM without and with the SIS assumption. We show the in-focus plane and two xz and yz slices below and to the left of each lateral slice, respectively. In part (b), we show additional slices for different depths. All the distances are measured in μm. The settings used to capture both the LF image, and the wide-field image are specified in Section 4.3.

$z = 3\ \mu m$ and $z = -3\ \mu m$. In general, comparable performance is achieved for depths far from $z = 0\ \mu m$, but artifacts near the native object plane are clearly removed when using the proposed ADMM methods.

For the second experiment, we use a LF image captured from acute mouse brain slices expressing the calcium indicator NIR-GECO2G [76], [77]. Furthermore, the size of the input LF is 1881×1881 pixels. From this single image, we reconstruct a volume of size $1584 \times 1584 \times 96$ voxels covering a depth range $-30\mu m < z < 30\mu m$.

Figure 4.8 (a) shows the LF image and the corresponding wide-field image stack

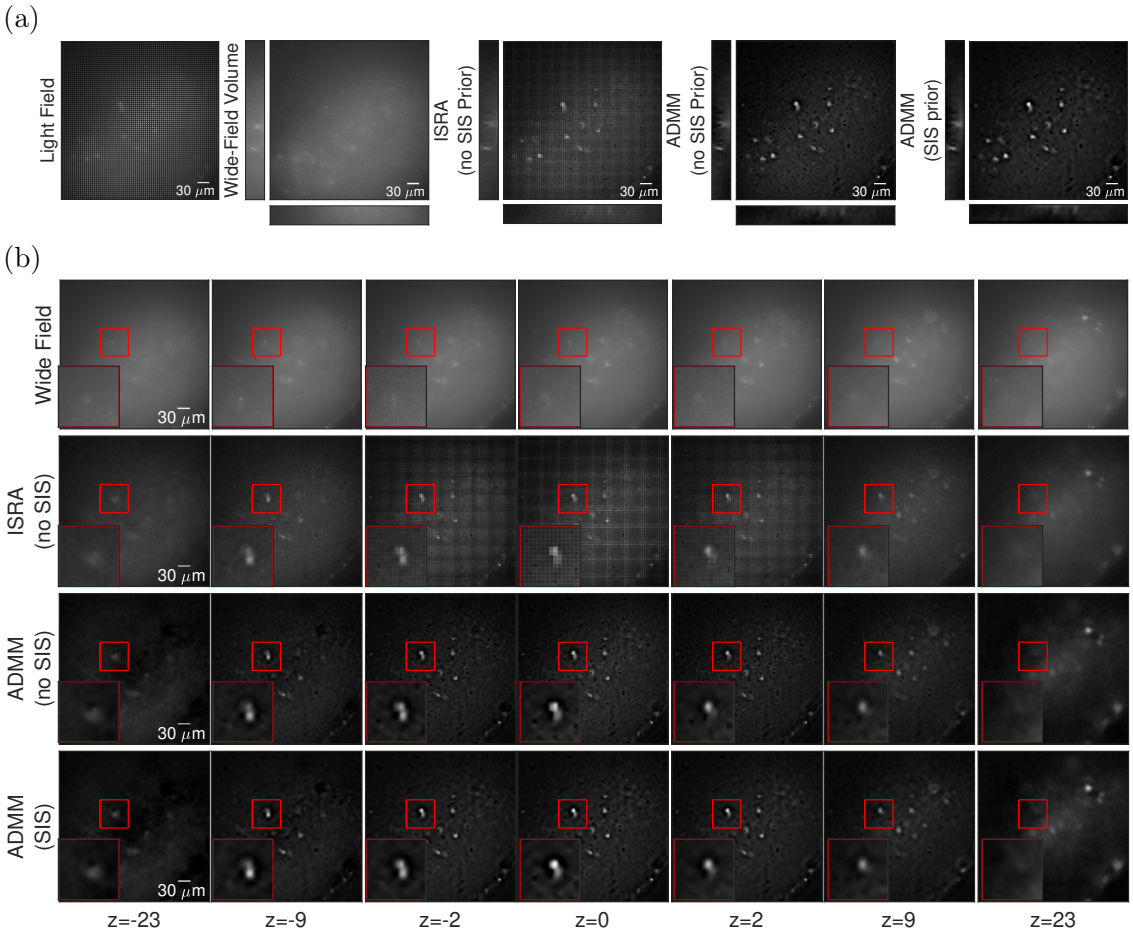


Figure 4.8: Reconstruction using real LF data from acute mouse brain slices expressing the calcium indicator NIR-GECO2G [76], [77]. In part (a) from left to right, we show the original LF image, a wide-field image taken with the microscope, the reconstruction using ISRA, and our ADMM without and with the SIS assumption. We show the in-focus plane and two xz and yz slices below and to the left of each lateral slice, respectively. In part (b), we show additional slices for different depths. All the distances are measured in μm. The settings used to capture both the LF image, and the wide-field image are specified in Section 4.3

(in-focus plane with two axial slices) captured from the same volume. Furthermore, different lateral slices corresponding to seven depths of the captured volume are shown in Figure 4.8 (b). In this case, we find similar results as in the previous example. We observe that both ADMM methods behave similarly but remarkably remove artifacts near the native object plane, specifically at depths $z = -2\ \mu m$, $z = 0\ \mu m$ and $z = 2\ \mu m$. ISRA instead produces strong square-like artifacts in these planes.

Finally, we show how our method adapts to different types of data by performing reconstruction using the LF data provided in [67]. For this experiment, part of the previous

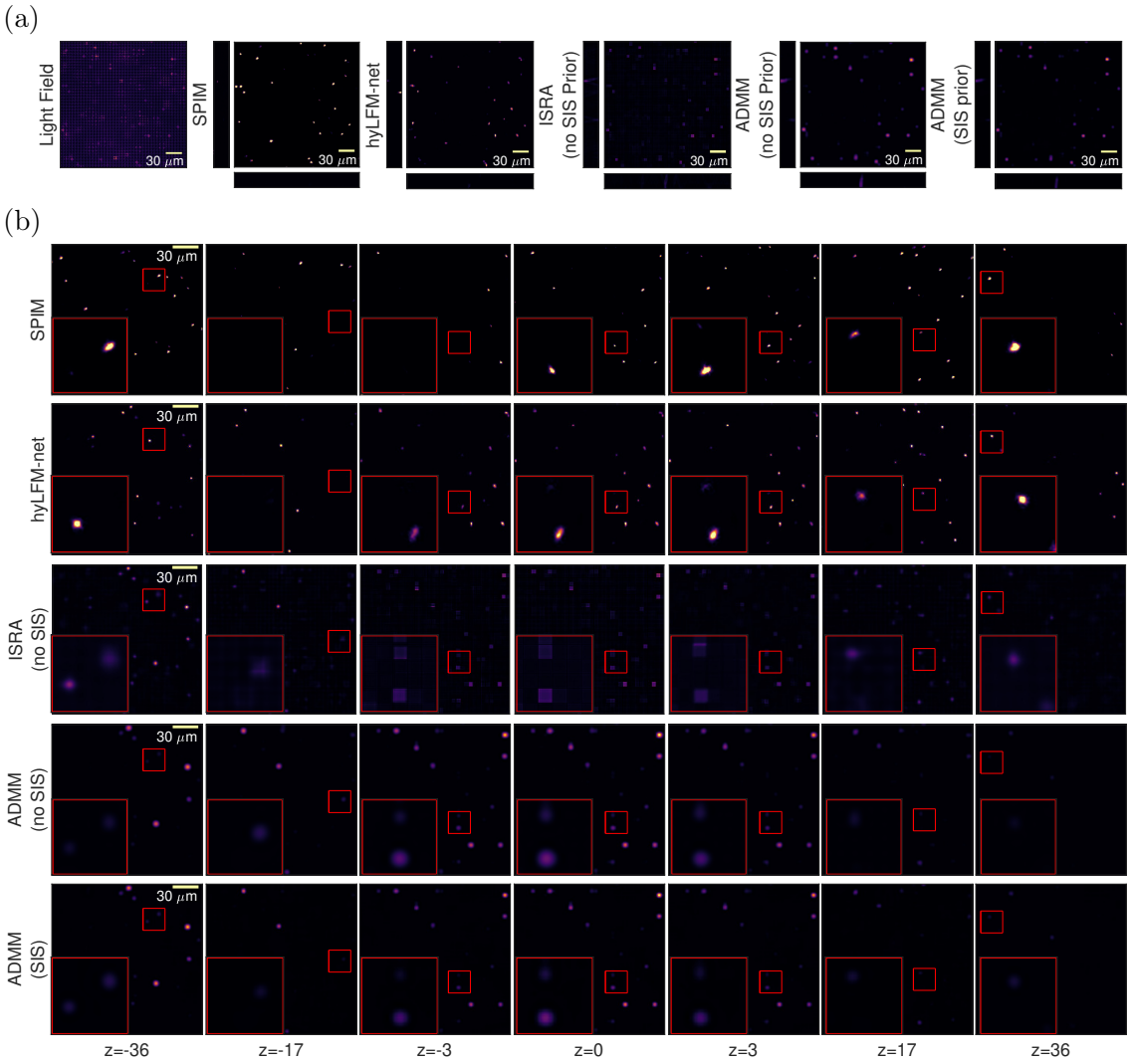


Figure 4.9: Reconstruction using real LF data from beads taken from [67]. In part (a), from left to right, we show the original LF image, a SPIM image, the reconstruction using the HyLFM-Net [67], the reconstruction using ISRA, and finally, our ADMM without and with the SIS assumption. We show the in-focus plane and two xz and yz slices below and to the left of each lateral slice, respectively. In part (b), we show additional lateral slices for different depths. All the distances are measured in μm. The settings used to capture the LF image are specified in Section 4.3 according to [67].

microscope settings is modified. According to [67], the wavelength is set to 488 nm, the numerical aperture is 0.5, the magnification is 22.5, the microlens focal length is 3125 μm, and the tube lens focal length is 0.3 m. From a single LF of size 931×931 pixels, we reconstruct a volume of size $752 \times 752 \times 96$ in the range $-50\mu m < z < 50\mu m$. The image corresponds to a group of beads, as shown in Figure 4.9 (a). Furthermore, in this case, a high-quality image stack of the volume obtained using selective plane illumination

microscopy (SPIM) is available. Additionally, the reconstruction obtained using the deep network hylfm-net from [67] is also available.

The results obtained using LF data from [67] are consistent with our previous conclusions. Figure 4.9 (a) and 4.9 (b), show that ISRA produces square-like artifacts near the native object plane while both ADMM methods do not introduce noticeable artifacts. In addition, this figure show that the hyLFM-net can reconstruct sharp beads that are visually similar to the SPIM images because SPIM volumes were used as labels in the training stage. However, when the trained hyLFM-net [67] and VCDNet [65] were tested on synthetic LF images of pollens and neurons (shown in Figure 4.3), the performance of both networks degraded dramatically, as shown by the poor PSNR and SSIM scores shown in Table 5.2. Even though both learning approaches can perform well if tested with similar data as used in the training stage [67], [65], producing thousands of labelled data pairs for each type of sample is challenging and even unfeasible under many realistic conditions. Furthermore, in our scenario, a network trained on synthetic data will not perform well on real data without proper modelling of the background noise and light scattering, as opposed to [65]. This highlights the relevance of model-based methods in scenarios like those discussed in this chapter.

Table 4.6: Performance of learning methods on synthetic data.

	PSNR	SSIM
	Pollen	
VCDNet	22.76	0.49
hyLFMNet	22.84	0.57
	Neurons	
VCDNet	23.03	0.23
hyLFMNet	22.54	0.10

Deep-Learning prior. In this section, we perform additional experiments using a self-supervised learning technique that aims to solve problems related to the lack of data for training.

Specifically, we use a deep-learning prior for reconstruction [78]. In Section 4.2, we proposed to use a customized ADMM algorithm to perform volume reconstruction. However, ADMM is a general tool that can also be used to impose a deep prior. Therefore,

another approach to solve the LFM problem is to replace the proposed SIS prior with a deep-learning prior and use the ADMM algorithm for reconstruction, as proposed in [79] for medical image restoration. Thus, in this method, we aim to solve the inverse problem by assuming that the reconstructed volume is the output of a neural network with input $\mathbf{z}$. Furthermore, an additional isotropic total variation prior is imposed as follows:

$$\begin{aligned} \min_{\theta} \quad & \|\mathbf{H}_\delta f_\theta(\mathbf{z}) - \mathbf{g}\|_2^2 + \|\mathbf{t}\|_2, \\ \text{s.t.} \quad & \mathbf{D} f_\theta(\mathbf{z}) = \mathbf{t}, \end{aligned} \tag{4.8}$$

where $f_\theta(\mathbf{z})$ is a neural network with learnable parameters θ, $\mathbf{z}$ is a fixed random vector and $\mathbf{D}$ is the discrete gradient operator. To solve this problem we use ADMM, as proposed in [79]. The following sub-problems must be solved for each iteration k:

$$\begin{aligned} \theta^{k+1} \in \arg\min_{\theta} \; & \frac{1}{2}\|\mathbf{H}_\delta f_\theta(\mathbf{z}) - \mathbf{g}\|_2^2 + \\ & + \frac{\beta}{2}\|\mathbf{D} f_\theta(\mathbf{z}) - \mathbf{t}^k + \frac{\lambda^k}{\beta}\|_2^2, \end{aligned} \tag{4.9}$$

$$\begin{aligned} \mathbf{t}^{k+1} = \arg\min_{\mathbf{t}} \; & \|\mathbf{t}\|_2^2 + \\ & + \frac{\beta}{2}\|\mathbf{t} - (\mathbf{D} f_{\theta^{k+1}}(\mathbf{z}) + \frac{\lambda^k}{\beta})\|_2^2, \end{aligned} \tag{4.10}$$

$$\lambda^{k+1} = \lambda^k + \beta(\mathbf{D} f_{\theta^{k+1}}(\mathbf{z}) - \mathbf{t}^k), \tag{4.11}$$

where Equation (4.9) is solved inexactly by applying a prefixed number of iterations of a gradient-based method using Adam optimizer. Equation (4.10) is solved directly since it is a least-square minimization, and the last equation (Eq. (4.11)) simply updates the dual variable λ.

The deep-learning-prior approach is evaluated using the pollen dataset described in the previous section. For this experiment, $\mathbf{z}$ is a random vector sampled from a uniform distribution. The structure of the network $f_\theta(\cdot)$ is a 2D-UNet, as proposed in previous works [43], [65] for supervised learning for LFM.

In Figure 4.10, we show the results after solving Equation (4.8) iteratively. The recovered volume has limited quality (PSNR=23.47) and suffers from reconstruction artifacts at different depths, for instance, at $z = -2$ and $z = 9$. These artifacts are not

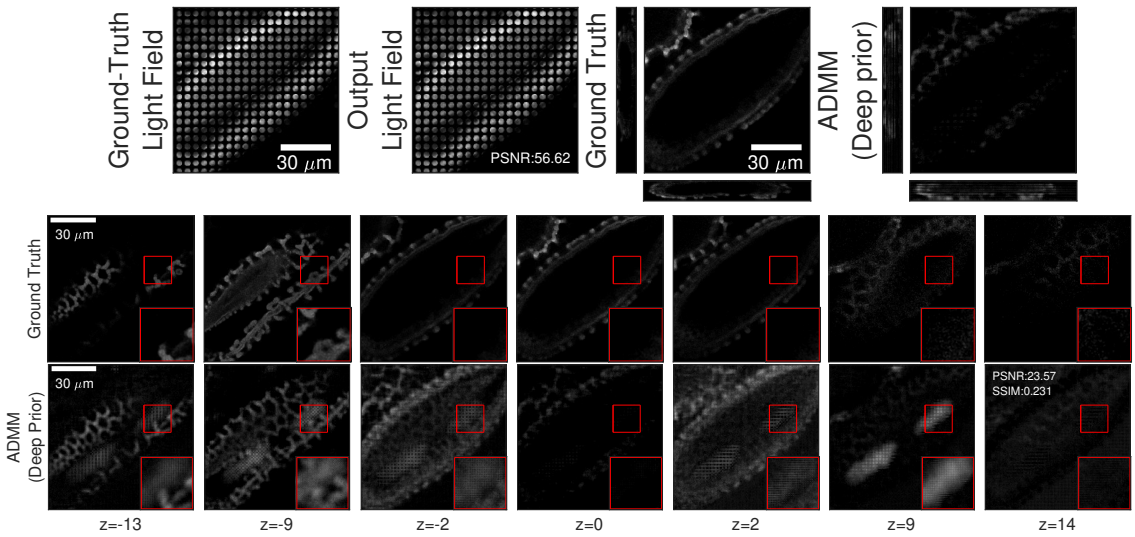


Figure 4.10: Reconstruction using ADMM with deep prior. Top, we show the in-focus plane, one xz, and one yz slice of the ground truth volume, the reconstruction using ADMM with deep prior. All distances are measured in μm. Furthermore, below, we show additional slices for different depths. The shown PSNR and SSIM correspond to the whole volume. We use the Lilium Longiflorum Pollen [69] dataset. The LF images were created synthetically from the 3D volumes.

the classic square-like artifacts close to the native object plane; they appear to be due to the structure of the network used. However, part of the volume structure is recovered at specific depths, for instance, at $z = -9$ and $z = -13$. As shown in Appendix A.4, in the same conditions our model-based ADMM achieves PSNR 31.32. Furthermore, note in Figure 4.10 that the LF image re-synthesised from the reconstruction is almost identical to the input LF image (PSNR=56.62) since the problem is highly ill-posed. Hence, the proposed priors and settings are not enough to avoid obtaining undesired solutions for this particular experiment.

In addition, we would like to emphasize that this approach is highly demanding in terms of computation since one artificial neural network must be “trained" for each reconstruction. Specifically, to obtain the result shown in Figure 4.10, we fixed the number of inner iterations to 100, and we ran 1000 iterations for the outer loop [79]. As mentioned in the introduction, LFM usually seeks to reconstruct volume time series. Therefore, reconstruction approaches for LFM should take computational complexity into account. Other approaches attractive for reconstruction are those exploiting denoisers for regularization, such as Plug-and-Play (PnP) [80] or Regularization by Denoising (RED) [81]. These ap-

proaches usually exploit state-of-the-art denoisers or pre-trained denoising networks acting as a regularizer in an optimization framework such as gradient descent or ADMM. However, typically available denoisers are not designed to remove artifacts exclusive to LF reconstruction or to work with 3D data in bio-imaging applications. Furthermore, the computation cost is still a drawback when reconstructing a stack of LF images since a 3D denoiser must be evaluated at every iteration. In our approach, we successfully achieved a balance between complexity and reconstruction performance.

4.4 Summary

We have experimentally shown that in an ideal scenario, perfect reconstruction can be achieved by using the pseudoinverse; however, this depends on the proper selection of the template functions and shifts defining the shift-invariant subspace. Furthermore, we propose a new 3D reconstruction algorithm for LFM with improved PSNR and SSIM, fewer artifacts, and faster speed (lower computational complexity) than conventional methods. This is achieved by incorporating additional priors and using a specific algorithm based on ADMM. The improvement in computation time is due to the exploitation of the underlying low-rank property of the measurement matrix by either using a specific SVD or assuming that the reconstructed volume lies in a SIS. Experimentation on both synthetic and real data demonstrate that our approach consistently outperforms conventional volume reconstruction approaches.

Chapter 5

Physics-based Deep Learning for Imaging Neuronal Activity via Two-photon and Light Field Microscopy

STUDYING the rapid dynamics of hundreds of neurons in brain tissue poses a challenge for conventional methods used in microscopy imaging. Typical optical techniques struggle to achieve simultaneous 3D imaging of multiple neurons since they focus on a single plane or point in space. In this chapter, we explore two-photon microscopy to regularize 3D reconstruction for light field microscopy.

Two-photon microscopy is a point-scanning imaging technique that achieves deep tissue penetration and low scattering levels. In traditional microscopy, visible light excites the atom to emit fluorescence in the visible wavelength range. In two-photon microscopy, a laser exits the atom with light with double the visible wavelength. Laser source in two-photon microscopy is in the infrared spectrum and has less energy per photon than visible light [82]. Thus, to make the labelled sample emit fluorescence, two photons must hit the molecule within a short time-lapse ($\sim$ 1 femtosecond). Light with a larger wavelength penetrates deeper in the tissue, is less prone to scattering and two-photon excitation makes it less likely to excite undesired regions in the tissue beyond the focal volume [21]. However, it also requires more power and expensive specialized equipment [21], [83]. Even though 3-

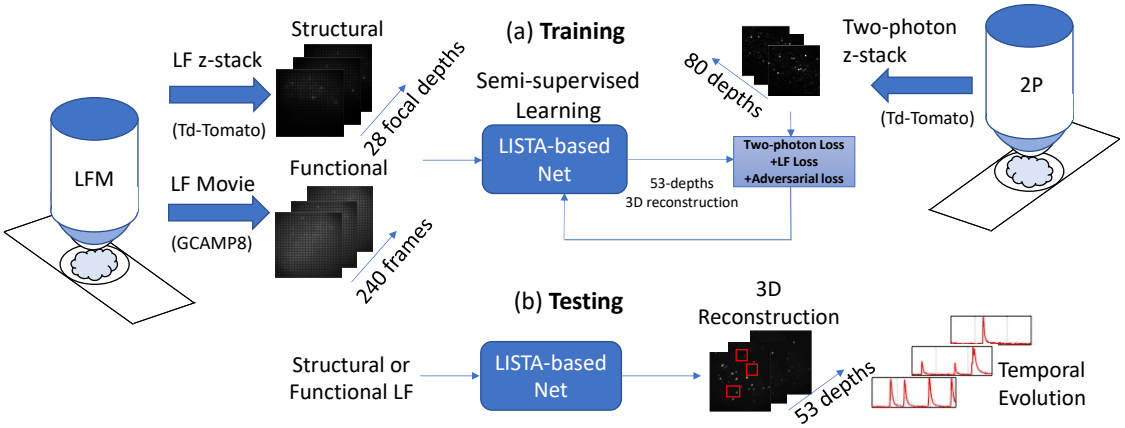


Figure 5.1: Overview of our approach. As shown in (a), we train our network using a small dataset of 28 training pairs. We effectively collect a 28-slice focal stack of LF images and the corresponding 2P 3D image of 80 depths for a single brain sample labelled with the TdTomato fluorophore. In addition, we use unlabelled data for training. We use three sequences of 80 frames each from 3 different samples for training. These brain samples are labelled with the jGCaMP8f protein which encodes calcium responses in the brain, which is an indirect measurement of electrical brain activity. The architecture of our network is based on the unfolding of the ISTA algorithm. We use a training loss that exploits the knowledge of the forward model and an adversarial regularizer. The testing is performed on LF stacks from unseen samples or LF sequences, and we produce one volume per frame from which we extract neuronal activity, as shown in part (b).

photon microscopy and onwards is possible, tissue damage is more likely to occur due to the higher density of photons [83]. Moreover, the point-scanning nature makes multi-photon microscopy unattractive for studying the fast dynamics of neurons.

Deep learning is a potential tool to enable efficient 3D imaging of large population of neurons. In particular, it has shown potential to alleviate typical problems in LFM [65], [43], [67]. However, current learning approaches for LFM are tested under idealized settings that are difficult to achieve in many realistic situations. For instance, when studying neuronal activity in mammalian brain tissue, the sample is highly scattering, non-transparent and contains high background noise, which makes training artificial neural networks (ANN) challenging. On the other hand, model-based optimization approaches have shown to be more robust under these adverse experimental conditions [59], [60].

This chapter proposes a novel multimodal imaging approach leveraging the respective strengths of 2P microscopy and LFM. We label neurons in the mouse brain tissue using two types of fluorescent proteins: TdTomato and jGCaMP8f. TdTomato captures the static neuron distribution in space, disregarding its activity. On the other hand, the

jGCaMP8f is an indicator of calcium concentration which indirectly measures electrical, and therefore functional, activity in the brain. In our setting, we capture the distribution of the neurons in a 400-μm-thick brain slice labelled with the TdTomato protein at high resolution using 2P microscopy. Similarly, the LF microscope captures the corresponding LF images for different focal depths. This approach gives us a labelled dataset. In addition, multiple LF temporal sequences are captured from different brain samples using jGCaMP8f protein. The small labelled dataset and a fraction of the LF temporal sequences are used to train our network in a semi-supervised manner, as shown in Figure 5.1. After training, our network can reconstruct volume time series from LF sequences with high accuracy and speed, despite the temporal LF sequences being obtained with the jGCaMP8f protein, for which we do not have the ground truth volume.

We achieve these results by introducing a physics-driven deep neural network whose architecture is driven by precise modelling of the forward model in LFM and the fact that labelled neurons in tissue are sparse. We leverage the sparsity assumption to design the architecture by unfolding the ISTA algorithm. Finally, a Generative Adversarial Network (GAN) ensures achieving 2P level resolution in the reconstruction while exploiting the knowledge of the forward model, as in model-based approaches. Overall, this approach allows us to exploit the best of two optical techniques to achieve the accurate and fast reconstruction of volume time series of neuronal activity in mammalian brain tissue.

5.1 Problem Formulation

As mentioned in Chapter 2, a LF microscope can be described as a linear operator. In most cases, it is safe to ignore non-linear effects such as occlusion or non-constant refractive indexes of the medium [51]. Therefore, after discretization, a monochromatic (plenoptic 1.0) LF microscope can be represented in matrix form as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{x}, \tag{5.1}$$

where matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ maps a vectorized volumetric input $\mathbf{x} \in \mathbb{R}^n$ into a LF image $\mathbf{y} \in \mathbb{R}^m$. The number n of voxels of the volume is usually much larger than the number

m of pixels of the LF image. In general, the size of $\mathbf{H}$ depends on the input and output sampling intervals, which are commonly chosen to be T/s and T/N, respectively (assuming unit lens magnification for simplicity). The constant T is the microlens pitch, s is an arbitrarily chosen upsampling factor, and N is the number of pixels under each microlens (per lateral axis).

The LF system is periodically shift-invariant per depth. If one shifts the input of the system laterally by T, the output is shifted by N pixels. This behaviour only occurs for shifts which are multiple of T. Therefore, for a fixed depth, the system can be modelled by using a filter bank. Furthermore, the impulse response is unique for each depth, which is the property that allows for localization of sources at different depths. Thus, to describe the whole system, we need one different filter bank per depth. The input volume first passes through a slicing operator S_i that selects the depth $z = i$ for $i = 1, 2, .., D$, where D is the number of depths. Then, each slice is the input to the corresponding filter bank, i.e., for each $z = i$, the i-th filter bank outputs a LF image $\mathbf{y}_i$. The final output of the microscope $\mathbf{y}$ is the summation of all the $\mathbf{y}_i$. Note that the structure of each filter bank is related to the measurement matrix of the system. Suppose there are $N \times N$ pixels under each microlens and the volume sampling interval for both lateral dimensions is T/s, in this case, the filter bank structure has $s \times s$ branches, the downsampling factor is s, and the upsampling factor is N. As explained in Chapter 4, the input and output filter of each branch can be obtained from the measurement matrix $\mathbf{H}$. Note that the convolutions and filters are two-dimensional. Similarly, the downsampling factor s and upsampling factor N refer to both lateral dimensions.

In this chapter, we aim to solve the inverse problem derived from Equation (5.1) using a deep-learning approach. Furthermore, studying the spatial and temporal behavior of neurons in brain tissue requires a reconstruction method that performs fast 3D reconstruction from LF sequences. The reconstruction of a 3D volume $\mathbf{x}$ from a single LF image $\mathbf{y}$ is traditionally solved using RL-like algorithms. Significant improvements in performance and speed have been achieved with model-based reconstruction, e.g [64], [84]. However, learning-based methods can potentially achieve better reconstruction quality and faster speed when trained properly in controlled scenarios.

5.2 Efficient Implementations of the Forward Model

As mentioned in Chapter 3, the LF system can be described by a linear CNN. This description is fundamental for the derivation and implementation of the reconstruction method shown in this chapter. In this section, we convert the filter-bank model to efficient CNNs architectures to accelerate the computation of the forward model. Due to memory and computational-complexity constraints, it is useful to find efficient implementations of the forward CNN.

In Chapter 3, we used a CNN with two convolutional layers to help map the forward model to the filter banks. This description is shown in Figure 5.2 (a). However, we can find other alternative architectures that simplify the implementation. First, the 3D input $\mathbf{x}$ is reshaped to obtain the new input $\mathbf{x}_r$, which has $D \times s \times s$ channels, as shown in Figure 5.2 (b). Then, one can replace the two convolutional layers in Figure 5.2 (a) with a single convolutional layer, as shown in Figure 5.2 (b). Furthermore, we propose two different simplifications of the architecture based on two observations:

(a) Convolutional layers with large filters can usually be well-approximated by a sequence of convolutional layers with smaller filters. Therefore, it is realistic to describe the original system with a series of convolutional layers, as shown in Figure 5.2 (d). We ensure this architecture uses fewer parameters than the single convolution and also ensure that the size of the equivalent filter is the same as the original one by choosing the filter size l and the number of channels c accordingly. Note that the weights of the network can be learned from the theoretical matrix $\mathbf{H}$. We use this architecture when we need to apply the forward CNN $f(\cdot)$ to a given input volume.

(b) The number of channel outputs greatly impacts the number of parameters. For instance, in our setting $N \times N = 19 \times 19$. Therefore, we can reduce the number of views to a smaller value V to reduce computational complexity. Then, to restore the original number of views ($N \times N$), we can add a linear convolutional layer with filters of unit size, as shown in Figure 5.2 (e). Since the sub-aperture images are usually highly correlated, it is feasible to perform this linear approximation without significantly impairing the accuracy of the model. The definition of the compressed forward CNN $h(\cdot)$ in Figure 5.2 (e), is relevant since it is connected to our reconstruction approach.

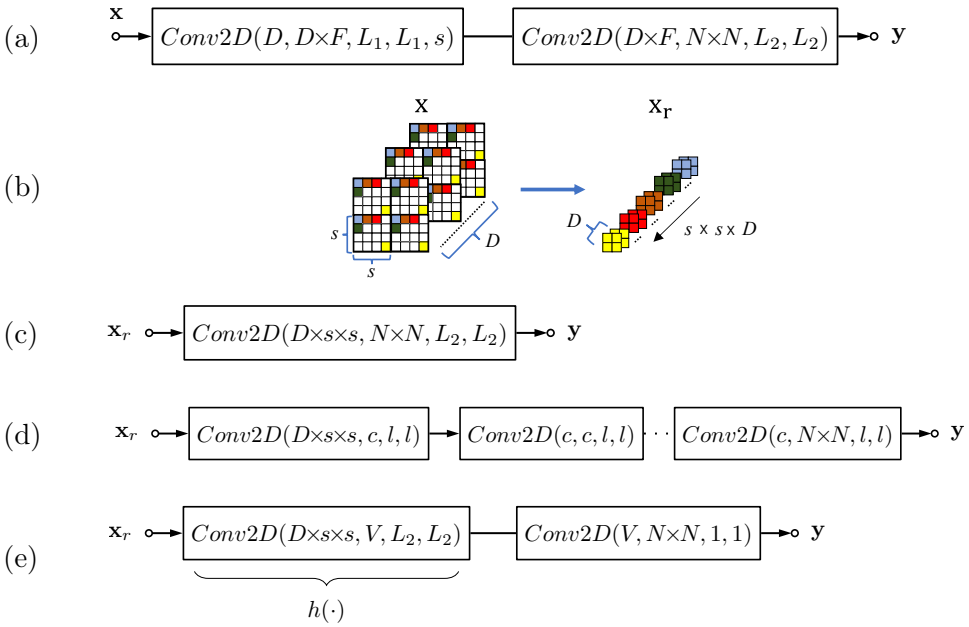


Figure 5.2: Forward model as a CNN. We show four different linear CNNs architectures that transform a 3D input image x with D depths into a LF image y with $N \times N$ views, where $N \times N$ is the number of pixels under each microlens. The model (a) is a novel form to described the forward model without losing accuracy. Note that in (a) the image x is input directly while in models (c)-(d) the input x is first reshaped to obtain $\mathbf{x}_r$, as shown in (b). In (c) we use a single layer to describe the system. The model in (a) can always be converted into the single convolutional layer shown in (c) while in (d) we approximate the system by a sequence of convolutional layers with filters of smaller size l. Finally, a second approximation is shown in (e), where the first layer $h(\cdot)$, named compressed forward CNN, outputs V channels, and the second layer recovers the $N \times N$ views with filters of unit size. The notation $Conv2D(\cdot, \cdot, \cdot, \cdot, \cdot)$ means a 2D convolutional layer with parameters ordered as follows: number of input channels, number of output channels, the height of the filter, width of the filter, and stride. If the stride is omitted, it means unit stride.

In the sequel we use $f(\cdot)$ to compute the forward model, while the compressed forward CNN $h(\cdot)$ is used in the reconstruction network. This will be clarified in the following section.

5.3 3D Reconstruction

In this section, we design a CNN that considers the physics of the system to perform the reconstruction. The architecture of our network is constructed using the unfolding technique [85] and is obtained by unrolling sparsity-driven algorithms for reconstruction. Furthermore, our network is trained in a semi-supervised manner to alleviate the lack of data which is a typical issue for applications in neuroscience.

5.3.1 Deep Neural Network Architecture for Volume Reconstruction

Large distributions of labelled neurons can be modelled as compact cell bodies sparsely distributed in brain tissue [19]. Therefore, to reconstruct high-quality 3D volumes, we can consider the following optimization approach that promotes sparsity in the reconstruction:

$$\arg\min_{\mathbf{x}} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \|\mathbf{x}\|_1, \tag{5.2}$$

where $\mathbf{y}$ is a given LF image, $\mathbf{x}$ is the reconstructed volume. This problem can be solved using the Iterative Shrinkage-Thresholding Algorithm (ISTA) [86] by computing at each iteration:

$$\mathbf{x}^{k+1} = \mathcal{T}_\lambda(\mathbf{x}^k - \mathbf{H}^\mathsf{T}\mathbf{H}\mathbf{x}^k + \mathbf{H}^\mathsf{T}\mathbf{y}), \tag{5.3}$$

where $\mathcal{T}_\lambda$ is the soft-thresholding operator with parameter λ. One can interpret each iteration of ISTA as a layer of a neural network with fixed weights. Therefore, it is possible to design a neural network architecture based on ISTA. LISTA [85] (the learned version of ISTA) is a neural network built such that each layer corresponds to one iteration of ISTA. Effectively, each layer of LISTA implements the following step:

$$\mathbf{x}^{k+1} = \mathcal{T}_\lambda(\mathbf{x}^k - \mathbf{H}_1^\mathsf{T}\mathbf{H}_2\mathbf{x}^k + {\mathbf{H}_3}^\mathsf{T}\mathbf{y}), \tag{5.4}$$

where $\mathbf{H}_1$,$\mathbf{H}_2$ and $\mathbf{H}_3$ are matrices of same size and structure as $\mathbf{H}$. These matrices are the parameters of the network that can be learned using a proper loss function. Note that, contrary to [85], we do not fuse the product $\mathbf{H}_1^\mathsf{T}\mathbf{H}_2$ into a single matrix since we want to keep the structure of each factor. This version of LISTA uses the soft-thresholding as the element-wise non-linearity due to the l_1 constraint in Equation (5.2). However, ISTA can be used with different types of non-linearities related to the prior imposed, as explained in [87]. For instance, replacing $\mathcal{T}_\lambda$ with a rectified linear unit (ReLU) imposes non-negativity, and replacing it with a ReLU with a bias term imposes sparsity and non-negativity. In our case, $\mathbf{x}$ is sparse and non-negative. Therefore, we propose a LISTA network that uses a ReLU with a bias term as non-linearity:

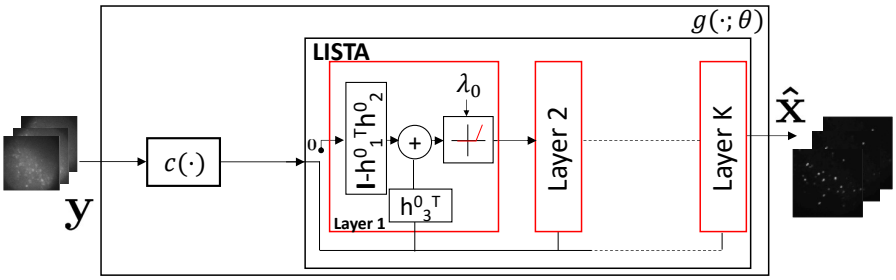


Figure 5.3: CNN architecture. Our reconstruction network $g(\cdot)$ is composed of (1) a compression layer $c(\cdot)$, which is a linear convolutional layer with $N \times N$ input channels and V output channels and (2) a LISTA network. At each layer of LISTA we use the architecture of the compress forward CNN $h(\cdot)$ shown in Figure5.2 and the adjoint operator $h^T(\cdot)$. The LISTA network is composed of K layers.

$$\mathbf{x}^{k+1} = ReLU(\mathbf{x}^k - \mathbf{H}_1^{\mathsf{T}^k}\mathbf{H}_2^k\mathbf{x}^k + \mathbf{H}_3^{\mathsf{T}^k}\mathbf{y} + \lambda^k), \tag{5.5}$$

where λ^k is a learnable bias. The custom $\{\mathbf{H}_i^k\}_{i=1}^3$ for each unfolded iteration k gives the network more capabilities without compromising its simplicity. Furthermore, the original LISTA with shared weights is a type of recurrent neural network (RNN) [88]. Therefore, it can suffer from typical RNN issues such as vanishing gradients, exploding gradients [89] and large memory consumption [90]. In particular, memory consumption during training is a major problem due to the LF data size. LISTA with custom weights for each layer helps to alleviate this issue.

In many practical cases, the described LISTA network cannot be used directly to solve the volume reconstruction problem. The size and structure of the matrix $\mathbf{H}$ make it computationally prohibitive to perform matrix multiplications repeatedly. Therefore, we propose using the compressed forward CNN $h(\cdot)$ proposed in Section 5.1 to reduce the computational complexity. The final architecture of our network is, therefore, described as follows:

$$\mathbf{x}^{k+1} = ReLU(\mathbf{x}^k - h_1^{k\mathsf{T}}(h_2^k(\mathbf{x}^k)) + h_3^{k\mathsf{T}}(c(\mathbf{y})) + \lambda^k), \tag{5.6}$$

where we have replaced matrices $\mathbf{H}_i^k$ in Equation (5.5) with the linear mappings $\{h_i\}_{i=1}^3$. The computation of all the $\{h_i\}_{i=1}^3$ is determined by the architecture of the compressed forward CNN derived from physics and explained in Section5.1. Note that the structure of

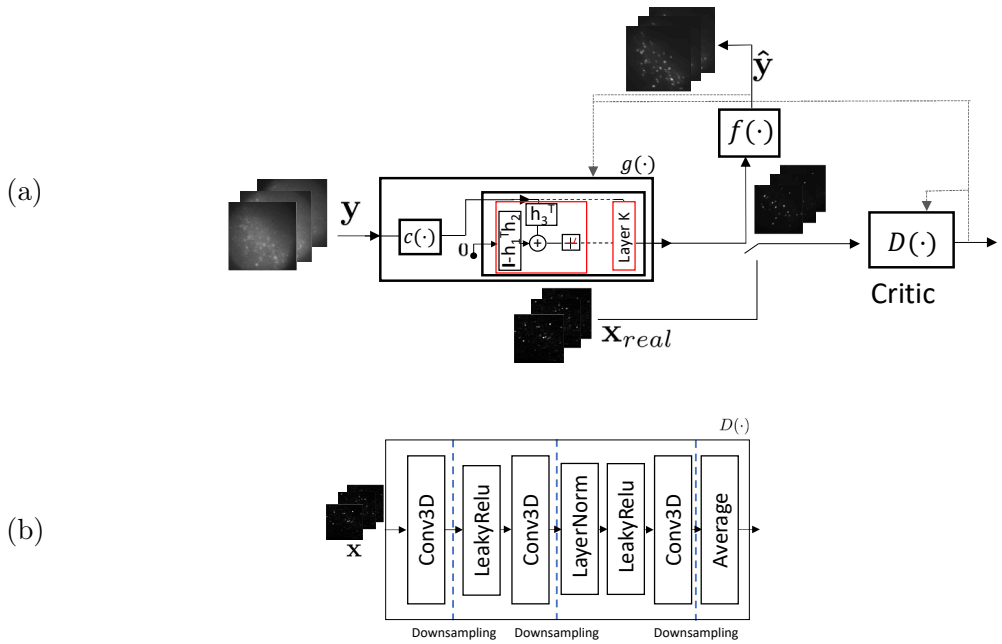


Figure 5.4: Training of our GAN architecture. In (a), we show how the LISTA network $g(\cdot)$ is trained using a content loss and an adversarial loss computed from a critic $D(\cdot)$. The content loss is computed using a few labelled data pairs, unlabelled LF data, and the known forward model $f(\cdot)$. In (b), we show the architecture of the critic $D(\cdot)$ designed following typical techniques for 3D GANs [91].

the adjoint operators (transpose) $\{h_i^{\mathsf{T}}\}_{i=1}^{3}$ in Equation (5.6) can be easily computed from the permutation of the weights of $h(\cdot)$. Furthermore, the input of the network is $c(\mathbf{y})$ rather than $\mathbf{y}$. The mapping $c(\cdot)$ is defined as a single linear convolutional layer with $N \times N$ input channels and V output channels and filters of unit size. By having V output channels, $c(\cdot)$ is compatible with the input size of the operators $\{h_i^{\mathsf{T}}\}_{i=1}^{3}$. For this compression step, we found unit-size filters to be effective; however, filters of any size could be used. We highlight that the coefficients of the compression layer $c(\cdot)$ are learned together with LISTA. The end-to-end network $g(\cdot;\theta)$, where θ represents the learnable parameters of the network, is shown in Figure 5.3. If additional simplification is needed, some convolutional layers in $g(\cdot)$ can be replaced by a sequence of convolutional layers with smaller filters.

5.3.2 Training Strategy

We learn the parameters θ of our LISTA network $g(\cdot;\theta)$ with a proper loss function and a mixture of labelled and unlabelled datasets. In our scenario, a labelled dataset comprises LF images and the corresponding 2P volumes. For many applications in LFM, capturing a

huge labelled dataset is too expensive or even unfeasible. For instance, when studying the behavior of neurons in mammalian tissue, capturing a clean 3D label is challenging due to the scattering media. Furthermore, using only synthetic data for training is problematic if noise and other imperfections are not appropriately modelled.

In our setting, we propose acquiring a very small labelled training dataset. We label neurons in a single brain sample using TdTomato fluorophore. The TdTomato allows capturing the static distribution of the neurons in space using both 2P and LF modalities. The 2P raster scanning modality provides the ground truth volume that can be paired with the LF images acquired with the same fluorophore. Therefore, to train LISTA we exploit the small labelled dataset, the large amount of unpaired LF images, and the knowledge of the forward model. The training loss is stated as follows:

$$\frac{1}{M}\sum_{i=1}^{M}\mathcal{L}_{c1}(\mathbf{x}^i,\hat{\mathbf{x}}^i)+\frac{1}{K}\sum_{j=1}^{K}\mathcal{L}_{c2}(\mathbf{y}^j,\hat{\mathbf{y}}^j)+\mathcal{L}_{adv}(g(\mathbf{y}^j)), \tag{5.7}$$

where $\mathbf{x}^i$ is the 2P 3D image, $\hat{\mathbf{x}}^i$ is the network reconstruction, $\mathbf{y}^i$ is a LF image, M is the number of 3D samples, K is the number of LF samples and $\hat{\mathbf{y}}^i = f(g(\mathbf{y}^i))$, where $f(\cdot)$ is the known forward CNN. Notice that the operator $f(\cdot)$ is fixed since it is already known and is based on the model [51]. The loss $\mathcal{L}_{c1}(\cdot)$ is the 2P content loss computed on the labelled dataset. The loss $\mathcal{L}_{c2}(\cdot)$ is the LF content loss, which ensures that the re-synthesized LF computed from the recovered volume is close to the original LF image. The adversarial loss $\mathcal{L}_{adv}(\cdot)$ makes the recovered volume look realistic. The adversarial loss $\mathcal{L}_{adv}(\cdot)$ is computed from a trainable critic $D(\cdot)$, which works as a regularizer. We interpret this loss as a dynamic regularizer that is updated simultaneously with LISTA parameters during training. Note that only $\mathcal{L}_{c1}(\cdot)$ needs a labelled dataset, while $\mathcal{L}_{c2}(\cdot)$ and $\mathcal{L}_{adv}(\cdot)$ need LF images and unpaired 2P data, respectively.

The type of loss shown in Equation (5.7) has been first proposed as a supervised-learning technique for single image super-resolution [92]. Furthermore, learning regularizers to solve various standard inverse problems using stochastic gradient descent has been investigated in [93]. However, we highlight that our approach is a semi-supervised technique compared to [92]. Furthermore, the critic, or regularizer, is not pre-trained, as opposed to [93]. Here, the adversarial regularizer is learned simultaneously with the generator

(LISTA) by using the well-known adversarial training used for least squares GANs (LS-GANs) [94], as depicted in Figure 5.4 (a). In the next section, we explicitly define the training loss and the architecture of the critic used for the experiments.

5.4 Experiments and Results

In this section, we show the performance of our approach by imaging mouse brain tissue with both LF and 2P modalities (see Appendix B.1). We compare the performance of our method with state-of-the-art model-based and learning-based methods for reconstructing structural and functional LF data.

5.4.1 Experimental Setup

The LF microscope is modelled as follows: numerical aperture =1, refractive index =1.33, wavelength =514 nm for jGCaMP8f and 580 nm for TdTomato fluorophore, magnification =25, microlens pitch =125 μm, microlens focal length = 1250 μm, tube lens focal length = 0.18 m, pixels per microlens =19×19.

We compare our method with two model-based approaches: ISRA, a variant of RL algorithm [63] and the ADMM approach proposed in Chapter 4. Furthermore, we evaluate other learning-based approaches by adapting the LFMNet [43], HyLFM [67], and VCDNet [65] to work with our specifications based on the respective code made available online. To train our network and to evaluate model-based approaches, we use the conventional theoretical forward model used for reconstruction proposed by Broxton et al. [51].

We imaged mouse brain slices expressing TdTomato fluorescent protein with a 2P microscope to capture the spatial distribution of the network of neurons. The captured stack contains 80 planes taken at steps of 2 μm. Then, we generate 28 volumes with 53 slices each. The first volume includes planes 1 to 53. The second one contains planes 2 to 54, and so on. Similarly, we capture the corresponding LF stack with 28 images. Therefore, we have a training dataset with 28 training pairs taken from a single brain slice. For evaluation, we capture another dataset of the same size from a different brain sample.

In addition, we capture temporal LF sequences from samples labelled with genetically encoded calcium indicators (jGCaMP8f). We acquire 3 different temporal sequences

with 500 LF images each. We took the first 80 LF images from each one for training. This additional training dataset only contains LF images without any 2P label. In our experiments, the 2P data is acquired only once and is not updated when evaluating different sample tissues.

Due to the dimensionality of the dataset, data augmentation is needed to alleviate data over-fitting. Specifically, we perform data augmentation by using random reflections on the x, y, and z-axis and axes swapping on the x, y dimension of the volume. We modify the LF data accordingly as well. Furthermore, due to size of the data, we use patch-based training to alleviate memory requirements. To this end, we use 3D patches of size $141 \times 141 \times 53$ (height, width, depth) for the 2P-data and the corresponding LF patches of size $47 \times 47 \times 361$ (height, width, number of views).

5.4.2 Deep-learning Setup

To initialize the network, we pre-train it using only the labeled dataset. For this step, we use only the first content loss $\mathcal{L}_{c1}(\cdot)$ in Equation (5.7), which is chosen to be the normalized mean square error loss as follows:

$$\mathcal{L}_{c1}(\mathbf{x}, \hat{\mathbf{x}}) = \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|_2} \right\|_2^2, \tag{5.8}$$

where $\mathbf{x}$ is the 2P 3D image and $\hat{\mathbf{x}}$ is the 3D image reconstructed by the network.

Once the network is initialized, all the losses in Equation (5.7), including the previous $\mathcal{L}_{c1}(\cdot)$ are considered in the minimization. The second content loss named $\mathcal{L}_{c2}(\cdot)$ is given by the following equation:

$$\mathcal{L}_{c2}(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y}_n - \hat{\mathbf{y}}_n\|_2^2, \tag{5.9}$$

where the subscript n represents mean normalization, which is performed as follows:

$$\mathbf{y}_n = \mathbf{y} - \mathbb{E}_{d \times d}[\mathbf{y}], \tag{5.10}$$

where the notation $\mathbb{E}_{d \times d}[\cdot]$ means that the expected value is computed for regions of size $d \times d$ pixels in the spatial dimensions. Specifically, each sub-aperture image of the LF $\mathbf{y}$ is

divided into a grid of squares of size $d \times d$ pixels, then the expected value is subtracted from each square to obtain $\mathbf{y}_n$. This normalization allows focusing on reconstructing details in each sub-aperture image of the LF rather than background noise. In our experiments, the value d is experimentally chosen to be 8. For those familiar with wavelets, this procedure can be interpreted as subtracting the level 3 Haar approximation from each view in $\mathbf{y}$. Thus, the loss only considers the horizontal, vertical, and diagonal details of levels 1,2, and 3.

The adversarial loss in Equation (5.7), $\mathcal{L}_{adv}(\cdot)$ is determined by the discriminator $D(\cdot)$. The discriminator tries to assign different scores to the real 3D data and to the reconstruction from the network $g(\cdot)$. We name $\mathbb{P}_\theta$ the probabilistic distribution of real 3D volumes and $\mathbb{P}_r$ the distribution of 3D reconstructions from the network. Then, the adversarial loss is given by:

$$\mathcal{L}_{adv}(\mathbf{x}) = \mathbb{E}_{\mathbf{x}\sim\mathbb{P}_\theta}[(D(\mathbf{x}) - a)^2], \tag{5.11}$$

where $a = 1$ and the architecture of $D(\cdot)$ is depicted in Figure 5.4 (b). The discriminator was designed by following standard architectures used for 3D GANs [91]. The expected value $\mathbb{E}[\cdot]$ is approximated by computing the mean on a batch of volumes generated by LISTA. Finally, the discriminator is trained by using the loss

$$\mathbb{E}_{\mathbf{x}\sim\mathbb{P}_\theta}[(D(\mathbf{x}) - b)^2] + \mathbb{E}_{\mathbf{x}\sim\mathbb{P}_r}[(D(\mathbf{x}) - a)^2], \tag{5.12}$$

where $a = 1$, $b = -1$. As mentioned previously, the first expected value is computed on the batch of volumes generated with LISTA. Similarly, the second expected value is computed on batches of real data. Intuitively, the discriminator is trained to assign a 1 if the input has the 2P quality and a -1 if the input is generated by $g(\cdot)$ and does not have 2P quality. At the same time, $g(\cdot)$ tries to make $D(\cdot)$ to assign a 1 by improving the reconstruction quality. This training procedure is part of a standard technique proposed in [94] to train LSGANs.

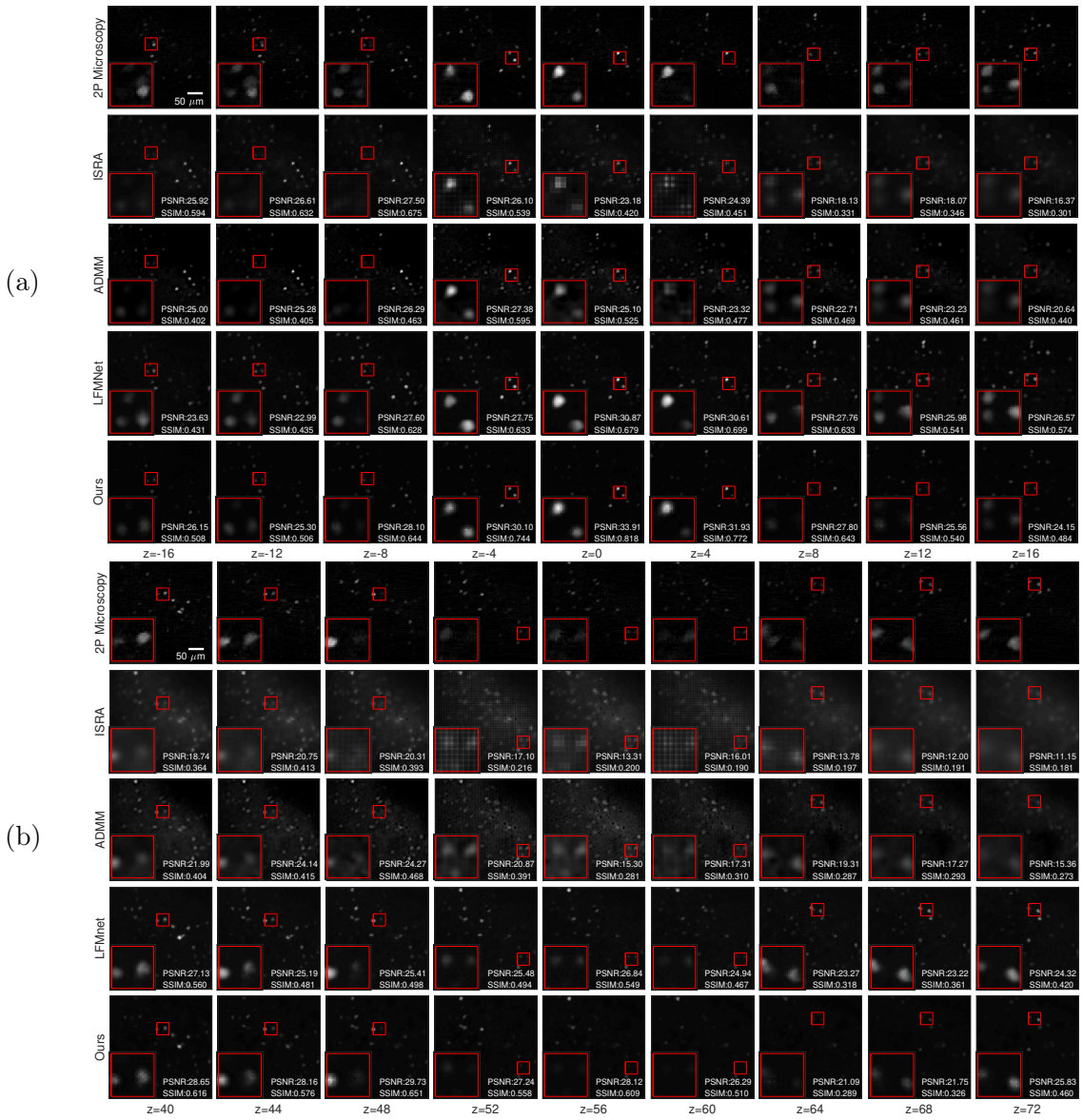


Figure 5.5: Reconstruction using real LF data from acute mouse brain slices expressing TdTomato fluorophore. In part (a), the first three rows show the 2P 3D image used as ground truth, and the reconstruction using two model-based approaches: ISRA and ADMM, respectively. Furthermore, in the next two rows we evaluate the state-of-the-art LFMNet proposed in [43] and we show our approach. We show several slices for different depths. This reconstruction corresponds to the performance shown in the first row in Table 5.1. In part (b), we show performance for a LF image with a deeper focal depth, corresponding to the row 28 in Table 5.1. The performance of all methods degrades when imaging deeper in the tissue. Note that our method achieves the best performance in terms of both PSNR and SSIM. The shown PSNR and SSIM are measured for the whole plane at each depth. Measures on the whole volume are shown in Table 5.1. All the distances are measured in μm. The settings used to capture both the LF image and 2P image are specified in Section 5.4.

5.4.3 Reconstruction of Structural 3D Images from Ligh Field Images

In this section, we evaluate the performance of our method for reconstructing 3D images from a single LF image. We use unseen samples to measure the reconstruction performance by measuring the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). The neurons in this sample are labelled with the TdTomato fluorescent fluorophore. Since this fluorophore is bright every time it is illuminated, these LF images only give structural information and do not show neuronal activity. As mentioned previously, we captured 28 LF images for training and 28 from a different sample for testing. Each LF image corresponds to a focal depth ranging from 0 to $54\mu m$. From every LF image, we reconstruct a volume of size $321 \times 321 \times 53$ voxels covering a range $533.3 \times 533.3 \times 104\ \mu m^3$. The size of each LF image is 2033×2033 pixels.

The number of iterations used for model-based reconstruction must be chosen properly to avoid noise amplification. As mentioned in previous works [17], [64], [18], a typical empirical number of iterations used for ISRA is between 8 and 10. We fixed this value to 8 for both ISRA and ADMM. Our LISTA network comprises 6 unfolded iterations. Remarkably, 6 unfolded iterations of LISTA are enough to outperform competing methods.

The LISTA network achieves better average performance than competing methods in a focal depth range of 54 μm in terms of PSNR and SSIM. In Table 5.1 we show the performance for LF images taken at different focal depths; they are taken at steps of $2\mu m$. The depth index in Table 5.1 increases as the depth increases. Note that all methods are affected by scattering as the depth increases. Even though all learning methods outperform model-based reconstruction approaches, our method achieves the best performance in both PSNR and SSIM. The shown PSNR and SSIM are measured on the whole volume. In our experiments, the LFMNet achieved better performance among competing methods. Since the learning-based methods are trained with a very small dataset compared to the size of the dataset used in [43], [67], [65], their performance may be affected. In contrast, our method is more robust under this adverse condition. Furthermore, deep-learning methods are much faster than model-based approaches. Table 5.2 shows the average computation time to reconstruct a volume from a single LF image. All the methods were evaluated on a GeForce GTX 1080 Ti.

	PSNR						SSIM					
Depth Index	ISRA	ADMM	VCDNet	HyLFM	LFMNet	Ours	ISRA	ADMM	VCDNet	HyLFM	LFMNet	Ours
1	28.78	29.55	30.84	31.48	31.02	**34.19**	0.61	0.62	0.65	0.67	0.64	**0.82**
2	28.84	30.16	32.09	32.04	30.98	**34.34**	0.60	0.68	0.72	0.70	0.64	**0.82**
3	27.32	28.32	31.09	31.89	30.53	**33.88**	0.55	0.61	0.66	0.68	0.62	**0.79**
4	26.22	28.54	29.99	32.89	29.98	**33.20**	0.50	0.63	0.62	0.74	0.60	**0.75**
5	24.95	27.54	29.82	32.67	29.64	**33.18**	0.46	0.59	0.61	0.73	0.60	**0.75**
6	25.37	26.67	28.79	**33.00**	29.95	32.99	0.48	0.57	0.57	**0.75**	0.60	0.74
7	26.36	27.56	28.94	31.68	31.18	**33.33**	0.51	0.58	0.57	0.68	0.65	**0.76**
8	26.58	27.90	29.35	29.97	31.57	**33.62**	0.53	0.58	0.59	0.61	0.67	**0.78**
9	26.08	28.07	29.69	30.09	32.41	**33.84**	0.51	0.60	0.61	0.62	0.71	**0.80**
10	25.14	27.47	29.77	30.22	31.16	**33.38**	0.48	0.56	0.61	0.62	0.65	**0.77**
11	25.49	27.70	31.12	30.27	30.94	**33.29**	0.49	0.60	0.68	0.62	0.65	**0.77**
12	24.95	26.77	30.24	29.38	29.03	**32.65**	0.47	0.55	0.63	0.58	0.56	**0.73**
13	25.06	26.43	30.03	29.79	31.36	**32.25**	0.48	0.55	0.63	0.60	0.67	**0.72**
14	25.08	27.65	29.80	29.96	30.17	**31.77**	0.47	0.59	0.61	0.60	0.61	**0.68**
15	25.41	27.68	29.51	29.26	29.11	**31.48**	0.48	0.60	0.60	0.57	0.57	**0.67**
16	24.88	27.48	29.19	28.13	28.39	**31.10**	0.47	0.58	0.58	0.52	0.53	**0.65**
17	24.50	27.11	28.20	27.12	28.54	**30.69**	0.45	0.56	0.53	0.48	0.54	**0.63**
18	23.91	26.71	29.76	26.94	28.75	**31.54**	0.43	0.55	0.60	0.48	0.55	**0.67**
19	22.63	25.42	25.93	26.09	27.28	**28.89**	0.39	0.50	0.44	0.44	0.49	**0.56**
20	22.28	25.03	27.54	27.45	27.85	**31.53**	0.38	0.49	0.50	0.49	0.51	**0.66**
21	20.80	23.57	25.86	27.00	27.31	**30.08**	0.33	0.45	0.43	0.47	0.49	**0.60**
22	21.19	23.06	26.74	26.86	27.89	**31.42**	0.34	0.45	0.47	0.47	0.51	**0.66**
23	21.60	23.37	25.80	26.15	28.32	**30.00**	0.35	0.44	0.43	0.44	0.52	**0.60**
24	21.43	23.51	25.36	26.42	27.74	**30.31**	0.35	0.44	0.42	0.45	0.50	**0.61**
25	20.21	23.32	26.28	26.02	27.92	**30.54**	0.31	0.47	0.45	0.43	0.51	**0.62**
26	20.46	23.19	27.35	25.47	27.88	**28.99**	0.32	0.46	0.49	0.41	0.50	**0.56**
27	19.80	22.62	27.80	24.58	26.70	**30.66**	0.30	0.43	0.50	0.38	0.45	**0.62**
28	19.30	22.09	26.26	25.05	27.16	**31.06**	0.28	0.42	0.44	0.39	0.47	**0.63**
Mean	24.09	26.25	28.68	28.85	29.31	**31.94**	0.44	0.54	0.56	0.56	0.57	**0.69**

Table 5.1: PSNR and SSIM for real LF data of neurons imaged using tdTomato fluorophore.

Table 5.2: Computation time.

	ISRA	ADMM	VCDNet	HyLFM	LFMNet	Ours
Time (s)	234.69	237.63	0.022	0.195	0.111	0.026

As shown in Figure 5.5, the LISTA network achieves better qualitative reconstruction performance than other methods. In Figure 5.5 (a) and (b), we show visual results for two different depths corresponding to index 1 and 28 in Table 5.1, respectively. ISRA introduces square-like artifacts strongly present near the in-focus plane, approximately from $z = -8\ \mu m$ to $z = 8\ \mu m$ in part (a) and from $z = 48\ \mu m$ to $z = 64\ \mu m$ in (b). The ADMM can effectively remove these artifacts; however, both ISRA and ADMM are affected by background noise and scattering. As one goes deeper into the tissue, the model-based methods are more affected by scattering. It is notable that learning methods achieve better performance than model-based approaches and are remarkably less affected by noise. However, our approach is visually closer to the ground truth and achieves higher PSNR

and SSIM than other learning methods. For instance, see plane $z = 48\ \mu m$. Also, note in $z = 64\ \mu m$ that the LFMNet incorrectly reconstructs neurons from neighbour depths.

5.4.4 Reconstruction of Volume Time Series from LF Images

In this section, we evaluate the performance of our method for reconstructing a temporal sequence of 3D volumes from temporal sequences of LF images. The LF sequence captures the activity of neurons labelled with the jGCaMP8f calcium indicator at different times focused at a fixed focal depth. The jGCaMP8f is a fluorophore that increases its fluorescence intensity when the neurons fire. In this case, the ground truth data is unavailable since it is impossible to capture the activity of many neurons in 3-D with scanning-based techniques. We evaluate our approach on 3 LF sequences with 500 frames. As mentioned previously, only the first 80 LF images per sequence were used for training (with no labels), while the rest of LF images were unseen by the network.

Our LISTA network performs better than model-based approaches, while the state-of-the-art neural networks fail to reconstruct volumes for jGCaMP8f-labelled brain tissues. In Figure 5.6, we show the visual performance of ISRA, ADMM, and our method for reconstruction of one frame of the sequence ($300th$ frame). We show two different samples in part (a) and (b). Even though the LFMNet [43] achieved satisfactory performance in the previous section, it generalizes poorly to the reconstruction of the temporal sequence due to the small training dataset. A more specific reason is that the use of a different fluorophore, the jGCaMP8f, implies samples with different noise levels and light sources with different wavelengths than those used for training. Figure 5.6 suggests that model-based methods are more robust under these adverse conditions than learning-based approaches, as mentioned in the introduction. However, model-based methods are heavily affected by scattering. In addition, ISRA introduces strong artifacts near the plane $z = 0$. In our approach, we exploit the knowledge of the forward model and the few available labels to achieve remarkable reconstruction performance.

Our training loss is designed to avoid amplifying noise from scattering. Contrary to conventional model-based approaches, our objective function avoids amplifying noise by regularizing the reconstruction using real 2P data with a discriminator, as shown in Equation (5.7). In Figure 5.7, we show the LF images synthesized from the reconstructed

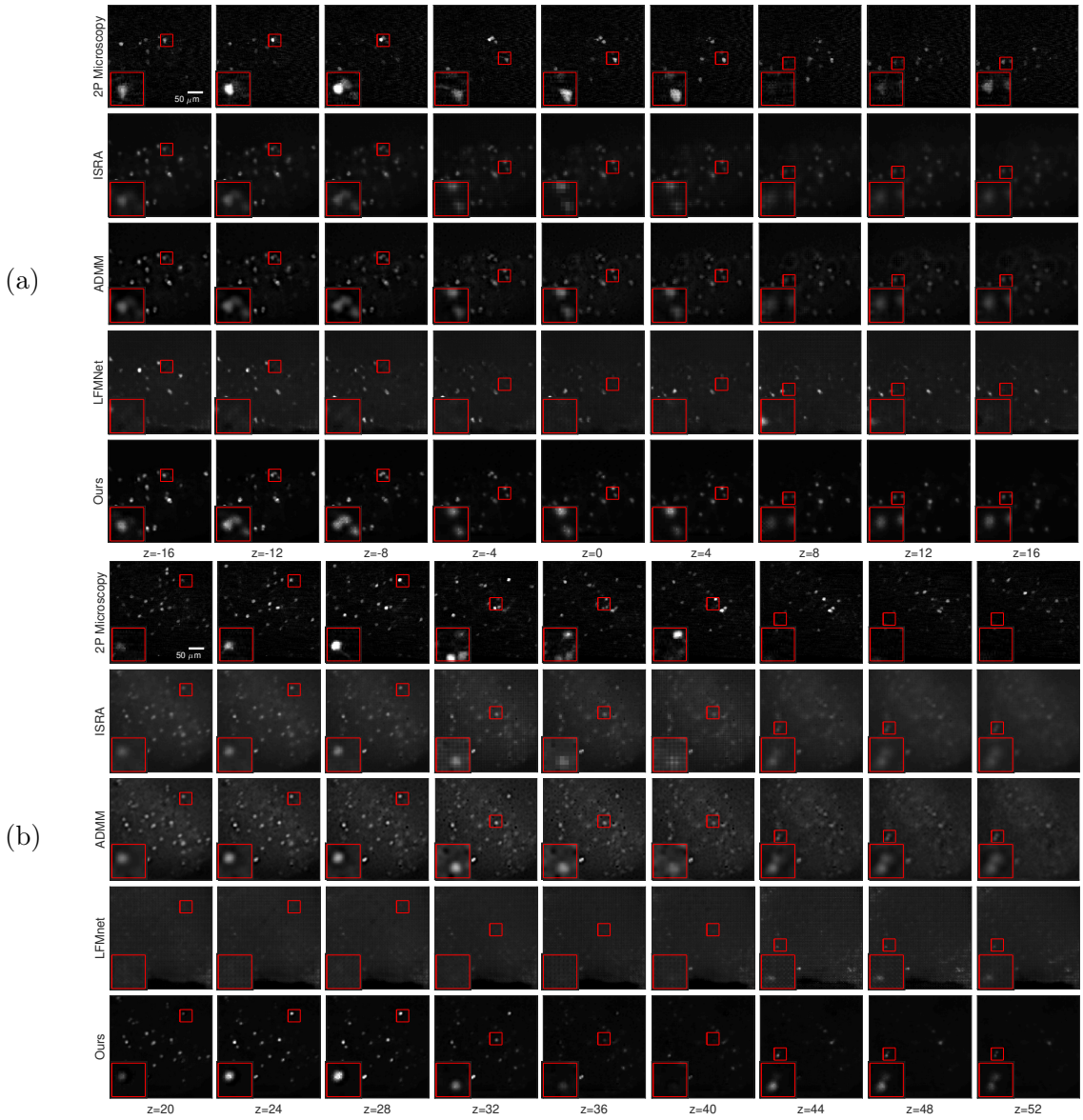


Figure 5.6: Reconstruction using real LF data from acute mouse brain slices expressing the calcium indicator jGCaMP8f. The indicator jGCaMP8f is suitable to perform functional imaging of neurons. The labels (a) and (b) indicate two different samples. The first row shows the 2P 3D image of the static tdTomato fluorophore used as a reference since the ground truth (2P jGCaMP8f volume) is unavailable. We show a particular reconstructed frame from a 500-frames sequence, see Section 5.4.4 for details. We included the performance of the LFMNet [43] and two model-based approaches for comparison: ISRA and ADMM. We show different slices corresponding to different depths. All the distances are measured in μm. The settings used to capture both the LF image and the 2P 3D image are specified in Section 5.4.

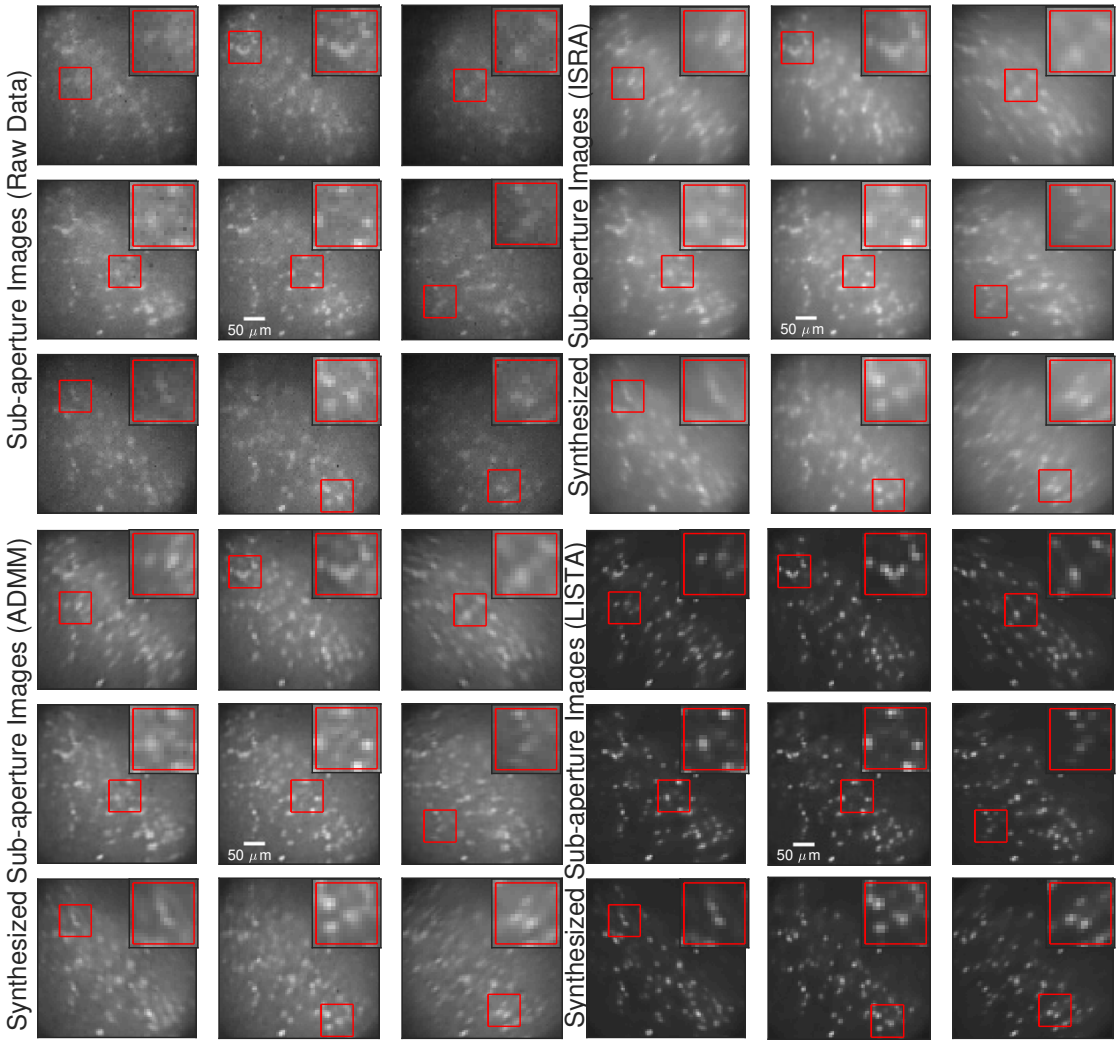


Figure 5.7: Visual comparison of sub-aperture images. We show 9 different sub-aperture images (views) for 4 different cases: views from the raw LF data, LF synthesised from ISRA reconstruction, LF synthesised from ADMM method, and LF synthesised from our LISTA approach. Both model-based methods reconstruct noisy regions in the sub-aperture images that do not carry any meaningful information, which translates into a noisy 3D reconstruction. On the other hand, our approach can accurately reconstruct every neuron's footprint while significantly reducing noise from scattering.

volumes. We display 3×3 views per LF image from the total 19×19 views. The re-synthesized LF image from ADMM shows that noise is reduced compared to the ISRA approach since the ADMM method imposes additional regularizers in the objective function. However, noise in the tissue region is still significant. Our approach greatly reduces noise while reconstructing every neuron footprint in the ground truth LF image. We achieve this remarkable performance due to the adversarial regularizer and the specialized

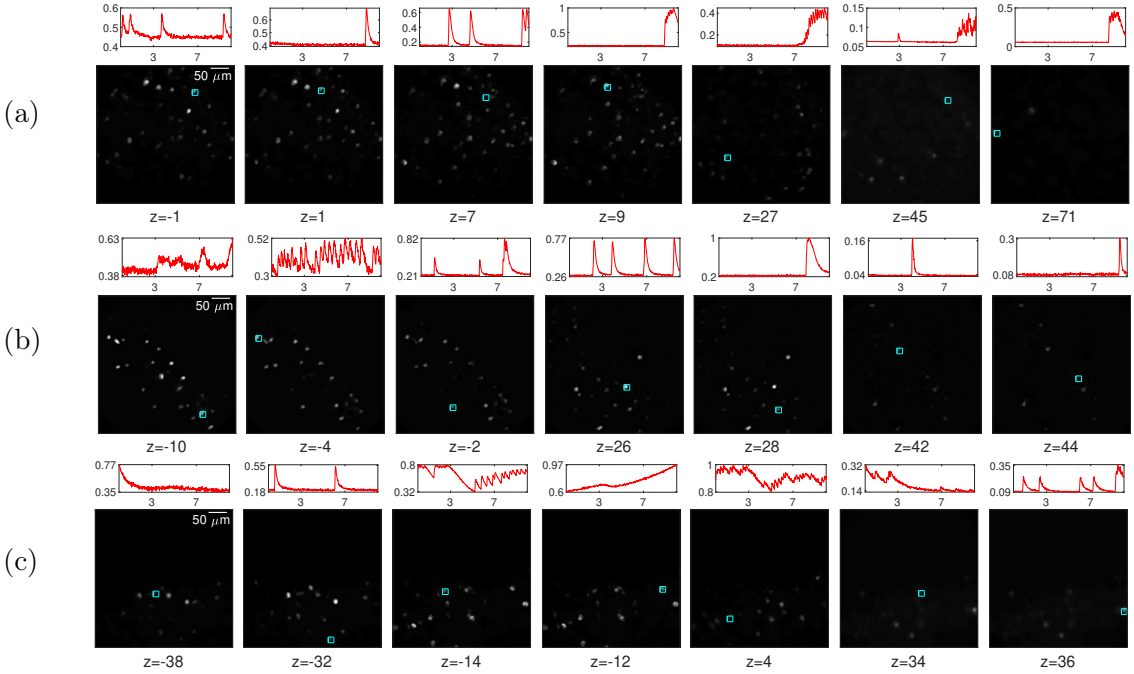


Figure 5.8: Reconstruction of the temporal evolution of neuron activity in brain tissue. Our LISTA network allows fast reconstruction of volume time series from LF data of mammalian brain tissue expressing the calcium indicator jGCaMP8f. The labels (a), (b), and (c) indicate three different sample tissues. We show several slices containing active neurons, the neuron of interest is marked in cyan. On top of each slice, we show the temporal evolution of the neuron. The horizontal axis shows time in seconds, the vertical axis shows the normalized intensity. The normalization is performed per sample between the selected neurons. Our method reconstructs 500 frames (500 3D images) in this experiment, a computationally demanding task with conventional model-based methods. The LF imaging rate is 50 Hz. All the distances are measured in μm. Other additional settings are specified in Section 5.4.

content losses used for training.

Our approach provides a new powerful tool to study fast temporal evolution of neurons in mammalian brain tissue. In Figure 5.8, we show reconstruction of temporal evolution of neurons. We show three brain samples in parts (a), (b), and (c). For this experiment, we perform reconstruction from LF sequences of 500 frames. Therefore, we reconstruct 500 3D volumes per LF sequence with the same size as in previous experiments. Since the imaging rate is 50 Hz, we show 10 seconds of neuronal activity. Note how neurons located at different positions in the 3D space show different activity.

5.5 Summary

We have introduced a physics-driven deep neural network to reconstruct 3D volumes from LF sequences. The architecture of the network is based on the observation that labelled neurons in tissues are sparse, which naturally leads to an architecture based on unfolding the ISTA algorithm. Moreover, we show how the forward model of a LF microscope can be described using CNNs, and this physics-driven architecture is also included in the network. Finally, we use GANs and exploit the theoretical forward model to train our network with a small labelled dataset.

We show that our method achieves better reconstruction quality for reconstructing temporal LF sequences imaged with the jGCaMP8f indicator than other state-of-the-art methods. None of the competing learning-based methods can perform this task. Furthermore, we also showed better performance than standard model-based and learning-based methods in terms of PSNR and SSIM for reconstructing structural 2P volume imaged using the TdTomato fluorophore. Although LFM cannot penetrate as deeply into scattering tissue as functional 2P imaging, our 2P-enhanced LFM strategy enables light-efficient volume acquisition at fast rates, essential to capturing neuronal dynamics transduced by fast sensors such as jGCaMP8f [74].

This chapter described a practical method to perform 3D reconstruction for LF microscopy under adverse acquisition conditions when imaging mammalian brain tissue. We believe that the proposed method could be helpful even beyond this specific scenario, and it could inspire similar solutions for other types of inverse problems.

Chapter 6

Conclusions

6.1 Summary of Thesis Achievements

In this thesis, we considered the problem of reconstructing 3D images from single LF images. We first studied methods for discretization based on generalized sampling theory and proposed simplifying the forward model to accelerate computation. Then, we examined aliasing-free reconstruction under ideal conditions. For real acquisition conditions, we proposed a reconstruction method based on ADMM that achieves better quality than conventional approaches. Furthermore, simplifying the forward model makes our approach faster than conventional RL-based reconstruction. Finally, we proposed a deep-learning method for reconstruction. Our method uses a multi-modal imaging approach based on LFM and 2P microscopy, which is tested under adverse conditions when imaging mammalian brain tissue. The network architecture is derived from the sparsity-based reconstruction algorithm named ISTA. In addition, we proposed a semi-supervised approach that uses a few labelled data and exploits the knowledge of the forward model and GANs for training.

In Chapter 3, we proposed a novel way to discretize the LF system that allows for more diverse sampling kernels and densities. Traditional reconstruction approaches use a discretization method based on the computation of the system impulse response with restricted sampling intervals, which is a special case of our framework. Our approach allows diverse sampling kernels such as Dirac deltas, sinc functions, or splines, and the lateral sampling intervals do not need to be a multiple of the microlens pitch. Furthermore, since periodic-shift invariance still holds after discretization, the system can be modelled using

filter banks. Specifically, the system is described using a custom filter bank per depth. The filter-bank description also leads to a simplification method to accelerate the computation. We found that the truncated SVD allows finding a set of optimal synthesis and analysis filters that minimize the approximation error and can progressively approximate the system. Due to the low-rank nature of the LF system, we can accelerate the computation of the forward model without significantly impairing the accuracy. Finally, we propose a novel method to compute the forward model by reshaping the filter-bank description into a linear CNN. The CNN allows taking advantage of existing deep-learning software for faster computation or calibration by exploiting deep-learning optimization algorithms.

In Chapter 4, we explored model-based reconstruction for LFM. In an ideal scenario, perfect reconstruction (artifact-free) can be achieved by exploiting the shift-invariant subspace assumption and the pseudoinverse. However, perfect reconstruction depends on the proper selection of the template functions and shifts defining the shift-invariant subspace. Traditional approaches do not impose enough regularization to prevent the solution from lying in the row space of the system matrix. The row space intrinsically generates square-like artifacts in the native object plane. The shift-invariant subspace assumption ensures that the reconstruction lies in a different subspace which implies artifacts-free reconstruction. However, this idea works under idealized settings that do not hold in practice due to system imperfections and noise from scattering, which is common when imaging mammalian brain tissue. Therefore, we propose a new reconstruction algorithm that works under real acquisition conditions. Our method allows artifacts-free reconstruction, higher speed, and higher PSNR and SSIM than conventional methods. The improvement in quality is achieved by incorporating additional priors to the optimization approach and designing a specific algorithm based on ADMM. The improvement in computation time is due to the exploitation of the underlying low-rank property of the measurement matrix by either using our SVD approximation or assuming that the reconstructed volume lies in a SIS. Numerical results on synthetic and real data demonstrate that our approach consistently outperforms conventional volume reconstruction approaches based on the RL algorithm.

In Chapter 5, we presented efficient CNN architectures to describe the LF system.

Exploiting the periodically-shift invariance allows different CNN architectures to reproduce the forward model or perform accurate approximations. Furthermore, they allow solving the inverse problem using a deep unfolding network or enable efficient online computation of the forward model when training the reconstruction network.

The architecture of the reconstruction network is designed from the unfolding of the ISTA algorithm. Specifically, we exploit the assumption that 3D images of labelled neurons in tissue are sparse and non-negative. We propose a semi-supervised learning approach to avoid the need for huge datasets for training. Specifically, the proposed imaging setup operating in two-photon modality allows capturing a small labelled dataset of 3D images, which is complemented with the knowledge of the forward model to train the network. The training strategy is based on the adversarial training proposed for LSGANs.

We experimentally showed that our method performs better than standard model-based and deep-learning methods. Our method achieves better average performance in terms of PSNR and SSIM for neurons tagged with the TdTomato fluorophore, which shows structural information. Furthermore, the proposed CNN successfully reconstructs 3D volume time series of neuronal activity tagged with the GCAMP8 indicator. We found that other deep-learning methods perform well for data similar to the one used in training (emitting TdTomato fluorescence). However, they collapse when faced with LF data from neurons with the GCAMP8 indicator. Lack of generalization is typical for deep-learning approaches when trained with small datasets, while model-based approaches for LFM are more robust under these challenging conditions. Our method offers more robustness and better performance than typical reconstruction methods for fast 3D reconstruction from LF images under adverse acquisition conditions when imaging mammalian brain tissue.

6.2 Future Work

6.2.1 Forward Model Calibration or Learning

In this thesis, we showed that the theoretical model proposed by Broxton et al. [51] allows 3D reconstruction with both model-based and learning-based techniques even under adverse conditions. However, this model assumes ideal optical devices to simplify computation of equations which may lead to inaccuracies in the reconstruction. Therefore, one

interesting research direction is to use the proposed CNN architectures to calibrate the forward model for a given microscope. This calibration requires labelled data for learning but not necessarily data from neurons since the objective is to capture the impulse response of the system. For instance, one can image fluorescent beads to create a small labelled dataset for calibration. Furthermore, using a mixture of data created from the theoretical model and real images may alleviate the need for huge amounts of real labelled data for training. Another possible approach could be to simultaneously update the weights of the forward model and the reconstruction network. However, this approach could lead to instabilities in training, and the learned reconstruction may not be reliable due to uncertainty in the learned forward model. In bio-imaging applications, reconstruction accuracy is critical. In general, more precise modelling of the system may allow going beyond the results achieved in this thesis.

6.2.2 Exploitation of Temporal Correlation in LF Stacks

Our framework is proposed for the reconstruction of any LF stack. However, as mentioned in this work, the main goal of LFM is to study the temporal evolution of biological samples. Therefore, an interesting research direction is to exploit the temporal correlation between LF frames to achieve better performance in the reconstruction of temporal sequences. There are model-based reconstruction methods exploiting this correlation. However, they are highly demanding in terms of computation [17]. A possible direction is to adjust either the network architecture or the training loss to ensure temporal consistency. Regardless of the strategy, new reconstruction frameworks should consider computational complexity since LFM is intended to perform the reconstruction of large LF stacks. Even though reconstructing a high-quality 3D image from a single LF image is interesting, it is less relevant than the reconstruction of volume time series since other optical techniques, such as two-photon microscopy, can already perform accurate 3D imaging.

6.2.3 Design of Discriminators

A specialized discriminator architecture for neuronal imaging may help to improve reconstruction results. In our work, the discriminator is designed based on architectures proposed in previous works on 3D GANs. We show that using a simple architecture for the

discriminator allows enough regularization to perform satisfactory reconstruction results without increasing computation complexity. However, other types of discriminator may also give good regularization performance or improved stability in training. For instance, a Cycle GAN used for deconvolution microscopy proposed in [95] is regularized using a multi-patch-based discriminator that considers different scale levels for regularization. An interesting research direction could be studying new architectures for the discriminator based on model-based algorithms or studying the performance of other known discriminator architectures for bio-imaging.

6.2.4 Exploration of Alternative LFM designs

The techniques proposed in this thesis are conceived for the original LF microscope proposed in [45], known as Plenoptic 1.0. However, the ideas and methods presented here, such as discretization, SVD-based simplifications, or CNN-based models, could potentially be extended to other alternative optical configurations, such as Plenoptic 2.0 [96] and Fourier LFM [97]. Furthermore, investigating more efficient optical methods for fast 3D imaging is an interesting problem. Current hand-crafted optical designs such as plenoptic 1.0, plenoptic 2.0, or FLFM could be improved by incorporating data-driven approaches in the design of the optical system. For instance, one could use deep-learning frameworks to find an optimal optical architecture that allows 3D reconstruction with fewer artifacts under real acquisition conditions.

Appendix A

A.1 Fresnel-Kirchhoff Diffraction

A typical method used to find a solution to the Helmholtz Equation (2.6) is to make use of the modified version of Green's theorem from calculus [98]:

$$\iiint_V \left(U\nabla^2 G - G\nabla^2 U\right) dv = \iint_S \left(U\frac{\partial G}{\partial n} - G\frac{\partial U}{\partial n}\right) ds, \tag{A.1}$$

where U and G are two complex-valued scalar fields, S is a closed surface surrounding a volume V and $\frac{\partial}{\partial n}$ is the partial derivative in the outward normal direction The fields U and G and their partial derivatives must be continuous inside S and on S.

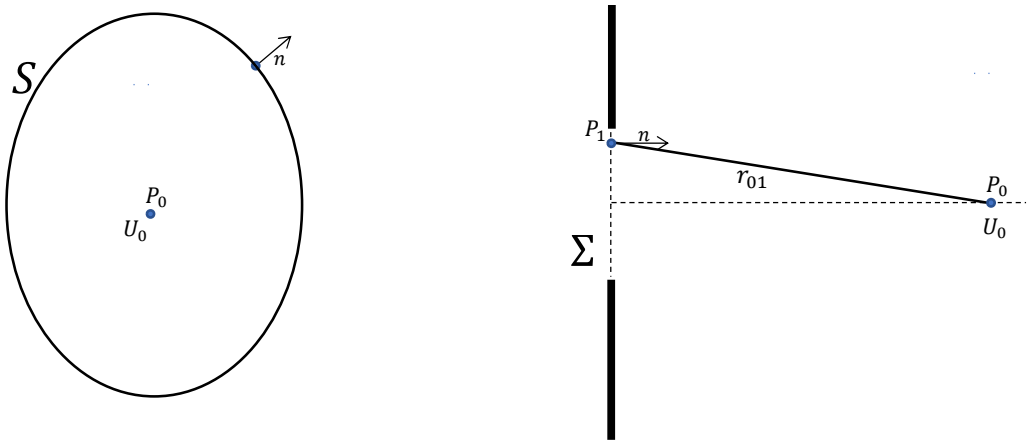
The integral theorem of Helmholtz and Kirchoff make use of Green's Theorem in Equation (A.1) to compute the field U_0 at a point P_0 in the space given its value on a close surface S enclosing P_0. See Figure A.1 (a). The field U_0 can be computed as follows:

$$U_0 = \frac{1}{4\pi}\iint_S \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n}\right) ds, \tag{A.2}$$

where the same notation used in Equation (A.1) is still valid. The function G is usually a Green function of the Helmholtz Equation (2.6). However, other strategies to choose G can be used according to the scenario [98], as mentioned in the following paragraphs. A well-known Green function of Equation (A.1) is the spherical wave satisfying the Sommerfeld's radiation conditions [98], [99]:

$$G(P_1) = \frac{exp(jkr_{01})}{r_{01}}, \tag{A.3}$$

where r_{01} is the distance between point P_0 and point P_1, P_0 is the center of the spherical



(a) (b)

Figure A.1: Scalar wave optics. The integral theorem of Helmotz and Kirchoff allows to compute the field U_0 at any point P_0 in the space, given the field over a closed surface S,as shown in (a). A particular case of interest is the diffraction of a wave over an aperture Σ. In (b), under some approximations, the knowledge of the field over the aperture is sufficient to recover the field U_0 at any point P_0 in the space with accuracy. The field U_0 can be computed using the the Fresnel-Kirchoff diffraction formula or the Raylegh-Sommerfeld solution.

wave and P_1 is an arbitrary point.

An important case of study is the diffraction by a screen. See Figure A.1 (b). The problem is to find the field U_0 at a point P_0 given that we know the field at the aperture Σ. There are two well-known strategies using the integral theorem of Helmholtz and Kirchoff to solve this problem.

A.1.1 Fesnel-Kirchoff Diffraction Formula

The Fesnel-Kirchoff diffraction formula computes the field as follows:

$$U_0 = \frac{1}{4\pi} \iint_{\Sigma} \left(G \frac{\partial U}{\partial n} - U \frac{\partial G}{\partial n} \right) ds, \tag{A.4}$$

where Σ is the aperture. As explained in [98], the Fesnel-Kirchoff diffraction formula uses assumptions that are contradictory themselves. However, this formula is sufficiently accurate to compute far fields.

A.1.2 Rayleigh-Sommerfeld Solutions

The Rayleigh-Sommerfeld solutions compute the same field in Equation (A.4) but solve contradictions in the Fresnel-Kirchoff assumptions. The first Rayleigh-Sommerfeld solution

is described by the following equation:

$$U_0 = -\frac{1}{2\pi}\iint_{\Sigma} U\frac{\partial G}{\partial n}ds, \tag{A.5}$$

similarly the second Rayleigh-Sommerfeld solution is described as follows:

$$U_0 = \frac{1}{2\pi}\iint_{\Sigma} G\frac{\partial U}{\partial n}ds. \tag{A.6}$$

Both Rayleigh-Sommerfeld solutions can be derived using the integral theorem of Helmholtz and Kirchoff by using a field G composed of a summation of an incoming and outgoing spherical wave rather than a single spherical wave (Equation (A.3)). The first Rayleigh-Sommerfeld solution assumes a field G that vanishes on the diffraction aperture, while the second solution assumes a field G such that the normal derivative vanishes on the aperture. Both solutions compute the same field, but one may be more convenient to use than the other, depending on the scenario. More details are given in [98]. Furthermore, the Rayleigh-Sommerfeld are more rigorous solutions of the Helmholtz Equation than the Fresnel-Kirchoff diffraction formula.

A.2 Angular Spectrum Representation

A simple but powerful idea to simplify the analysis of the field propagation is to study the Fourier transform of the scalar field rather than the field directly. The angular spectrum of a scalar field is:

$$\hat{U}(k_x, k_y, z) = \iint_{-\infty}^{+\infty} U(x, y, z)e^{-j(k_x x + k_y y)}dxdy, \tag{A.7}$$

which is simply the 2D Fourier transform of the field with fixed z. It is possible to show that the angular spectrum of a scalar field satisfying the Helmholtz Equation (2.6) also satisfies a simple relationship [48]:

$$\hat{U}(k_x, k_y, z) = \hat{U}(k_x, k_y, 0)\hat{G}\left(k_x, k_y; z\right), \tag{A.8}$$

where $\hat{G}\left(k_x, k_y; z\right) = e^{jk_z z}$, which shows that the propagation can be described as a linear shift-invariant system with a transfer function $\hat{G}(k_x, k_y; z)$. We clarify that k_z is a function of k_x and k_y as follows:

$$k_z = \sqrt{k^2 - k_x^2 - k_y^2} \approx k - \frac{(k_x^2 + k_y^2)}{2k}, \tag{A.9}$$

where $k = 2\pi/\lambda$ is the angular wavenumber, and λ is the wavelength. The last approximation is called paraxial approximation, which is useful to find a close form of the impulse response $G(x, y)$ of the propagation process as follow:

$$G(x, y; z) \approx \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)}. \tag{A.10}$$

The Fourier pair of the shown impulse response is:

$$\hat{G}\left(k_x, k_y; z\right) \approx e^{jkz} e^{-\frac{jz}{2k}(k_x^2+k_y^2)}. \tag{A.11}$$

Therefore, the propagation of a scalar field along the z-axis can be computed from a convolution with the propagation kernel $G(x, y; z)$ or a multiplication with the transfer function $\hat{G}\left(k_x, k_y; z\right)$ in the Fourier domain. We note that the constant factors in Equations (A.10) and (A.11) can be safely ignored in many applications.

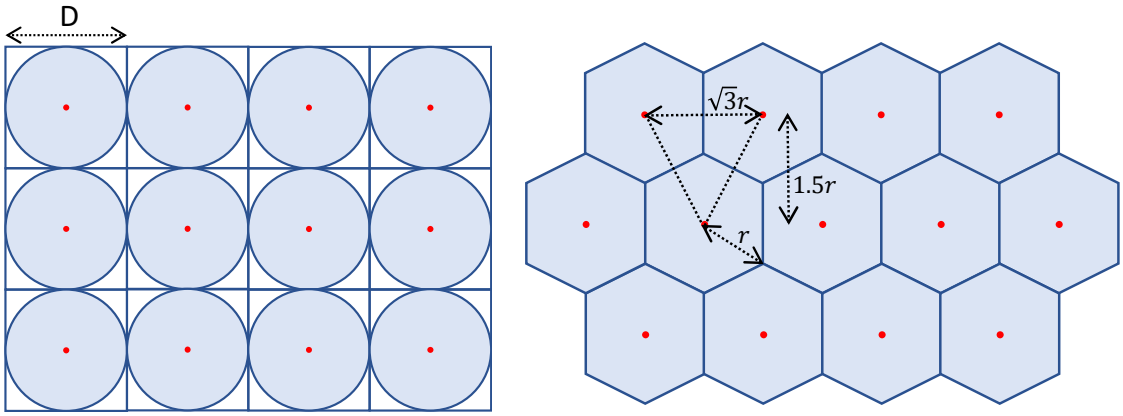
A.3 Lenses

A.3.1 Simple Lens

The effect of a lens over the light field is usually modelled by using the concept of transmittance. The transmittance is a function that models the relationship between the field just after the lens and the incident field. As shown in [50], the transmittance of a lens can be expressed as:

$$t(x, y) = P(x, y) e^{\frac{jk(x^2+y^2)}{2f}}, \tag{A.12}$$

where the function $P(x, y)$ is the pupil function, and f is the focal length of the lens. The pupil function defines the physical size and the shape of the lens. In simple models, the pupil function is always real and non-negative. However, it can be a complex function to



(a) (b)

Figure A.2: Microlens arrays. In part (a), a circular-shaped and square-packed microlens array is shown. This configuration is defined by the microlens pitch D. In (b), a hexagonal shaped and hexagonal packed microlens array is shown. This microlens type can be characterized by the maximal radius r.

model aberrations [50].

A particular type of lens is the microlens array. A microlens array is a set of typically thousands of tiny lenses packaged together. Two diagram of different microlens arrays are shown in Figure A.2. Note that they can be arranged in different ways, mainly rectangular, squared and hexagonal [100]. The whole array can be seen as a lens. Therefore, it is possible to compute its transmittance $\Phi(x, y)$ as follows:

$$\Phi(x, y) = t(x, y) * \Delta(x, y), \tag{A.13}$$

where the operator $*$ represents a 2D convolution, $t(x, y)$ is the transmittance of a single microlens, modelled as in Equation (A.12) and $\Delta(x, y)$ is a 2D function that indicates the location of the microlens centers. Specifically, $\Delta(x, y)$ is a set of Dirac functions centered at the center of each microlens. For instance, it could be a 2D comb function, as in [51]. Note that the shape of each microlens is modelled with the transmittance $t(x, y)$, while the packing of the array is modelled with the function $\Delta(x, y)$. For instance, one can have a squared-packed array of circular-shaped lens, as shown in Figure A.2.

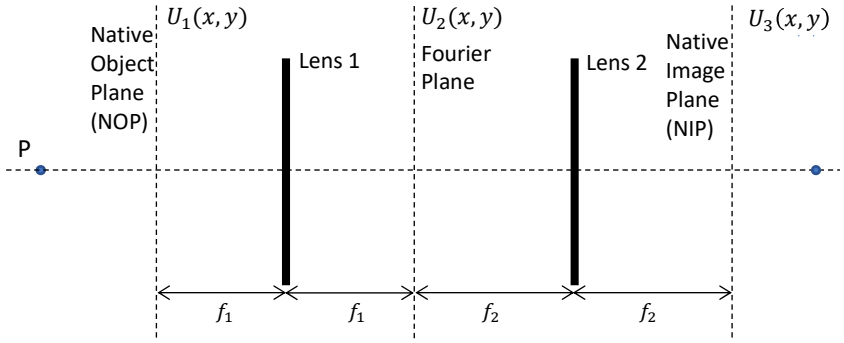


Figure A.3: 4F system diagram. A 4f system replicates the input field U_1 located at the native object plane into the Native image plane $U_3(x_3, y_3) = U_1(-x_1M, -y_1M)$, where M is the magnification factor $M = f_1/f_2$, f_1 and f_2 are the focal length of lens 1 and lens 2 , respectively. At the intermediate plane (Fourier Plane), the field U_2 is the Fourier transform of the input U_1 (up to a constant factor).

A.3.2 4f System

A 4f-system is an array of two lens in cascade. The distance between the first and the second lens is equal to the sum of the focal length of both lenses $f_1 + f_2$, as shown in Figure (A.3). The field in the intermediate focal plane (at f_1 after the first lens), is defined by the Fourier transform as follows:

$$U_2(x,y) = \frac{e^{j2kf_1}}{j\lambda f_1}\hat{U}_1\left(kx/f_1, ky/f_1\right), \tag{A.14}$$

where $\hat{U}_1\left(k_x, k_y\right) = \mathcal{F}\{U_1(x,y)\}$, $\mathcal{F}\{\cdot\}$ is the Continuous Time Fourier Transform (CTFT). Similarly, the field measured at the output focal plane $U_3(x,y)$ is given by the CTFT of $U_2(x,y)$. Therefore, it holds that:

$$U_3(x,y) = -Me^{j2k(f_1+f_2)}U_1(-xM, -yM), \tag{A.15}$$

where $M = f_1/f_2$ is the magnification factor. A detailed derivation of Equation (A.14) is shown in [50]. The derivation uses the Fresnel diffraction formula and assumes the pupil function is uniform and of infinite extent. The properties of the Fourier transform allow to go from Equation (A.14) to Equation (A.15). Note that to deal with lenses of finite extent one can use Raylegh-Sommerfeld solutions, as explained in Chapter 2.

Even though microscope objectives can be approximated with a 4f system, they

are more complex optical systems. An important element in a microscopic objective is a telecentric stop placed in the Fourier plane, as mentioned in [45]. The telecentric stop ensures that the system is object space telecentric, which helps shift-invariance to hold in practice [45].

A.4 Sampling the WDF

This section aims to find the relationship between the measured intensity at the sensor plane (SP) and the WDF computed at the microlens array (NIP). For this derivation, we follow an analysis similar to [57]. However, we propose a more direct approach that avoids the use of the slope-form WDF and we give more details not specified in [57]. For simplicity, we ignore one spatial dimension. We focus on the field modulated by single microlens located at the position s. Then, the intensity at the SP can be computed by multiplication with the microlens transmittance followed by convolution due to propagation. See Figure 2.8 for clarification. Therefore, the field intensity at the SP is:

$$\hat{I}_f(s,\hat{x}) = \left| \int_{-\infty}^{+\infty} U_i(x) T_c(x-s) e^{jk(\hat{x}-x)^2/2z} dx \right|^2, \tag{A.16}$$

where T_c is the complex transmittance of a single microlens, s is the location of the microlens, z is the distance between the microlens and the sensor plane, and $U_i(x)$ is the field just before the microlens. Note that constant factors of the propagation transfer function in Equation (A.10) can be safely ignored. By manipulating Equation (A.16), it follows that:

$$\hat{I}_f(s,\hat{x}) = \left| \int_{-\infty}^{+\infty} U_i(x) e^{jkx^2/2z} T_c(x-s) e^{-jk\hat{x}x/z} dx \right|^2. \tag{A.17}$$

Then, we can name $U_{ic}(x)$ the factor $U_i(x)e^{-jkx^2/2z}$ for simplicity. Furthermore, Equation (A.17) can be written as:

$$\hat{I}_f(s,\hat{x}) = \iint_{-\infty}^{+\infty} U_{ic}(x_1) T_c(x_1-s) e^{-jk\hat{x}x_1/z} U_{ic}^*(x_2) T_c^*(x_2-s) e^{jk\hat{x}x_2/z} dx_1 dx_2. \tag{A.18}$$

The factors in the integral can be reordered as follows:

$$\hat{I}_f(s,\hat{x}) = \iint_{-\infty}^{+\infty} U_{ic}(x_1)U_{ic}^*(x_2)T_c(x_1-s)T_c^*(x_2-s)e^{-j(k/z)\hat{x}(x_1-x_2)}dx_1dx_2. \tag{A.19}$$

We can redefine the integration variables $x_1 := x_1 + x_2/2$,$x_2 := x_1 - x_2/2$ as follows:

$$\begin{aligned}\hat{I}_f(s,\hat{x}) = \iint_{-\infty}^{+\infty} & U_{ic}(x_1+x_2/2)U_{ic}^*(x_1-x_2/2) \\ & T_c(x_1+x_2/2-s)T_c^*(x_1-x_2/2-s)e^{-j2\pi\frac{\hat{x}}{\lambda z}x_2}dx_2dx_1.\end{aligned} \tag{A.20}$$

Note that the integral over x_2 is simply a Fourier transform evaluated at $\frac{\hat{x}}{\lambda z}$. Thus, we can use the definition of WDF and the fact that multiplication in one domain is equal to multiplication in the Fourier conjugate to write the following equation:

$$\hat{I}_f(s,\hat{x}) = \int_{-\infty}^{+\infty} W_{U_{ic}}(x_1,\frac{\hat{x}}{\lambda z}) \circledast_{\hat{x}} W_{T_c}(x_1-s,\frac{\hat{x}}{\lambda z})dx_1, \tag{A.21}$$

where $\circledast_{\hat{x}}$ means convolution along the second variable. The remaining integral is also a convolution. Therefore, we have a 2D dimensional convolution as follows:

$$\hat{I}_f(s,\hat{x}) = W_{U_{ic}}(s,\frac{\hat{x}}{\lambda z}) \circledast W_{T_c}(-s,\frac{\hat{x}}{\lambda z}). \tag{A.22}$$

Now, we can split again the field $U_{ic}(x)$ into two factors using the definition $U_{ic}(x) = U_i(x)e^{-jkx^2/2z}$. The WDF of a field $U_i(x)$ modulated by a complex exponential $e^{-jkx^2/2z}$ translates to a shearing in the WDF domain [55]. Therefore, we have that:

$$\begin{aligned}\hat{I}_f(s,\hat{x}) &= W_{U_i}(s,\frac{\hat{x}}{\lambda z}-\frac{s}{\lambda z}) \circledast W_{T_c}(-s,\frac{\hat{x}}{\lambda z}) \\ &= W_{U_i}(s,\frac{\hat{x}-s}{\lambda z}) \circledast W_{T_c}(-s,\frac{s+(\hat{x}-s)}{\lambda z}).\end{aligned} \tag{A.23}$$

Note that the quantity $(\hat{x}-s)$ is the relative location of a point in the sensor plane with respect to the microlens location s. This relative location is the variable used to describe the light field captured with a camera or microscope. Based on this observation, the

captured light field $L(s,u)$ can be defined as follows:

$$L(s,u) = W_{T_c}(-s, \frac{s}{\lambda z} + u) \circledast W_{U_i}(s,u), \tag{A.24}$$

where $u = \frac{\hat{x}-s}{\lambda z}$. Since the transmittance of a single microlens is $T_c(x) = T(x)e^{-j(x^2)/\lambda 2f}$, the effect of the exponential can be simplified by performing a shearing in the WDF domain. Furthermore, since the distance between the MLA and the sensor z is the same as the focal length of the microlens array for a plenoptic 1.0 design, the shearing in Equation (A.24) cancels out with the effect of the microlens leading to the following equation:

$$L(s,u) = W_T(-s,u) \circledast W_{U_i}(s,u). \tag{A.25}$$

Note that now $W_T(\cdot,\cdot)$ is simply the WDF of $T(x)$, which is an indicator function that is one inside the lens and zero outside the microlens. Finally, the sampling process of the pixels in the sensor can be modelled with a convolution with a kernel $P(\cdot)$ followed by sampling. Also, note that the shown derivation can be easily extended to the case with three dimensions in the object space. Therefore, we have that:

$$L[m,n,p,q] = L(mT_s, nT_t, pT_u, qT_v), \tag{A.26}$$

where $L(s,t,u,v) = K(s,t,u,v) \circledast W_{U_i}(s,t,u,v)$, $K(s,t,u,v) = W_T(-s,-t,u,v) \circledast P(u,v)$, $P(u,v)$ is an indicator function describing each pixel in the sensor, similarly $T(x,y)$ is the indicator function describing a single microlens. The discrete light field $L[m,n,p,q]$ is simply a rearrangement of the aquired 2D image. This is always true if there is no overlapping from light passing through different microlenses, which is ensured if the microlens and objective numerical apertures are matched by design [30], [45]. The sampling intervals $T_s = T_t$ are usually equal to the microlens pitch T, and T_u and T_v are related to the pixel pitch as follows: $T_u = T_v = \frac{T}{N\lambda f}$,where N is the one-dimensional number of pixels under each microlens. Note that $\frac{T}{N}$ is the pixel pitch. In LFM, if the numerical apertures are matched, it also holds that $T_u = T_v = \frac{2NA}{M\lambda N}$, where NA is the numerical aperture of the

microscope and M is the magnification factor.

Appendix B

B.1 Stopping Criteria

In Section 4.3, we proposed to run the reconstruction method for a fixed number of iterations. However, the algorithm can also be stopped using any other stopping criterion based on the computation of the objective function since it is an optimization-based approach. In this section, we investigate using the fidelity term of the objective function to monitor the convergence of the algorithm. The fidelity term measures the error between the input LF image and the synthesised LF image from the reconstructed volume. Thus, when the SIS assumption is not used, the fidelity term is:

$$\mathcal{L}_F = \|\mathbf{H}_\delta \mathbf{f} - \mathbf{g}\|_2. \tag{B.1}$$

Similarly, using the SIS assumption, we have that:

$$\mathcal{L}_F = \|\mathbf{H}_\delta \mathbf{S}_\varphi \mathbf{A}_\varphi \mathbf{f} - \mathbf{g}\|_2. \tag{B.2}$$

We propose to stop the algorithm when the fidelity term is close to stabilizing. Therefore, we can stop the algorithm when $\mathcal{L}_f^k - \mathcal{L}_f^{k-1} < \tau$, where k is the current iteration, and τ is a pre-defined threshold.

To exemplify this idea, we use the pollen dataset. We fixed the threshold τ to 0.1 and ran our ADMM reconstruction methods. In this experiment, the volume is resized to $336 \times 336 \times 92$ voxels, and the corresponding light field image is of size 399×399 pixels to simplify the computation.

Figure B.1 shows that the ADMM methods are close to convergence after approximately 700 iterations, which is when the term $\mathcal{L}_f^k - \mathcal{L}_f^{k-1}$ achieves the threshold τ. Notice

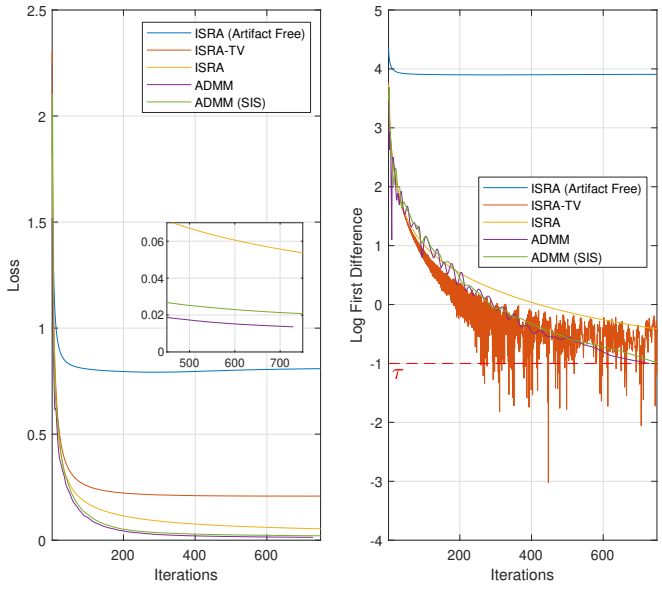


Figure B.1: Fidelity loss. To the left we shown the fidelity loss (error between input light field image and synthesized light field image) of 4 iterative method: ISRA, ISRA with total variation [17], (artifact-free) ISRA [64], ADMM without SIS assumption, and ADMM with SIS assumption. Both ADMM methods where run until the first difference of the loss achieves the pre-set threshold value $\tau = 0.1$. The ISRA methods where run for 700 iterations for a fair comparison. To the right we show the first difference of the loss (the logarithm is used for better visualization).

that even though the fidelity loss from the artifact free ISRA stabilizes faster, its value is very high compare to other methods. Our ADMM with the SIS assumption achieves the smaller error after ~ 700 iterations followed by our second ADMM method.

In Figure B.2, we show a visual comparison of our methods and RL-based methods. The RL methods were evaluated running 700 iterations for a fair comparison. Our experiments show that the ADMM methods perform better than RL-based methods qualitatively and quantitatively. As shown in Figure B.2, the highest PSNR and SSIM were achieved by our ADMM methods. In particular, the highest PSNR is achieved by our ADMM method with the SIS assumption. Notice that our methods can significantly reduce artifacts while ISRA, and TV-ISRA introduce square-like artifacts near $z = 0$. Furthermore, the artifact-free RL is over-smoothing the solution giving a poor performance (PSNR=24.70), which coincides with the behaviour of the loss curve. In addition, recall that our methods have an additional advantage that allow faster computation than RL methods, as described in Section 4.3.

Finally, we remark that in real applications of LFM, volume reconstruction methods

are not run for a large number of iterations since it is computationally demanding when working with large datasets. Therefore, limiting the number of iterations of reconstruction methods is a useful criterion in practice, as proposed in previous works [17], [64], [18].

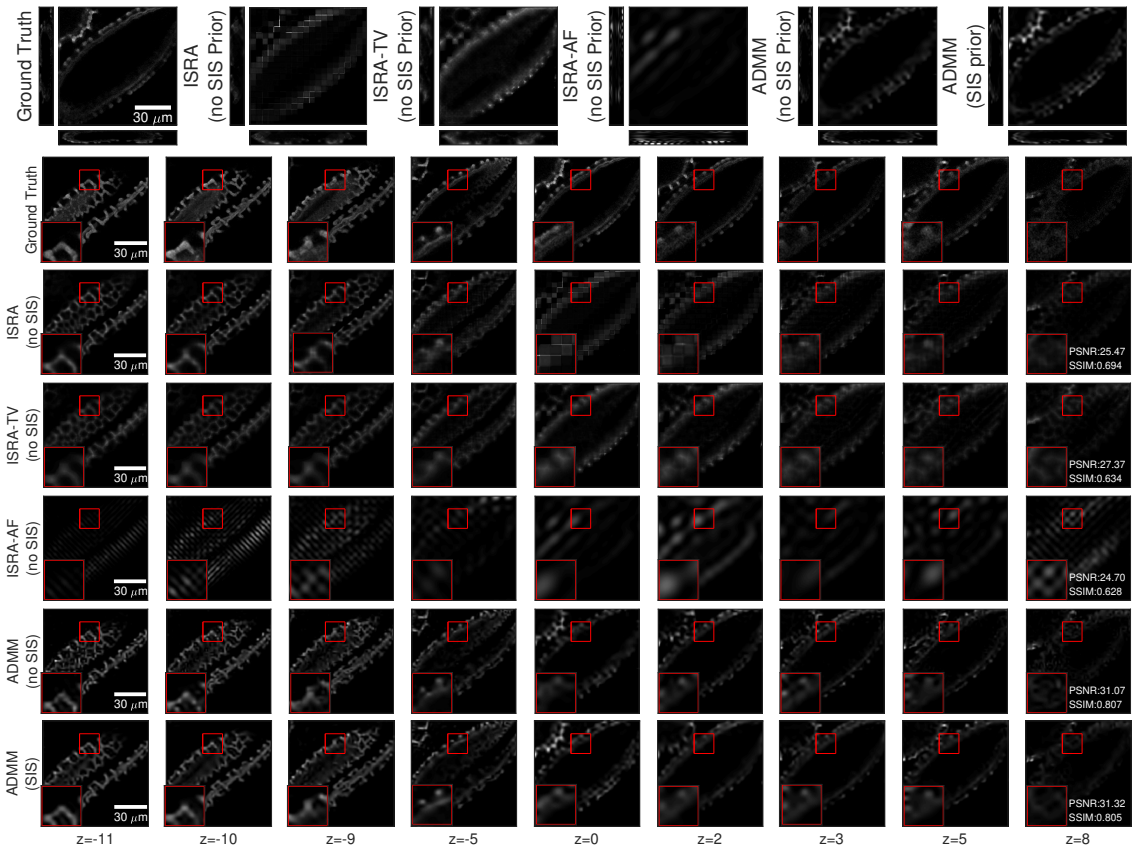


Figure B.2: Reconstruction using a loss-based stopping criterion. Top, we show the in-focus plane, one xz, and one yz slice of the ground truth volume, the reconstruction using ISRA, ISRA with total variation [17], artifact-free (AF) ISRA [64], ADMM without SIS assumption, and ADMM with SIS assumption. All distances are measured in μm. Furthermore, below, we show additional slices for different depths. The shown PSNR and SSIM correspond to the whole volume. The data is from the Lilium Longiflorum Pollen [69] dataset. The light field images were created synthetically from the 3D volumes.

Appendix C

C.1 Experimental Setup

We performed multimodal light field and two-photon imaging of mouse live brain slices in which layer 2/3 excitatory neurons were co-transfected with the fluorophores jGCaMP8f (calcium indicator) and tdTomato (static, morphology label). Imaging was performed with a custom-built microscope featuring coaligned light-field and two-photon imaging paths. Specifically, imaging was performed with a custom-built epifluorescence microscope with an MLA (125 μm pitch, ff/10, RPC Photonics) placed at the imaging plane of a 25×, numerical aperture (NA) =1.0 water immersion objective lens (XLPLN25XSVMP, Olympus), and 180-mm tube lens (TTL180-A, Thorlabs). The MLA was imaged onto a scientific complementary metal-oxide-semiconductor (sCMOS) camera (ORCA Flash 4 V2 with Camera Link, 2048 × 2048 pixels, 6.5 μm pixel size, Hamamatsu) with a 1:1 relay macro lens (Nikon 60 mm f2.8 D AF Micro Nikkor Lens). The MLA was aligned following the advice in [101]. TdTomato fluorescence was excited with a 530 nm LED, and jGCaMP8f fluorescence was excited by a 470 nm LED, both powered by an OptoLED current driver (P1110/002/000, Cairn Research). Light-field images were collected at $50-100$ frames per second using Micromanager 2.0-gamma [101].

C.1.1 Light-field Modality Specifications

Imaging was performed with a custom-built epifluorescence microscope with an MLA (125μm pitch, ff/10, RPC Photonics) placed at the imaging plane of a 25×, numerical aperture (NA) =1.0 water immersion objective lens (XLPLN25XSVMP, Olympus), and 180-mm tube lens (TTL180-A, Thorlabs). The MLA was imaged onto a scientific com-

plementary metal-oxide-semiconductor (sCMOS) camera (ORCA Flash 4 V2 with Camera Link, 2048×2048 pixels, 6.5 μm pixel size, Hamamatsu) with a 1:1 relay macro lens (Nikon 60 mm f2.8 D AF Micro Nikkor Lens). The MLA was aligned following the advice in [101]. TdTomato fluorescence was excited with a 530 nm LED, and jGCaMP8f fluorescence was excited by a 470 nm LED, both powered by an OptoLED current driver (P1110/002/000, Cairn Research). Light-field images were collected at $50 - 100$ frames per second using Micromanager 2.0-gamma [101].

C.1.2 Two-photon Modality Specifications

The two-photon imaging laser (Coherence Monaco 1035-40-40, central wavelength 1035 nm, pulse frequency 10 MHz) was introduced between the light-field imaging system objective and tube lens with a dichroic mirror (DI03-R785-T3 $25 \times 36 \times 3$ mm, Semrock). The laser was focused by a f=30 mm scan lens, f=300 mm tube lens, and the common objective lens (XLPLN25XSVMP, Olympus). The laser was scanned laterally (x,y) by two 3 mm mirrors (6M2003S-S, Cambridge Technologies) driven by high power servo electronics (671315K-1HP, Cambridge Technologies). The plane of focus was adjusted by moving the objective with a stepper motor (SliceScope, Scientifica). tdTomato fluorescence was collected by a $50 \times 70 \times 2$ mm dichroic mirror (T750lpxrxt-UF2, Chroma) positioned directly above the objective back aperture only during two-photon imaging. The objective back pupil was conjugated and demagnified onto the active area of a photomultiplier tube with integrated transimpedance amplifier (PMT, Hamamatsu H10722-20-10MHz). The tdTomato fluorescence passed through a 750 nm shortpass filter (Semrock FF01-750/SP-25) before the PMT. A National Instruments PCI-6110 drove the scan mirrors and digitized the PMT/amplifier output through ScanImage version 3.8 software.

C.1.3 Fluorophore Transfection

Mouse layer 2/3 cortical neurons were transfected via in-utero electroporation (IUE) with soma-targetedw [102] jGCaMP8f [74] (pAAV-CAG-RiboGCaMP8f) and tdTomato (pCAG-tdTomato [103], Addgene). On embryonic day (E)15.5 timed-pregnant female CD-1 mice (Charles River UK) mice were deeply anaesthetized with 2% isoflurane. Uterine horns were exposed and periodically rinsed with warm sterile PBS. Plasmid DNA, 1-2 μg total at a

final concentration of 1 $\mu g/\mu l$ (a 6:1 ratio of jGCaMP8f:tdTomato) diluted in sterile PBS was injected into the lateral ventricle of one cerebral hemisphere of an embryo. Five voltage pulses ($50V$, $50ms$ duration, $1Hz$) were delivered using 5-mm round plate electrodes (ECM 830 electroporator, Harvard Apparatus), with the anode or cathode placed on top of the skull to target the cortex or hippocampus, respectively. Electroporated embryos were placed back into the dam, and allowed to mature to delivery. Brain slices were prepared from electroporated mice at postnatal day (P)12-P30.

C.1.4 Brain Slice Preparation

This study was carried out in accordance with the recommendations of the UK Animals (Scientific Procedures) Act 1986 under Home Office Project and Personal Licenses (project license 70/9095). 400-μm slices were prepared from $12-30$ day old mice Slices were cut in choline chloride cutting solution containing (in mM): 110 choline-Cl, 25 $NaHCO_3$, 20 glucose, 2.5 KCl, 1.25 NaH_2PO_4, 0.5 $CaCl_2$ and 7 $MgSO_4$. After cutting, the slices were transferred to a solution containing (in mM) 125 NaCl, 25 $NaHCO_3$, 20 glucose, 2.5 KCl, 1.25 NaH_2PO_4, 2 $MgSO_4$, 2 $CaCl_2$, adjusted 300 to 310 mOsm/kg, pH 7.3 to 7.4 with HCl at $36°C$. All solutions were oxygenated with $95\%O_2/5\%CO_2$.

Bibliography

[1] K. P. Lillis, A. Eng, J. A. White, and J. Mertz, "Two-photon imaging of spatially extended neuronal network dynamics with high temporal resolution," *Journal of neuroscience methods*, vol. 172, no. 2, pp. 178–184, 2008.

[2] A. J. Sadovsky, P. B. Kruskal, J. M. Kimmel, J. Ostmeyer, F. B. Neubauer, and J. N. MacLean, "Heuristically optimal path scanning for high-speed multiphoton circuit imaging," *Journal of neurophysiology*, vol. 106, no. 3, pp. 1591–1598, 2011.

[3] R. Schuck, M. A. Go, S. Garasto, S. Reynolds, P. L. Dragotti, and S. R. Schultz, "Multiphoton minimal inertia scanning for fast acquisition of neural activity signals," *Journal of neural engineering*, vol. 15, no. 2, p. 025003, 2018.

[4] K. N. S. Nadella, H. Roš, C. Baragli, V. A. Griffiths, G. Konstantinou, T. Koimtzis, G. J. Evans, P. A. Kirkby, and R. A. Silver, "Random-access scanning microscopy for 3D imaging in awake behaving animals," *Nature methods*, vol. 13, no. 12, p. 1001, 2016.

[5] S. J. Yang, W. E. Allen, I. Kauvar, A. S. Andalman, N. P. Young, C. K. Kim, J. H. Marshel, G. Wetzstein, and K. Deisseroth, "Extended field-of-view and increased-signal 3d holographic illumination with time-division multiplexing," *Optics express*, vol. 23, no. 25, pp. 32 573–32 581, 2015.

[6] A. Cheng, J. T. Gonçalves, P. Golshani, K. Arisaka, and C. Portera-Cailliau, "Simultaneous two-photon calcium imaging at different depths with spatiotemporal multiplexing," *Nature methods*, vol. 8, no. 2, p. 139, 2011.

[7] P. Quicke, S. Reynolds, M. Neil, T. Knöpfel, S. R. Schultz, and A. J. Foust, "High speed functional imaging with source localized multifocal two-photon microscopy," *Biomedical optics express*, vol. 9, no. 8, pp. 3678–3693, 2018.

[8] M. Ducros, Y. G. Houssen, J. Bradley, V. de Sars, and S. Charpak, "Encoded multisite two-photon microscopy," *Proceedings of the National Academy of Sciences*, vol. 110, no. 32, pp. 13 138–13 143, 2013.

[9] M. L. Castanares, V. Gautam, J. Drury, H. Bachor, and V. R. Daria, "Efficient multi-site two-photon functional imaging of neuronal circuits," *Biomedical optics express*, vol. 7, no. 12, pp. 5325–5334, 2016.

[10] V. Nikolenko, B. O. Watson, R. Araya, A. Woodruff, D. S. Peterka, and R. Yuste, "SLM microscopy: scanless two-photon imaging and photostimulation using spatial light modulators," *Frontiers in neural circuits*, vol. 2, p. 5, 2008.

[11] S. Bovetti, C. Moretti, S. Zucca, M. Dal Maschio, P. Bonifazi, and T. Fellin, "Simultaneous high-speed imaging and optogenetic inhibition in the intact mouse brain," *Scientific reports*, vol. 7, no. 1, p. 40041, 2017.

[12] P. Pozzi, D. Gandolfi, M. Tognolina, G. Chirico, J. Mapelli, and E. D'Angelo, "High-throughput spatial light modulation two-photon microscopy for fast functional imaging," *Neurophotonics*, vol. 2, no. 1, p. 015005, 2015.

[13] R. Prevedel, A. J. Verhoef, A. J. Pernia-Andrade, S. Weisenburger, B. S. Huang, T. Nöbauer, A. Fernández, J. E. Delcour, P. Golshani, A. Baltuska *et al.*, "Fast volumetric calcium imaging across multiple cortical layers using sculpted light," *Nature methods*, vol. 13, no. 12, p. 1021, 2016.

[14] A. J. Foust, V. Zampini, D. Tanese, E. Papagiakoumou, and V. Emiliani, "Computer-generated holography enhances voltage dye fluorescence discrimination in adjacent neuronal structures," *Neurophotonics*, vol. 2, no. 2, p. 021007, 2015.

[15] D. Tanese, J.-Y. Weng, V. Zampini, V. de Sars, M. Canepari, B. J. Rozsa, V. Emiliani, and D. Zecevic, "Imaging membrane potential changes from dendritic spines using computer-generated holography," *Neurophotonics*, vol. 4, no. 3, p. 031211, 2017.

[16] F. Anselmi, C. Ventalon, A. Bègue, D. Ogden, and V. Emiliani, "Three-dimensional imaging and photostimulation by remote-focusing and holographic light patterning," *Proceedings of the National Academy of Sciences*, vol. 108, no. 49, pp. 19 504–19 509, 2011.

[17] T. Nöbauer, O. Skocek, A. P.-A. J., L. Weilguny, F. M. Traub, M. I. Molodtsov, and A. Vaziri, "Video rate volumetric ca2+ imaging across cortex using seeded iterative demixing (sid) microscopy," *Nature Methods*, vol. 14, p. 811, 2017. [Online]. Available: https://doi.org/10.1038/nmeth.4341

[18] R. Prevedel, Y. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. SchrÃ¶del, R. Raskar, M. Zimmer, E. S. Boyden, and A. Vaziri, "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," *Nature methods*, vol. 11, no. 7, p. 727, 2014.

[19] N. C. Pegard, H.-Y. Liu, N. Antipa, M. Gerlock, H. Adesnik, and L. Waller, "Compressive light-field microscopy for 3D neural activity recording," *Optica*, vol. 3, no. 5, pp. 517–524, May 2016. [Online]. Available: http://www.osapublishing.org/optica/abstract.cfm?URI=optica-3-5-517

[20] S. R. Schultz, C. S. Copeland, A. J. Foust, P. Quicke, and R. Schuck, "Advances in two-photon scanning and scanless microscopy technologies for functional neural circuit imaging," *Proceedings of the IEEE*, vol. 105, no. 1, pp. 139–157, 2017.

[21] S. Weisenburger and A. Vaziri, "A guide to emerging technologies for large-scale and whole brain optical imaging of neuronal activity," *Annual review of neuroscience*, vol. 41, p. 431, 2018.

[22] H. Miyazawa, K. Okumura, K. Hiyoshi, K. Maruyama, H. Kakinuma, R. Amo, H. Okamoto, K. Yamasu, and S. Tsuda, "Optical interrogation of neuronal circuitry

in zebrafish using genetically encoded voltage indicators," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[23] M. Inoue, A. Takeuchi, S. Manita, S.-i. Horigane, M. Sakamoto, R. Kawakami, K. Yamaguchi, K. Otomo, H. Yokoyama, R. Kim *et al.*, "Rational engineering of xcamps, a multicolor geci suite for in vivo imaging of complex brain circuit dynamics," *Cell*, vol. 177, no. 5, pp. 1346–1360, 2019.

[24] S. Weisenburger, F. Tejera, J. Demas, B. Chen, J. Manley, F. T. Sparks, F. M. Traub, T. Daigle, H. Zeng, A. Losonczy *et al.*, "Volumetric ca2+ imaging in the mouse brain using hybrid multiplexed sculpted light microscopy," *Cell*, vol. 177, no. 4, pp. 1050–1066, 2019.

[25] T. Knöpfel and C. Song, "Optical voltage imaging in neurons: moving from technology development to practical tool," *Nature Reviews Neuroscience*, vol. 20, no. 12, pp. 719–727, 2019.

[26] L. Cong, Z. Wang, Y. Chai, W. Hang, C. Shang, W. Yang, L. Bai, J. Du, K. Wang, and Q. Wen, "Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (danio rerio)," *Elife*, vol. 6, 2017.

[27] S. Aimon, T. Katsuki, L. Grosenick, M. Broxton, K. Deisseroth, T. Sejnowski, and R. Greenspan, "Fast whole brain imaging in adult drosophila during response to stimuli and behavior," *Doi. Org*, vol. 33803, 2017.

[28] D. G. Dansereau, "Plenoptic signal processing for robust vision in field robotics," Ph.D. dissertation, 2014.

[29] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Australian Centre for Field Robot., Mech. Mechatron. Eng. Univ. of Sydney, Sydney, NSW, Australia. Los Alamitos, CA, USA: IEEE Computer Society, 23-28 June 2013 2013, pp. 1027–34, t3: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [Online]. Available: http://dx.doi.org/10.1109/CVPR.2013.137

[30] R. Ng, M. Levoy, M. B. 'edif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Tech. Rep., apr 2005. [Online]. Available: http://graphics.stanford.edu/papers/lfcamera/

[31] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42. [Online]. Available: http://doi.acm.org/10.1145/237170.237199

[32] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 43–54. [Online]. Available: http://doi.acm.org/10.1145/237170.237200

[33] I. Ihrke, J. Restrepo, and L. Mignard-Debise, "Principles of light field imaging: Briefly revisiting 25 years of research," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 59–69, 2016.

[34] Y. Zhang, H. Lv, Y. Liu, H. Wang, X. Wang, Q. Huang, X. Xiang, and Q. Dai, "Light-field depth estimation via epipolar plane image analysis and locally linear embedding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 739–747, 2017.

[35] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.

[36] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, 2014.

[37] S. Heber and T. Pock, "Shape from light field meets robust pca," in *European Conference on Computer Vision*. Springer, 2014, pp. 751–767.

[38] R. Farrugia and C. Guillemot, "Light field super-resolution using a low-rank prior and deep convolutional neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, Jan 18, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8620368

[39] B. Goldlucke, O. Klehm, S. Wanner, and E. Eisemann, "Plenoptic cameras," in *Digital Representations of the Real World: How to Capture, Model, and Render Visual Reality*, 1st ed., M. A. Magnor, O. Grau, O. Sorkine-Hornung, and C. Theobalt, Eds. A K Peters/CRC Press, 2015, ch. 5, http://www.crcpress.com/product/isbn/9781482243819. [Online]. Available: http://graphics.tudelft.nl/Publications-new/2015/GKWE15

[40] D. Lanman, R. Raskar, A. Agrawal, and G. Taubin, "Shield fields: modeling and capturing 3d occluders," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5, pp. 1–10, 2008.

[41] A. Manakov, J. Restrepo, O. Klehm, R. Hegedus, E. Eisemann, H.-P. Seidel, and I. Ihrke, "A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 47–1, 2013.

[42] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *2009 IEEE International Conference on Computational Photography (ICCP)*, 2009, pp. 1–8.

[43] J. Page, F. Saltarin, Y. Belyaev, R. Lyck, and P. Favaro, "Learning to reconstruct confocal microscope stacks from single light field images," 2020.

[44] T. Georgiev and A. Lumsdaine, "Resolution in plenoptic cameras," in *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest*. Optica Publishing Group, 2009, p. CTuB3. [Online]. Available: http://opg.optica.org/abstract.cfm?URI=COSI-2009-CTuB3

[45] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 924–34, 2006. [Online]. Available: http://dx.doi.org/10.1145/1141911.1141976

[46] Z. Xue, "Sampling models in light fields," Ph.D. dissertation, 2016.

[47] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge,UK: Cambridge University Press, 2003.

[48] L. Novotny, "Lecture notes on electromagnetic fields and waves," *Switzerland: ETH Zurich*, pp. 82–83, 2013.

[49] J. W. Goodman, *Introduction to Fourier Optics 2nd Edition.* New York: McGraw-Hill, 1998.

[50] M. Gu, *Advanced optical imaging theory.* Berlin; London: Springer, 2000.

[51] M. Broxton, L. Grosenick, S. Yang, A. A. N. Cohen and, K. Deisseroth, and M. Levoy, "Wave optics theory and 3-d deconvolution for the light field microscope," *Optics express*, vol. 21, no. 21, pp. 25 418–25 439, 2013.

[52] M. Testorf, B. Hennelly, and J. Ojeda-Castañeda, *Phase-space optics: fundamentals and applications.* McGraw-Hill Education, 2010.

[53] M. J. Bastiaans, "Application of the wigner distribution function to partially coherent light," *Journal of the Optical Society of America A*, vol. 3, no. 8, pp. 1227–1238, 1986, j2: J. Opt. Soc. Am. A. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-3-8-1227

[54] H. Liu, E. Jonas, L. Tian, J. Zhong, B. Recht, and L. Waller, "3D imaging in volumetric scattering media using phase-space measurements," *Optics express*, vol. 23, no. 11, pp. 14 461–14 471, Jun 2015. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-23-11-14461

[55] A. Papoulis, "Ambiguity function in fourier optics∗," *J.Opt.Soc.Am.*, vol. 64, no. 6, pp. 779–788, Jun 1974. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=josa-64-6-779

[56] D. Dragoman, "Applications of the wigner distribution function in signal processing," *EURASIP J.Adv.Signal Process*, vol. 2005, pp. 1520–1534, 2005. [Online]. Available: https://doi.org/10.1155/ASP.2005.1520

[57] Z. Zhang and M. Levoy, "Wigner distributions and how they relate to the light field," pp. 1–10, 2009.

[58] Z. Lu, J. Wu, H. Qiao, Y. Zhou, T. Yan, Z. Zhou, X. Zhang, J. Fan, and Q. Dai, "Phase-space deconvolution for light field microscopy," *Opt.Express*, vol. 27, no. 13, pp. 18 131–18 145, Jun 2019. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-27-13-18131

[59] P. Quicke, C. L. Howe, P. Song, H. V. Jadan, C. Song, T. Knöpfel, M. Neil, P. L. Dragotti, S. R. Schultz, and A. J. Foust, "Subcellular resolution three-dimensional light-field imaging with genetically encoded voltage indicators," *Neurophotonics*, vol. 7, no. 3, p. 035006, 2020.

[60] C. L. Howe, P. Quicke, P. Song, H. V. Jadan, P. L. Dragotti, and A. J. Foust, "Comparing volumetric reconstruction algorithms for light field imaging of high signal-to-noise ratio neuronal calcium transients," *bioRxiv*, 2020.

[61] W. H. Richardson, "Bayesian-based iterative method of image restoration," *JoSA*, vol. 62, no. 1, pp. 55–59, 1972.

[62] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The astronomical journal*, vol. 79, p. 745, 1974.

[63] M. E. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorthm suitable for volume ect," *IEEE transactions on medical imaging*, vol. 5, no. 2, pp. 61–66, 1986.

[64] A. Stefanoiu, J. Page, P. Symvoulidis, G. G. Westmeyer, and T. Lasser, "Artifact-free deconvolution in light field microscopy," *Opt. Express*, vol. 27, no. 22, pp. 31 644–31 666, Oct 2019. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-27-22-31644

[65] Z. Wang, L. Zhu, H. Zhang, G. Li, C. Yi, Y. Li, Y. Yang, Y. Ding, M. Zhen, S. Gao *et al.*, "Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning," *Nature Methods*, vol. 18, no. 5, pp. 551–556, 2021.

[66] Z. Wang, H. Zhang, Y. Yang, Y. Li, S. Gao, and P. Fei, "Deep learning light field microscopy for rapid four-dimensional imaging of behaving animals," *bioRxiv*, p. 432807, 2018. [Online]. Available: http://biorxiv.org/content/early/2018/10/02/432807.1.abstract

[67] N. Wagner, F. Beuttenmueller, N. Norlin, J. Gierten, J. C. Boffi, J. Wittbrodt, M. Weigert, L. Hufnagel, R. Prevedel, and A. Kreshuk, "Deep learning-enhanced light-field imaging with continuous validation," *Nature Methods*, vol. 18, no. 5, pp. 557–563, 2021.

[68] N. Chacko, M. Liebling, and T. Blu, "Discretization of continuous convolution operators for accurate modeling of wave propagation in digital holography," *JOSA A*, vol. 30, no. 10, pp. 2012–2020, 2013.

[69] A. Carpenter, "Drosophila melanogaster kc167 cell." [Online]. Available: http://cellimagelibrary.org/images/21766

[70] X. Li, H. Qiao, J. Wu, Z. Lu, T. Yan, R. Zhang, X. Zhang, and Q. Dai, "DeepLFM: Deep learning-based 3D reconstruction for light field microscopy," pp. NM3C–2, 04/14 2019, j2: NTM; T3: The Optical Society. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=NTM-2019-NM3C.2

[71] M. Eberhart, "Efficient algorithm for calculating transposed psf matrices for 3d light field deconvolution," *arXiv e-prints*, pp. arXiv–2003, 2020.

[72] L. Donati, E. Soubies, and M. Unser, "Inner-loop-free ADMM for cryo-em," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 307–311.

[73] L. Madisen, A. R. Garner, D. Shimaoka, A. S. Chuong, N. C. Klapoetke, L. Li, A. Van Der Bourg, Y. Niino, L. Egolf, C. Monetti *et al.*, "Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance," *Neuron*, vol. 85, no. 5, pp. 942–958, 2015.

[74] Y. Zhang, M. Rózsa, D. Bushey, J. Zheng, D. Reep, Y. Liang, G. J. Broussard, A. Tsang, G. Tsegaye, R. Patel, S. Narayan, J.-X. Lim, R. Zhang, M. B. Ahrens, G. C. Turner, S. S.-H. Wang, K. Svoboda, W. Korff, E. R. Schreiter, J. P. Hasseman, I. Kolb, and L. L. Looger, "jgcamp8 fast genetically encoded calcium indicators," 12 2020. [Online]. Available: https://janelia.figshare.com/articles/online_resource/jGCaMP8_Fast_Genetically_Encoded_Calcium_Indicators/13148243

[75] S. W. Hasinoff, "Photon, poisson noise." *Computer Vision, A Reference Guide*, vol. 4, 2014.

[76] Y. Qian, D. M. O. Cosio, K. D. Piatkevich, S. Aufmkolk, W.-C. Su, O. T. Celiker, A. Schohl, M. H. Murdock, A. Aggarwal, Y.-F. Chang *et al.*, "Improved genetically encoded near-infrared fluorescent calcium ion indicators for in vivo imaging," *PLoS Biology*, vol. 18, no. 11, p. e3000965, 2020.

[77] Y. Qian, K. D. Piatkevich, B. Mc Larney, A. S. Abdelfattah, S. Mehta, M. H. Murdock, S. Gottschalk, R. S. Molina, W. Zhang, Y. Chen *et al.*, "A genetically encoded near-infrared fluorescent calcium ion indicator," *Nature methods*, vol. 16, no. 2, pp. 171–174, 2019.

[78] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9446–9454.

[79] P. Cascarano, A. Sebastiani, M. C. Comes, G. Franchini, and F. Porta, "Combining weighted total variation and deep image prior for natural and medical image restoration via admm," *arXiv preprint arXiv:2009.11380*, 2020.

[80] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.

[81] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

[82] M. Carter and J. Shieh, "Chapter 5 - microscopy," in *Guide to Research Techniques in Neuroscience (Second Edition)*, second edition ed., M. Carter and J. Shieh, Eds. San Diego: Academic Press, 2015, pp. 117–144. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128005118000058

[83] L. A. Ware, "Three photons are better than two," 2014.

[84] H. Verinaz-Jadan, P. Song, C. L. Howe, A. J. Foust, and P. L. Dragotti, "Shift-invariant-subspace discretization and volume reconstruction for light field microscopy," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 286–301, 2022.

[85] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th international conference on international conference on machine learning*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, pp. 399–406.

[86] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[87] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: https://doi.org/10.1137/080716542

[88] X. Chen, J. Liu, Z. Wang, and W. Yin, "Theoretical linear convergence of unfolded ista and its practical weights and thresholds," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[89] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*. PMLR, 2013, pp. 1310–1318.

[90] A. Gruslys, R. Munos, I. Danihelka, M. Lanctot, and A. Graves, "Memory-efficient backpropagation through time," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[91] G. Kwon, C. Han, and D.-s. Kim, "Generation of 3D brain mri using auto-encoding generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 118–126.

[92] C. Ledig, L. Theis, F. HuszÃ¡r, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114.

[93] S. Lunz, O. Öktem, and C.-B. Schönlieb, "Adversarial regularizers in inverse problems," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018.

[94] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.

[95] S. Lim, H. Park, S.-E. Lee, S. Chang, B. Sim, and J. C. Ye, "Cyclegan with a blur kernel for deconvolution microscopy: Optimal transport geometry," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1127–1138, 2020.

[96] H. Li, C. Guo, and S. Jia, "High-resolution light-field microscopy," in *Frontiers in Optics 2017*. Optical Society of America, 2017, p. FW6D.3. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=FiO-2017-FW6D.3

[97] C. Guo, W. Liu, X. Hua, H. Li, and S. Jia, "Fourier light-field microscopy," *Opt. Express*, vol. 27, no. 18, pp. 25 573–25 594, Sep 2019. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-27-18-25573

[98] J. Goodman, *Introduction to Fourier Optics*, ser. McGraw-Hill physical and quantum electronics series. W. H. Freeman, 2005. [Online]. Available: https://books.google.co.uk/books?id=ow5xs_Rtt9AC

[99] E. Soubies, T.-A. Pham, and M. Unser, "Efficient inversion of multiple-scattering model for optical diffraction tomography," *Opt. Express*, vol. 25, no. 18, pp. 21 786–21 800, Sep 2017. [Online]. Available: http://opg.optica.org/oe/abstract.cfm?URI=oe-25-18-21786

[100] D. Daly, *Microlens Arrays*, ser. Microlens Arrays. Taylor & Francis, 2000. [Online]. Available: https://books.google.co.uk/books?id=xCWONq0kmr0C

[101] A. D. Edelstein, M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, and N. Stuurman, "Advanced methods of microscope control using μmanager software," *Journal of biological methods*, vol. 1, no. 2, 2014.

[102] Y. Chen, H. Jang, P. W. Spratt, S. Kosar, D. E. Taylor, R. A. Essner, L. Bai, D. E. Leib, T.-W. Kuo, Y.-C. Lin *et al.*, "Soma-targeted imaging of neural circuits by ribosome tethering," *Neuron*, vol. 107, no. 3, pp. 454–469, 2020.

[103] M. Pathania, J. Torres-Reveron, L. Yan, T. Kimura, T. V. Lin, V. Gordon, Z.-Q. Teng, X. Zhao, T. A. Fulga, D. Van Vactor *et al.*, "mir-132 enhances dendritic morphogenesis, spine density, synaptic integration, and survival of newborn olfactory bulb neurons," *PloS one*, vol. 7, no. 5, p. e38174, 2012.