Quantitative Analysis of Mitochondrial Shape and Ultrastructure Based on Expansion Microscopy

Alexander Sauer

Keble College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2024

Abstract

Expansion microscopy (ExM) and the recently developed pan-staining allow researchers to produce images of cells at super-resolution similar to electron microscopy (EM). ExM does not require specialized hardware, making super-resolution microscopy more accessible to biology laboratories and enabling significantly higher throughput than electron microscopy (EM).

However, large amounts of 3D imaged samples pose a challenge for transforming the images into biological insights as manually annotating and analysing them is infeasible. This thesis therefore develops the first machine learning pipeline that automatically segments mitochondria and their ultrastructure in pan-stained ExM images and enables a scalable quantitative analysis. ExM allows combining the pan-staining with additional dyes and immunolabeling and I demonstrate how this additional information can be utilised to decrease the amount of human annotation required to train a segmentation model. Additionally, I develop a novel algorithm that exploits inter-annotator disagreement to improve the performance of semi-supervised segmentation models. I also present a method to automatically learn a prior for biologically plausible segmentation masks and apply it selectively to regions of high uncertainty, where it corrects segmentation errors that violate biological constraints.

Finally, I apply the segmentation pipeline to over 60,000 mitochondria to characterize their shape and ultrastructure in HeLa cells, investigate the morphological effects of different drug treatments and quantify the change in the mitochondrial ultrastructure induced by acute kidney injury in mouse kidney tissue.

Quantitative Analysis of Mitochondrial Shape and Ultrastructure Based on Expansion Microscopy



Alexander Sauer

Keble College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2024

Acknowledgements

I thank my supervisor Professor Jens Rittscher for his advice, guidance and feedback throughout this project. Moreover, I want to thank Professor Joerg Bewersdorf and Yuan Tian from Yale University for numerous discussions, providing the imaging data for this project and trusting me with the analysis.
I also thank my family and friends for their support, especially Varnika, Peter and Carolin.
Finally, I thank the Health Data Science Centre of Doctoral Training and EPSRC for providing the funding for the project.

Abstract

Expansion microscopy (ExM) and the recently developed pan-staining allow researchers to produce images of cells at super-resolution similar to electron microscopy (EM). ExM does not require specialized hardware, making super-resolution microscopy more accessible to biology laboratories and enabling significantly higher throughput than electron microscopy (EM).

However, large amounts of 3D imaged samples pose a challenge for transforming the images into biological insights as manually annotating and analysing them is infeasible. This thesis therefore develops the first machine learning pipeline that automatically segments mitochondria and their ultrastructure in pan-stained ExM images and enables a scalable quantitative analysis. ExM allows combining the pan-staining with additional dyes and immunolabeling and I demonstrate how this additional information can be utilised to decrease the amount of human annotation required to train a segmentation model. Additionally, I develop a novel algorithm that exploits inter-annotator disagreement to improve the performance of semi-supervised segmentation models. I also present a method to automatically learn a prior for biologically plausible segmentation masks and apply it selectively to regions of high uncertainty, where it corrects segmentation errors that violate biological constraints.

Finally, I apply the segmentation pipeline to over 60,000 mitochondria to characterize their shape and ultrastructure in HeLa cells, investigate the morphological effects of different drug treatments and quantify the change in the mitochondrial ultrastructure induced by acute kidney injury in mouse kidney tissue.

Contents

List of Figures **vii**

List of Abbreviations **ix**

1 Introduction **1**
1.1 Overview . . . 1
1.1.1 Seeing the Invisible . . . 2
1.1.2 Overcoming the Diffraction Limit . . . 2
1.1.3 The Evolution of Machine Learning . . . 4
1.1.4 Advancing Microscopy with Machine Learning . . . 6
1.1.5 Understanding Mitochondria . . . 7
1.2 Aims and Contributions . . . 9
1.3 Publications . . . 12

2 Background **13**
2.1 Biological Context and Imaging Technique . . . 13
2.1.1 Mitochondria and their Role in the Cell . . . 13
2.1.2 Expansion Microscopy . . . 15
2.2 Machine Learning and Image Segmentation . . . 18
2.2.1 Overview . . . 18
2.2.2 Deep Learning Based Approaches . . . 19
2.2.3 Semantic Segmentation . . . 22
2.2.4 Pre-training and Semi-supervised Approaches . . . 23
2.3 Automated Analysis of Mitochondria . . . 26
2.3.1 Analysis of Mitochondria in Light Microscopy . . . 26
2.3.2 Analysis of Mitochondria in Electron Microscopy . . . 28
2.3.3 Analysis of Mitochondrial Ultrastructure . . . 31

3 Pan-Stained ExM and Uncertainty Analysis **34**
3.1 Introduction . . . 35
3.2 Expansion Microscopy with PanExM . . . 35
3.3 Difference to EM . . . 38
3.4 Annotation and Analysis of Uncertainty . . . 39
3.5 Multiple Annotators for Semi-Supervised Learning . . . 44
3.5.1 Method . . . 46

3.5.2 Experiments . . . 49
3.5.3 Results . . . 50
3.5.4 Conclusion . . . 52
3.6 Analysis of Dice Loss with Noisy Labels . . . 52
3.6.1 Theory . . . 53
3.6.2 Experiments . . . 55
3.7 Discussion . . . 57

4 Mitochondria Segmentation **59**
4.1 Introduction . . . 60
4.2 MitoTracker as a Shape Prior . . . 61
4.2.1 Mask Based Approach . . . 63
4.2.2 Architecture . . . 65
4.2.3 Pre-processing and Data Augmentation . . . 66
4.2.4 Results . . . 68
4.3 Drug Treated Cells . . . 74
4.3.1 Drug Inspection . . . 74
4.3.2 Drug-Conditional Model . . . 77
4.3.3 Experimental Results . . . 80
4.4 Mitochondria in Tissue . . . 83
4.4.1 Kidney Tissue Data . . . 85
4.4.2 Training with Sparse Antibodies . . . 87
4.4.3 Results . . . 93
4.5 Antibody Requirements for Future Applications . . . 93
4.6 Discussion . . . 95

5 Post-Processing and Parameter Extraction **99**
5.1 Introduction . . . 100
5.2 Refining Biologically Inconsistent Segmentation Masks . . . 100
5.2.1 Uncertainty Detection . . . 103
5.2.2 Shape Constraints . . . 104
5.2.3 Method . . . 104
5.2.4 Experiments . . . 109
5.2.5 Postprocessing . . . 115
5.3 Parameter Extraction . . . 116
5.3.1 Instance Segmentation . . . 116
5.3.2 Skeletonisation . . . 117
5.3.3 Length and Diameter Calculation . . . 119
5.4 Discussion . . . 121

6 Application to Biological Studies **123**
6.1 Mitochondria Analysis in HeLa Cells 124
6.1.1 Global Descriptors 125
6.1.2 Ultrastructure Analysis 129
6.1.3 Statistical Analysis 133
6.2 Mitochondria in Kidney Tissue 135
6.3 Immunolabeling Analysis 136
6.4 Discussion 141

7 Conclusion and Future Work **144**

References **150**

List of Figures

2.1 Schematic Illustration of a Mitochondrion 15
2.2 Mitochondria Imaged with a Fluorescent Microscope 15
2.3 Illustration of a CNN 20
2.4 Example Slice of the EPFL Hippocampus Dataset 28
2.5 Example from the MitoEM Dataset 30

3.1 Example Clearly Imaged Mitochondria 37
3.2 Example Poorly Imaged Mitochondria 37
3.3 Cyptoplasm Patterns 37
3.4 PanExm with Additional Dye and Immunolabeling 38
3.5 Visual Comparison to EM 39
3.6 Annotation Protocol 40
3.7 Annotator Agreement Across Categories 43
3.8 Disagreement at Boundaries 44
3.9 Method Overview 47
3.10 Example from ISIC and QUBIQ Datasets 50
3.11 MitoEM and MoNuSeg Datasets with Noise 55
3.12 Comparison of Dice Performance 56

4.1 2D vs 3D Mitochondria Shape 62
4.2 Workflow MitoTracker Based Model 63
4.3 Visualisation of MitoTracker Derived Masks 65
4.4 Diversity in FoV Appearances 67
4.5 Visualisation of Data Augmentation 68
4.6 Results Annotation Efficiency 69
4.7 Results MitoTracker for Sparse Annotations 70
4.8 Ensembling Results 71
4.9 3D Visualisation 73
4.10 Example Mitochondria from the Control group (DMSO) 75
4.11 Examples of Mitochondria Treated with Antimycin A 76
4.12 Examples of Mitochondria Treated with HBSS 76
4.13 Examples of Mitochondria Treated with Oligomycin 77
4.14 Drug Embedding 79
4.15 Evaluation of Conditional Model 81
4.16 Effect of Perturbing Embedding 81

4.17 Drug Treatment Segmentation Results . . . 82
4.18 3D Rendering Drug Segmentation . . . 83
4.19 Kidney WT and AKI Pan-staining . . . 86
4.20 Zoomed in Kidney WT and AKI Pan-staining and Anti-bodies . . . 87
4.21 Antibody Based Training . . . 88
4.22 Lamellar Cristae . . . 90
4.23 Fully Unsupervised Outline Prediction . . . 92
4.24 Finetuning of Kidney Predictions with HeLa Model . . . 92
4.25 Kidney Antibody Localisation . . . 94
4.26 Potential Biases . . . 97

5.1 Comparison of a Biologically Valid Segmentation Mask, an Invalid Mask and the Proposed Reconstruction . . . 101
5.2 Poor Segmentation as a Consequence of Low SNR . . . 102
5.3 Reconstruction Method Overview . . . 105
5.4 Simulation of Unclear Staining . . . 110
5.5 Validity Metrics . . . 111
5.6 Reconstruction Results . . . 114
5.7 Instance Separation . . . 116
5.8 Skeleton Comparison . . . 118
5.9 Convex Skeleton Branch Pruning . . . 119
5.10 Skeleton Topology Refinement . . . 120

6.1 Distribution of Length . . . 125
6.2 Mitochondria Networks . . . 126
6.3 Mitochondria Topology . . . 127
6.4 Mitochondria Volume . . . 127
6.5 Distribution of Diameter . . . 129
6.6 Bivariate Plot of Length and Sphericity . . . 130
6.7 Distribution of Cristae Area . . . 130
6.8 Distribution of Ambiguous Area . . . 131
6.9 Distribution of Cristae Distance . . . 132
6.10 Relation Macro Parameter and Cristae . . . 132
6.11 PCA Plot . . . 134
6.12 Quantification WT Mitochondria . . . 135
6.13 WT and AKI Results . . . 137
6.14 WT and AKI Morphology Comparison . . . 137
6.15 Antibody Labelling . . . 138
6.16 Antibody Intensities . . . 139
6.17 Immunolabeling Analysis . . . 140
6.18 Antibody Distance from Cristae Junctions and Visualisation . . . 141

List of Abbreviations

ACM . . . . . Active contour model

AKI . . . . . Acute kidney injury

ATP . . . . . Adenosine triphosphate

BCE . . . . . Binary cross entropy

BNN . . . . . Bayesian neural network

CE . . . . . Cross entropy

CNN . . . . . Convolutional neural network

COX . . . . . Cytochrome c oxidase

CRF . . . . . Conditional random fields

CV . . . . . Computer vision

DL . . . . . Deep learning

DMSO . . . . . Dimethylsulfoxide

EM . . . . . Electron microscopy

ER . . . . . Endoplasmic reticulum

ETC . . . . . Electron transport chain

ExM . . . . . Expansion microscopy

FIB-SEM . . . Focused ion beam scanning electron microscope

FoV . . . . . Field of view

GAN . . . . . Generative adversarial network

GPU . . . . . Graphics processing unit

HBSS . . . . . Hank's balanced salt solution

HCS . . . . . High-content screening

IoU . . . . . Intersection over Union

ISIC . . . . . International skin imaging collaboration

MAE . . . . . Masked autoencoder

MICOS . . . . Mitochondria contact site and cristae organisation system

ML . . . . . Machine learning

MLP . . . . . Multi-Layer-Perceptron

NHS ester . . N-Hydroxysuccinimide ester

PanExM . . . Pan-stained expansion microscopy (M'Saad et al., 2020)

PCA . . . . . Principal component analysis

SD . . . . . Standard deviation

TSS . . . . . Target specific staining

TOM . . . . . Translocase of the outer mitochondrial membrane

ViT . . . . . Vision transformer

WT . . . . . Wild type

1 Introduction

1.1 Overview

Vision is the primary sense through which humans perceive and engage with the world around them. This is also true for the scientific process and visualising objects and processes has been a primary source of discovery in biology. While it is possible to study animals and plants with the naked eye, researchers learned in the 18th century that there is a whole new world that becomes only accessible with the help of some magnification. These magnification devices started as simple drops of glass in the 17th century but developed into high-tech microscopes built on the fundamental insights of modern physics. Today's microscopes are capable of providing a resolution down to the nano-meter scale and output more data than can be analysed by any human biologist.

However, modern computers with advances in hardware and algorithms in the form of machine learning provide the opportunity to automate the analysis of microscopy data. The next sections summarise the developments in microscopy and their symbiosis with machine learning. Subsequently, I provide an outlook of this thesis in which I develop an analysis pipeline to analyse mitochondria and their ultrastructure at scale.

1.1.1 Seeing the Invisible

With the first microscopes going back as far as the 17th century in the form of Galileo's *occhiolino*, they allow researchers to investigate the building blocks of life that are too small to be observed by the naked eyes. Antoni van Leeuwenhoek (1632-1723), for example, was able to build microscopes with a magnification factor of over 250 and was able to observe bacteria, yeast and red blood cells for the first time (Allen, 2015). These discoveries lay the foundations for the new field of microbiology.

Biological cells are mostly colourless and therefore show little contrast under a microscope. Scientists, therefore, developed different staining and dyes which allow to distinguish between different parts of the cell. Famous examples include the dyes synthesized by Paul Ehrlich leading to the first treatment for syphilis and Christian Gram allowing to differentiate different bacteria (Allen, 2015).

By allowing to visualise biological samples, microscopes allow researchers to settles debates and support abstract theories by empirical observations. Examples include the rejection of miasma theory in favour of the germ theory of disease. Miasma theory was a popular medical theory until the 19th century that explained the origin and propagation of diseases like cholera and the Black Death by 'bad air' stemming from rotting organic matter (Last et al., 2018). Being able to see microorganisms was a decisive factor to convince scientists of the germ theory of disease as advocated prominently by Louis Pasteur and Robert Koch (Gaynes, 2023).

The first observation of mitochondria occurred in the 1840s and in 1898 the term mitochondria was coined by Carl Benda (Benda, 1898). It took until 1910 before Kingsbury suggested that the cellular respiration occurs actually in the mitochondria (Pagliarini et al., 2013). However, the microscopes at that time were not powerful enough to allow an investigation of the inner structure of the mitochondria.

1.1.2 Overcoming the Diffraction Limit

From the mid 19th century, Carl Zeiss improved the production process of microscopes and his business partner Ernst Abbe investigated the underlying scientific principles and established that the resolution of the microscope is fundamentally

limited by the wavelength of the illuminating light. The resolution is defined as the ability to resolve two point sources of equal brightness and with visible light, this diffraction limit for optical microscopes lies around 200-300nm (Abbe, 1882).

This limit was only overcome at the beginning of the 20th century with the invention of the electron microscope (EM) and other more recent super-resolution techniques (Schermelleh et al., 2019).

EM is build on the idea of exploiting the much smaller wave-lengths of electrons which was postulated by Louis de Groglie around 1925. Building on the cathode ray tube developed by Karl Braun in 1896, which allows to generate and focus an electron beam, the first electron microscope was developed by Max Knoll and Ernst Ruska from 1929 and later improved to achieve a resolution of 50nm in 1933 (Masters, 2009). This led to the first electron micrograph of an intact cell in 1945 (Porter et al., 1945), the first imaging of viruses (Morgan et al., 1954) and a better understanding the mitochondria: With EM it was finally possible to image their ultrastructure and discover the double membrane character with the cristae invaginations of the inner membrane (Palade, 1952; Sjostrand, 1953).

However, EM comes with constraints with regard to sample preparation, imaging flexibility and target specificity and researcher therefore also investigated other approaches to overcome the classical diffraction limit (Schermelleh et al., 2019). This led to the development of Stimulated Emission Depletion Microscopy (STED, Bianchini et al. (2015)) and Photoactivated Localization Microscopy/ Stochastic Optical Reconstruction Microscopy (PALM/STORM, Henriques et al. (2011)) among others.

Another techniques, which follows a completely different approach, is expansion microscope (ExM) which was introduced by Chen et al. (2015). The idea to increase the size of the sample is deceptively simple but its implementation relies on substantial research into swellable hydrogels (Cohen et al., 1992; Hausen et al., 1981; Tanaka et al., 1980) which allows for an isotropic expansion. One of the main advantages of this approach is that the expanded sample can be imaged with a conventional confocal microscope but still achieving super resolution. This implies that there is no need for specialised, expensive hardware (Gao et al., 2017).

Therefore ExM has the potential to make super-resolution microscopy available to more biology labs and has seen rapid advances since 2015 (Wen et al., 2023).

However as discussed above, biological samples are mostly transparent when viewed with a microscope and this is especially true for the expanded cells because of their dramatically increase volume. While most ExM approaches (Gambarotto et al., 2019; Ku et al., 2016) rely on specific fluorescent labels to tag molecules of interest, the Bewersdorf group at Yale recently discovered that NHS ester acts as a pan-staining for ExM (M'Saad et al., 2020) which allows ExM to produces EM-like images visualising the whole proteome of the cell called panExM. Combing the pan-staining with other dyes or immunlabels allows to visualise specific proteins and structures in their cellular context.

By being compatible with conventional confocal microscopes and high-throughput imaging platforms, panExM enables imaging a substantially larger number of cells and volume at super-resolution (Faulkner et al., 2020). Enabling a large number of cells is crucial in many settings as biological samples often show a large heterogeneity. For example, when investigating the interactions between genetic mutations, morphological changes and disease, it has been observed that even neighbouring cells can show substantially different mutation burdens which can bias the analysis if only a small number of cells is imaged (Broughton et al., 2006).

However, a high throughput poses the challenge of turning a large amount of 3D images into actual biological insight as the annotation of such data is prohibitive labour intensive. Luckily, the last years have seen an revolution in the capabilities of computers to process data and extract information from it.

1.1.3 The Evolution of Machine Learning

In the form of machine learning (ML), computers have been able to reach human-level performance in the classification of images, detection of items within images and segmentation of regions of interest (Dong et al., 2021). Machine learning has its roots in statistics but emerged as its own discipline defined by its greater emphasis of prediction tasks over statistical inference and establishing of relationships

between variables. Moreover, ML has a unique focus on computational and engineering problems.

The great advances in ML over the last years were facilitated by the unique interplay of new algorithms, large datasets and more available computing power.

Most traditional ML methods have primarily been developed for their application to tabular data. This implies that each observation consisted of one variable of interest (the prediction target) and several features associated with it. Examples include the detection of spam emails in which the prediction target is the binary label whether the email should be classified as spam and several features that describe a specific email like its length or the occurrence of specific words (Mansoor et al., 2021).

In order to apply this framework to the classification of images, researchers developed several approaches to extract useful information from the pixel intensity values. However, hand-crafting these features turned out to be laborious and not to generalise well over different datasets and tasks. This approach was challenged with the advent of deep learning (DL) which started showing superior results in 2012 as demonstrated by Krizhevsky et al. (2012). DL approaches relied on neural networks which are stacked perceptrons which have already been developed by Rosenblatt (1958). The fundamental change for vision applications was the introduction of convolutional neural networks (CNN) which are able to take the raw pixel values as input and output a prediction. This implies that while traditional machine learning methods only learned the optimal relationship between the hand-crafted features for the prediction task, DL would also learn the optimal features.

However, this increased flexibility comes at a cost: While in the traditional approach, ML practitioners could use their domain knowledge to create features that extract the most information according to their view, DL is mostly missing this prior information and has to learn all the features directly from the data. This insight helps to explain why advent of DL methods was reliant on the availability of larger datasets and more computing power to extract the information from these datasets.

These larger image datasets were only possible because the internet started to provide researchers with a seemingly unlimited amount of images. ImageNet

(Deng et al., 2009) initially collected around 3 million images and provided a label of what is depicted for each image.

However, processing these datasets of this scale with neural networks becomes infeasible without relying on specialised hardware. Fortunately, the video game industry had relied on graphic processing units (GPUs) which are optimised to compute matrix vector products in parallel on much higher speed than traditional CPUs. Researchers realised that this is the main computational operation required for CNNs and repurposed GPUs to train the networks (Mittal et al., 2019). In this way, DL was optimally suited for the application of GPUs and compared with other ML algorithms won the 'hardware lottery' (Hooker, 2021).

Additionally, large technology companies starting pouring additional resources into the field which allowed the computational power used to train largest models to grow with a factor of 10x per year between 2012 and 2019 (Thompson et al., 2023).

1.1.4 Advancing Microscopy with Machine Learning

Researchers were soon applying deep learning techniques to the microscopy domain for example in the form of general microscopy image analysis (Liu et al., 2021), the segmentation of EM data (Aswath et al., 2023) or histopathology (Banerji et al., 2022).

However, as described above the success of DL-based method is the availability of large datasets and in the microscopy domain, the creation of annotations for the training data is especially very laborious and therefore few large labelled datasets exists. One major focus of this thesis is how the annotation effort can be reduced by using additional immunolabeling to facilitate the training of a segmentation model for the pan staining. Moreover, compared to natural images, microscopy images often feature a higher noise to signal ratio, a higher ambiguity and mostly miss strong colour gradients that correspond to semantic changes which complicates the analysis.

The combination of modern ML algorithms with microscopes enables a fundamental transformation: Traditional microscopes are only concerned with the magnification of a sample and historically just visualising and qualitatively describing

the sample has been sufficient. Modern biology on the other hand has a growing interest in moving beyond simple descriptions to quantitative measurements. However, biology has its own uncertainty principle as every measurement and microscopy analysis means interfering with the biological system as it is removed from its normal environment and has to undergo some imaging preparation. These steps complicate the measurement process and can lead to substantial differences between imaged samples. Even without being able to completely eliminate all concerns, imaging a larger number of samples allows to reduce the influence of cell and batch effects. However, manually analysing this larger number of cells introduces a substantial bottleneck.

ML allows to overcome this obstacle and transform the nature of the analysis. Instead of simply returning visual representations of the sample, it is possible to build a complete analysis pipeline which directly reports a quantitative description of the sample to the researcher. This allows to scale up the analysis and automate the analysis of a large number of cells.

While this philosophy has been pursued in the context of normal confocal microscopy, panExM opens the door to extend it to super resolution. For researchers interested in the ultrastructure of the cell at a nano-meter resolution, the combination of panExM and machine learning therefore offers exciting new opportunities. PanExM allows to image super-resolution images with a high throughput and the possibility of applying additional immunolabeling offers a chance to alleviate the need for manual annotations. Machine learning can build on this to turn the large amount of data into actionable biological insight. This thesis therefore develops an analysis pipeline for the large scale investigation of mitochondria and their ultrastructure.

1.1.5 Understanding Mitochondria

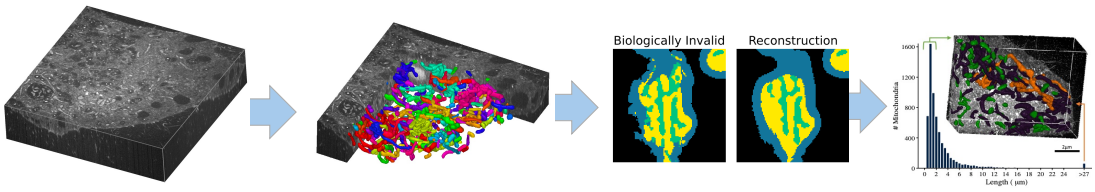
Mitochondria provide cells with energy in form of ATP and occupy up to 25 percent of the volume of a cell (Lodish et al., 2021). By breaking the linkage between cell size and energy demand, they have been hypothesised to be the enabler for all complex life on earth in form of eukaryotes (Lane et al., 2010). On top of that,

they participate in regulating the programmed cell death (Mohamad et al., 2005) and in the signalling mechanism of the cell (Tait et al., 2012).

Because of these crucial roles played by mitochondria, any malfunctioning potentially results in severe diseases (Picard et al., 2016). Changes in the mitochondria have been associated with myopathies (Holt et al., 1988), Parkinson's disease (Franco-Iborra et al., 2016), progressive-blindness disease (Yu-Wai-Man et al., 2011) or type 2 diabetes (Szendroedi et al., 2012).

The root of many of these diseases are suspected to lie in mutations of the mitochondria DNA which is more susceptible to mutations (Johns, 1995) because it is not protected by the nucleus. Mutations in the DNA also influence the shape of the mitochondria and their ultrastructure which have been associated with different disease types (Jenkins et al., 2024). For example, donut shaped mitochondria have been found to be more prevalent in liver cell under metabolic stress (Lionetti et al., 2014) and the cristae ultrastructure of the mitochondria has been observed to respond to abnormal conditions by changing its shapes (Cogliati et al., 2016; Vincent et al., 2016).

Nevertheless, our understanding of the role of mitochondria and the connection between function and shape is still incomplete, partly because genetic mutations, morphological changes and the reorganisation of the ultrastructure represent highly volatile events that show a large heterogeneity even between neighbouring organelles. Therefore, studying them requires large samples sizes to differentiate between signal and noise. Moreover, Jenkins et al. (2024) argue that it is crucial to consider the mitochondria morphology in the context of the entire cell. By combining the potential to facilitate large scale studies with the resolution to investigate the mitochondria ultrastructure while still visualising the whole cell, panExM therefore provides an excellent opportunity to study the morphology of mitochondria.



Uncertainty Analysis (Chapter 3) | Segmentation (Chapter 4) | Post-Processing (Chapter 5) | Analysis (Chapter 6)

1.2 Aims and Contributions

This thesis works towards the vision of enabling microscopes to directly extract semantic information from a sample with respect to the analysis of the mitochondrial ultrastructure with panExM. Moreover, the aim is to efficiently use all available information in the panExM data like additional immunolabeling in order to minimise the annotation effort required to adopt the method to new settings.

The presented work builds on a collaboration with the Bewersdorf group at Yale University in which all imaging was conducted by Yuan Tian at Yale University and all model development, image analysis, visualisations and other computational aspects were conducted by me at Oxford.

- **Chapter 2** provides background information on mitochondria and previous applications of machine learning to microscopy data with a special focus on the application to electron microscopy images and surveys the current state-of-the-art.

- Since this is the first work to apply machine learning techniques to panExM data, I analysis in **Chapter 3** the panExM data with respect to the annotation uncertainty and find that on a pixel level there is substantial ambiguity in the labelling of the mitochondrial ultrastructure. Motivated by this insight, I develop a new algorithm to exploit inter-annotator disagreement for semi-supervised learning which allows a more data-efficient training. Moreover, I investigate the popular Dice-loss function with regard to its properties in a noisy signal setting and explain its favourable performance.

- **Chapter 4** demonstrates how an additional dye like MitoTracker allows to alleviate the annotation requirement. Specifically, I demonstrate how an initial model can be trained which is capable of predicting the outline of the segmentation masks for the pan-staining. This model is then fine-tuned on human annotations to also predict the ultrastructure of the mitochondria. Moreover, I develop a model that improves the segmentation results for drug-treated HeLa cells and demonstrate how the model can be applied to mouse kidney cells in which the additional immunolabeling is affected by significantly higher noise levels than the MitoTracker. This allows to recover the mitochondria outline based purely on very spare immunolabeling. Subsequently, I provide guidance what quality the additional labels needs to have if the method is supposed to be applied to new applications.

- The resulting segmentation masks are further refined in **Chapter 5**. Microscopy images are often affected by local variations in the imaging quality due to uneven staining, the imaging process or the sample itself. These ambiguous areas pose a challenge for segmentation algorithms and can lead to segmentations that violate known biological constraints of the organelles. I propose a new reconstruction algorithm that is purely data driven to detect and correct implausible segmentation masks and I demonstrate its superior performance compared to competing approaches. Moreover, I describe the computation of several metrics derived from the segmentation masks and used in the subsequent analysis.

- Finally, based on the previous segmentation mask and computed metrics **Chapter 6** demonstrates how this pipeline can be used to investigate biological questions analysing over 65,000 mitochondria in total. I describe the mitochondria in untreated HeLa cells in terms of their global morphology and ultrastructure and demonstrate how my pipeline is always able to tie back every data entry to a concrete mitochondrion. Subsequently, I finalise the analysis by comparing the quantitative change in the mitochondria due

to the drug treatments and describe the baseline distribution we observe in the control group. For the mitochondria in mouse kidney tissue, I analyse the difference between healthy cells and tissue suffering from acute kidney injury which reveals drastic changes in the cristae ultrastructure which are not apparent in terms of lower-resolution macro parameters. Finally, I analyse the 3D localisation of three immunolabels (Tom20, Mic60 and COX IV) with respect to the segmentation of the mitochondria. In both steps, I am able to support vague visual impressions with a concrete quantitative measurements.

- **Chapter 7** summarises this thesis and gives an outlook where I expect this work to impact future research. Examples include the application to actual patient samples, the investigation of new organelles beyond mitochondria or the combination of my algorithms with new microscopy techniques to further increase the throughput.

1.3 Publications

This thesis has resulted in the following publications:

- **Sauer, A.**, Dmitrieva, M., Han, H. and Rittscher, J., 2023. Leveraging Inter-Annotator Disagreement for Semi-Supervised Segmentation. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5)

- **Sauer, A.**, Tian, Y., Bewersdorf, J. and Rittscher, J., 2024. Refining Biologically Inconsistent Segmentation Masks with Masked Autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6904-6912).

Additionally, we are currently in the process of publishing the combined results of Chapter 4 and 6 in a a journal specializing in biology. More indirectly, I was also involved in the following work:

- Han, H., Dmitrieva, M., **Sauer, A.**, Tam, K.H. and Rittscher, J., 2022. Self-supervised voxel-level representation rediscovers subcellular structures in volume electron microscopy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1874-1883).

- Wood, R., Sirinukunwattana, K., Domingo, E., **Sauer, A.**, Lafarge, M.W., Koelzer, V.H., Maughan, T.S. and Rittscher, J., 2022. Enhancing local context of histology features in vision transformers. In *MICCAI Workshop on Medical Image Assisted Blomarkers' Discovery* (pp. 154-163).

2
Background

Contents

2.1 Biological Context and Imaging Technique 13
2.1.1 Mitochondria and their Role in the Cell 13
2.1.2 Expansion Microscopy 15
2.2 Machine Learning and Image Segmentation 18
2.2.1 Overview 18
2.2.2 Deep Learning Based Approaches 19
2.2.3 Semantic Segmentation 22
2.2.4 Pre-training and Semi-supervised Approaches 23
2.3 Automated Analysis of Mitochondria 26
2.3.1 Analysis of Mitochondria in Light Microscopy 26
2.3.2 Analysis of Mitochondria in Electron Microscopy 28
2.3.3 Analysis of Mitochondrial Ultrastructure 31

2.1 Biological Context and Imaging Technique

2.1.1 Mitochondria and their Role in the Cell

Mitochondria were famously described as the 'powerhouse' of the cell (Siekevitz, 1957) as they provide energy by generating ATP through aerobic respiration. Mitochondria are contained in all eukaryotic cells and while their size and shape varies substantially and is highly dynamic, it is in general similar to bacteria with a length between 3-4µm (Scheffler, 2008). The similarity to bacteria goes beyond the shape and is no coincident as mitochondria are suspected to have been independent

organelles in the past that at one point merged with another cell even though the details of this origin story are still debated (Roger et al., 2017).

As illustrated in Figure 2.1, mitochondria consist of a double membrane structure. As discussed by Lodish et al. (2021), the outer membrane separates the mitochondrion from the surrounding cytoplasm of the cell. Within the mitochondrion and surrounded by the *intermembrane space*, a second continuous membrane is found with three distinct domains. Firstly, the part of this inner membrane adjacent to the outer membrane is called *boundary membrane*. Secondly, the inner membrane forms numerous invaginations inside the mitochondrion which are called *cristae*. These invaginations increase the surface area of the inner membrane by up to five times of the length of the outer membrane which is crucial for the production of ATP. Therefore, mitochondria with a greater demand for ATP are often found to have many times more cristae than others (Huang et al., 2023). Finally, the connections between the inner boundary membrane and the cristae are called a *cristae junction*.

The volume within the inner membrane is called *matrix* and the inner structure and its variations are discussed in more detail in Section 2.3.3.

As previously mentioned, eukaryotic cells are thought to be the symbiosis of ancestral cells with eubacteria that have been engulfed at some point in time (Lane, 2018). While most bacterial genes have been lost over time, there is still some DNA present in the mitochondria matrix that is distinct from the DNA found in the nucleus but necessary for a health mitochondrion (Johns, 1995).

The shape of the mitochondria is highly dynamic as they can merge (*fusions*) and separate (*fissions*) and form extensive networks in response to the metabolic state of the cell. Lodish et al. (2021) describe that stressed cells tend to have more fragmented mitochondria whereas mitochondria with a higher generation of ATP tend to form extensive networks of fused mitochondria as seen in Figure 2.2. Fission and fusion affect both the outer and inner membranes of the mitochondrion and are important when the cell divides to dispose of dysfunctioning mitochondria and to keep a functioning organisation throughout the cell. In this way, damaged

mitochondria can be separated from the main mitochondria network, surrounded by a membrane and degraded by lysosomes in a process called mitophagy.

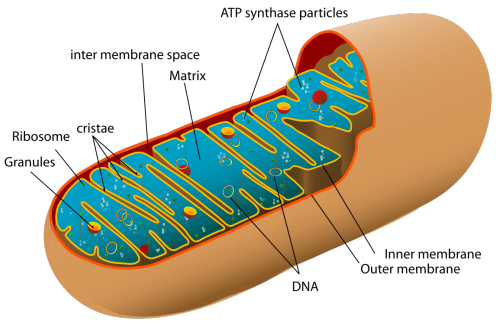


Figure 2.1: Schematic illustration of a mitochondrion (Villarreal, 2006).

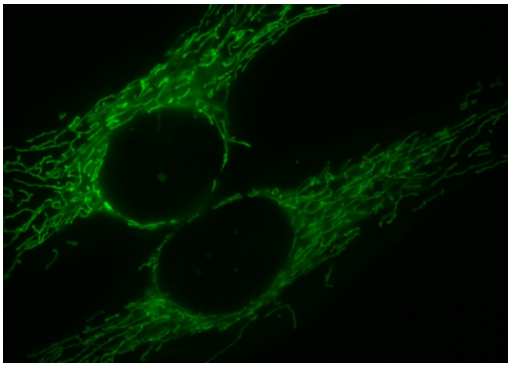


Figure 2.2: Mitochondria imaged with a fluorescent microscope (Troeder, 2014)

Beside its function to produce ATP, mitochondria play an important role in the sexual reproduction as in almost all multicellular eukaryotes only the female passes on its mitochondria to the offsprings which has important consequences for the transmission of genetic diseases (Perlman et al., 2015).

Additionally, mitochondria are capable of initiating apoptosis, which describes the process of cell death, and in this way contribute to removal of damaged cells (Lopez et al., 2015). Finally, mitochondria play a crucial role in the communication within the cell (McBride et al., 2006).

Because of all these important roles played by the mitochondria, any malfunctioning can result in severe diseases and researchers start to uncover the importance of mitochondria more and more (Lane, 2006). Several findings indicate that mitochondria are involved in Parkinson's disease (Franco-Iborra et al., 2016), progressive-blindness disease (Yu-Wai-Man et al., 2011) or type 2 diabetes (Szendroedi et al., 2012).

2.1.2 Expansion Microscopy

In order to circumvent the limitations of traditional light microscopy and the FIB-SEM's requirements for specialised equipment, several alternatives have been developed (Nieves et al., 2021) with one of the first examples being simulated emission depletion (STED, Hell et al. (1994)), structured illumination microscopy

(SIM, Gustafsson (2000)), stochastic optical reconstruction microscopy (STORM, Betzig et al. (2006)) and photoactivated localization microscopy (PALM, Rust et al. (2006)) which can achieve a resolution down a few nanometres.

In 2015, an additional new method was introduced called expansion microscopy (Chen et al., 2015) which builds on the idea of physically magnifying the sample through a chemical process. Following Gao et al. (2017), the process to make sure that the expansion is isotropic includes the following steps: Firstly, the biomolecules of the sample are chemically equipped with anchors that allow them to bound to a web of a swellable hydrogel or polymer which is synthesized throughout the specimen. This polymer is densely cross-linked to ensure that the biomolecules retain their spatial organisation relative to each other once the polymer threads swell apart. In order for them to do so, the polymer-embedded sample is mechanically homogenized to allow expansion and prevent fracturing and then added to a solvent which induces the swelling. By comparing their method with images from classical super-resolution microscopy, the distortion induced by the expansion is found to be only between 1 and 4 % (Gao et al., 2017).

While this first expansion increases the specimen size by a factor of 4-5, it was soon discovered that it is possible to conduct the expansion iteratively to increase the magnification factor up to 20 and achieve an effective resolution of 25nm (Chang et al., 2017). This improvement is part of several efforts to develop new protocols that improve anchoring strategies, spatial resolution, and fluorescence intensity (Wen et al., 2023).

In 2020, the Bewersdorf Lab developed the panExM protocol (M'Saad et al., 2020). For this method, the sample is expanded twice before it is stained with NHS ester dye conjugate and a final expansion is performed. This allows a resolution down to 20nm. Staining the cell with NHS ester dyes allows to visualise the full proteome within the sample as seen in figures shown in Chapter 3. Moreover, given its high resolution it can reveal nanoscopic structural hallmarks which allow researchers to identify organelles without the need for specific staining.

Nevertheless, conventional post-expansion antibody labelling or special dyes can be applied as well to reveal specific structures. One of the benefits of expansion microscopy is the decrowding of biomolecules through the expansion process from an average distance of 3nm between proteins to about 50nm after expansion (Gao et al., 2017). As the smallest fluorescent labels are comparable in size to the distance between proteins, this causes many binding sites to be masked or unreachable which fundamentally limits the resolution of traditional fluorescence microscopy. However, after the expansion phase this is no longer true and previously inaccessible binding sites are revealed.

Another crucial advantage of expansion microscopy is that even though it is able to provide resolution at the nanoscale, it only requires diffraction-limited microscopes hardware, incubators, freezers and other items which are commonly available in biology laboratories (Gao et al., 2017). This stands in contrast to other super-resolution methods or correlative light/electron microscopy (CLEM) which requires specialised and expensive hardware. Additionally, the sample preparation is relatively easy and the final imaging several times faster compared to other super-resolution methods (Gao et al., 2017) which allows for larger volumes to be imaged. However, one disadvantage to be noted is that the separation of proteins and fragmentation of lipid membranes means that expansion microscopy cannot be used on living specimens.

Because of the novelty of expansion microscopy, its potential in combination with machine learning methods remains mostly unexplored. In 2023, one workshop at the ISBI conference workshop *Robust Non-rigid Registration Challenge for Expansion Microscopy - Grand Challenge* (2024) organised a challenge to develop algorithms for robust non-rigid registration of ExM images. Moreover, Gatti et al. (2019) and Chang et al. (2024) develop algorithms to correct the anisotropic resolution. However, to the best of our knowledge, there is no work yet that explores segmentation algorithms for ExM or the automatic analysis of the images.

2.2 Machine Learning and Image Segmentation

2.2.1 Overview

As soon as it was feasible to store images digitally, computer scientists investigated ways to extract relevant information from these images. However, early algorithms were constrained by the limited available processing power and memory. In the case of semantic segmentation, simple thresholding based on the pixel intensity was often used to divide the image into fore- and background, possibly by adopting the threshold automatically to minimise the intra-class intensity variance (Otsu, 1979).

One of the early approaches that stood the test of time are linear filters. An image $I \in \mathbb{R}^{w \times h}$ is represented by a grid of w times h pixels where the intensity of the pixel in row i and column j is given by $I_{i,j}$. For a given response function or filter W specified by its weights $w \in \mathbb{R}^{k \times l}$, the transformed images is given by the convolution

$$[W * I]_{i,j} = \sum_{u=1}^{k} \sum_{v=1}^{l} w_{u,v} I_{i-u+\lfloor k/2 \rfloor, j-v+\lfloor l/2 \rfloor}$$

Applications of filters can either be used as pre- or post-processing step like the removal of noise if W is chosen as a Gaussian filter or as a first step to extract the presence of certain features like estimating the gradients in an image using Sobel filters (Sobel et al., 1968). This approach allows to combine several filters in a pipeline and achieve more complex tasks like the detection of edges in images (Canny, 1986).

In order to segment an image into its semantic components, one common approach was to use various filters to generate a diverse set of features and then train a traditional machine learning algorithm to assign each pixel a corresponding label like in the ilastik software (Berg et al., 2019). The advantages of this approach are that a small amount of annotations is often sufficient and the limited requirements in terms of memory and computing power imply short training times that enable fast and iterative corrections to the model.

The idea behind machine learning is to learn a relationship between the features $X \in \mathbb{R}^{n \times k}$ and an outcome $Y \in \mathbb{R}^{n}$ (like the pixel label in a segmentation task)

from noisy data where n is the number of observations and k the feature dimension. In the simplest case of linear regression $\hat{Y} = X\hat{\beta}$, this relationship is assumed to be well approximated by a linear function and only the coefficients $\hat{\beta} \in \mathbb{R}^k$ have to be estimated in order to minimise the difference between observed values and predictions $\|Y - X\hat{\beta}\|$.

However, several other algorithms have been developed that allow to learn much more flexible relationships $\hat{Y} = g(X)$ where g is some expressive, non-linear function. Popular approaches include tree based approaches like random forests (Breiman, 2001), boosted trees (Chen et al., 2016; Ke et al., 2017) or support vector machines with various kernels (Murphy, 2022). The ilastik software mentioned above for example in its initial version is based on a random forest (Berg et al., 2019).

As discussed later in Section 2.3.1, these approaches can be sufficient to handle simple cases like the mitochondria segmentation in fluorescent microscopy images in which the mitochondria represent a coloured foreground that needs to be separated from a mostly dark background.

2.2.2 Deep Learning Based Approaches

The task for EM and ExM images, however, is more complicated because the segmentation of the mitochondria is more involved due to the visibility of other organelles within the cell. Therefore, purely relying on the pixel intensities is not sufficient and the model rather has to learn the characteristic shape and texture of the segmentation target. Because the crucial part of the traditional machine learning pipeline is the feature extraction, designing the right filters to extract this information is determining the model performance. This approach is very labour intensive as finding the right features often involves several hyperparameters that need to be chosen by the researchers and it is not robust to changes in the data distribution. Therefore, this approach does not allow to scale.

Convolutional neural networks (CNNs) and the framework of deep learning (DL) promise to solve this problem by unifying the learning of the feature extraction and classification model. DL builds on the idea of stacking together several simple

linear models as proposed by Rosenblatt (1958) and combine them with a non-linear activation function as shown in Figure 2.3. The research on CNNs was pioneered by LeCun et al. (1995) and later accelerated by the availability of new computer hardware that substantially increase the available computational resources and was very well suited for the requirement of DL models (Hooker, 2021).

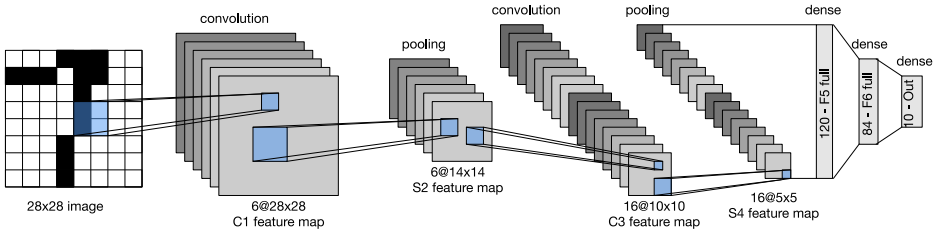


Figure 2.3: Illustration of the CNN proposed by LeCun et al. (1995) showing the cachining of different convolutional and pooling layers. Figure adopted from Zhang et al. (2023).

The parallel processing within the CNNs allowed to repurpose graphic processing units (GPUs) to dramatically decrease the training time and exploit a long history of hardware optimisations for the GPUs for the gaming industry. Combined with the availability of large benchmark datasets like ImageNet (Deng et al., 2009), CNNs were in a unique position to improve the performance of computer vision models (Krizhevsky et al., 2012).

CNNs originally targeted image classification and work by combining a larger number of filter operations both in series and parallel which are automatically optimised for the task at hand.

More specifically, each CNN consists of a sequence of m building blocks b which are applied sequentially to the raw input image X as $b_m \circ \cdots \circ b_1(X)$. In its basic form, each block $b(x) = h(c(x))$ consists of a convolution c and non-linearity activation function h. One of the most popular activation function is the simple function ReLu(x) $= \max(0, x)$ which can be interpreted as shutting down information in specific channels (Sandler et al., 2018). The convolution layer cross-correlates a

kernel W with the input of k channels into new feature tensor with l channels. By moving the kernel with a stride larger than 1, the spatial output of the input is reduced and a general guideline is that a decrease in the spatial resolution should be accompanied by an increase the number of channels l. This basic form of the blocks b is often augmented with features like dropout (Hinton et al., 2012) to increase the robustness of the model or some form of normalisation (Ioffe et al., 2015; Wu et al., 2018) to stabilise the training. In the classification setting, the spatial arrangement of the final output is then completely disregarded and passed through another linear layer before the outputs are transformed into probability predictions for the potential classes using a softmax function.

The crucial point is that the whole transformation from the raw input image to the predicted class probabilities is differentiable with respect to each parameter of all convolutional filters. This implies that using a suitable loss function like the cross-entropy between the target values and the predictions, it is possible to find the corresponding gradients for each kernel entry and iteratively update the weights via gradient descent to match the target predictions as closely as possible.

In order to increase the performance, the most successful direction has been to increase the size of the networks. Notable examples are ResNets (He et al., 2015) which allowed much deeper model through the introduction of residual connections and EfficientNet (Tan et al., 2019) which investigates and optimizes for different ways to scale a model in terms of depth, width and resolution. In the domain of language processing, the transformer architecture (Vaswani et al., 2017) showed impressive results and was adapted to the CV domain in the form of vision transformers (ViT) (Dosovitskiy et al., 2021) which allowed for even bigger models. For example, the largest ViT model presented by Dosovitskiy et al. (2021) included 632 million parameters compared to 1.7 million parameters of the best performing ResNet presented by He et al. (2015). Motivated by the success of these large ViTs, CNNs have been scale up to similar sizes and reported to be able to achieve a competitive performance (Smith et al., 2023). This suggests that the architectural decision

between CNNs and ViT is less important than having access to large amounts of data and enough computing power available to scale the model accordingly.

2.2.3 Semantic Segmentation

In the field of biomedical imaging, researchers are often interested in identifying and quantifying structures of interests which therefore need to be segmented. As described above, normal CNNs decrease the spatial resolution to encode the information which allows for a global representation of the input. Therefore, fully convolutional networks (Shelhamer et al., 2017) and UNets (Ronneberger et al., 2015) were developed which add an additional decoder that uses transposed convolutions to recover the original shape and allows predictions on the pixel level. Moreover, by replacing two dimensional with three dimensional kernels, the UNet architecture could easily be adopted for 3D data (Çiçek et al., 2016).

UNets have been widely adopted (Siddique et al., 2021a) and applied to various tasks like brain tumour diagnosis in MR images (Li et al., 2019), liver cancer in CT scans (Li et al., 2018), prenatal diagnosis with ultrasound (Włodarczyk et al., 2019) or the segmentation of mitochondria as described in Section 2.3. Moreover, an abundance of modifications have been proposed (Krithika et al., 2022) and claimed to increase the performance over the original architecture including attention UNets (Schlemper et al., 2019) that incorporate additional attention gates to focus on specific features or inception UNets (Zhang et al., 2020b) which factorises larger filters into several smaller ones.

One challenge is to make these powerful segmentation models accessible to biologists with little experience in ML models. Isensee et al. (2021) propose a framework built on top of UNets which aim to make the training and preprocessing more accessible after finding that the standard UNet architecture often provides competitive results (Isensee et al., 2019)

Other approaches like Cellpose (Stringer et al., 2021) or Stardist (Weigert et al., 2020) aim at instance segmentation instead of semantic segmentation and augment the mask prediction of the UNets for example by also predicting the distance map

or its gradients in order to recover individual instances. This follows the idea of inducing shape priors for semantic segmentation and has a long history starting with active contour models proposed by Kass et al. (1988) pre-dating any deep learning and was based on the idea of minimising a pre-defined energy function by following the corresponding differential equation. Optimisation approaches suffered long run-times however, and several attempt have been made to combine them with DL (Cheng et al., 2019; Gur et al., 2019; Marcos et al., 2018). Peng et al. (2020), for example, consider the bounding boxes provided by some object detection algorithm as graphs and transform them to fit the shape of the object of interest. Building on a similar idea, Wickramasinghe et al. (2020) propose an algorithm which outputs 3D meshes directly by transforming an initial input mesh using a mesh decoder that operates on the features learned from a voxel encoder. However, these approaches suffer from the fact that they only expect a small number of objects in each stack and either take an initial outline as given or assume only one object within a bounding box which is not a valid assumption for shapes highly elongated like mitochondria.

Other approaches try to capture the shape information by predicting additionally the signed distance maps (Wang et al., 2021a; Xue et al., 2020) or propose to impose a shape prior in the form of post-processing (Larrazabal et al., 2019).

Additionally, the transformer architecture has been adopted for segmentation tasks (Cao et al., 2023; Hatamizadeh et al., 2022a). However, Isensee et al. (2024) find that CNN-based UNet architectures still provide state-of-the-art performance compared to Transformer-based approaches and many papers claiming superiority of their proposed approach suffer from validation shortcomings like inadequate baselines, insufficient datasets and neglected differences in the employed computational resources.

2.2.4 Pre-training and Semi-supervised Approaches

One general theme that emerged was that sufficiently large networks can act as universal function approximators (Chen et al., 1995) regardless of the details of the network architecture. Tolstikhin et al. (2021) show for example that by sufficiently

scaling up and combining simple MLPs, it is possible to achieve competitive performance for classification task without the attention mechanism defining ViTs or explicit convolutional layers from CNNs.

Therefore, the big increases in performances have been achieved by being able to include larger amounts of data in the training process and using the available annotated data efficiently. Especially segmentation tasks typically suffer from a greater scarcity of labelled samples compared to classification problems due to the more time-consuming annotation process. This burden is even larger in 3D and in the biomedical domain due to the requirement for expert annotations.

Therefore, researchers explored ideas which allow to use completely unlabelled data in the form of self-supervised learning or combine a small amount of annotated data with a larger amount of unlabelled data from the same data source in the form of semi-supervised learning.

Pre-training Approaches

Across different domains of ML (Caron et al., 2021; He et al., 2022; Radford et al., 2018), it was shown that in order to find the intermediate representations that are crucial to the success of DL, self-supervised pre-training allows to substantially reduce the amount of required annotated data: Using only unlabelled data, a model is trained on some related proxy-task to learn a rough feature distribution and corresponding internal representations and it could be shown that once these representations are established, fine-tuning the model to the task at hand requires significantly less annotated data. Prominent proxy-tasks include filling gaps in some masked input (He et al., 2022), re-recognising an image instance after applying some form of data augmentation (like rotations, colour distortion, added noise) in the form of contrastive learning (Chen et al., 2020; He et al., 2020) or keeping the latent representations after these augmentations consistent (Caron et al., 2021; Grill et al., 2020). In this way, it has been possible to utilise large parts of the internet to pre-train language models on hundreds of billions of words (Brown et al., 2020) and image classification models on billion of images (Dehghani et al., 2023).

The pre-training approach, described above, of taking a large dataset (like ImageNet) to facilitate the feature learning of the deeper layers and then fine-tune it on the application at hand however faces several challenges with respect to semantic segmentation. Most fundamentally, He et al. (2019) find that pretraining on classification tasks does not provide the same benefits when later adopted to segmentation problems. Moreover, most of the large datasets depict natural images and the model therefore has to cover a large domain shift for medical images (Ke et al., 2021). Finally, this work is interested in segmenting 3D data for which less annotated data for pre-training is available.

Semi-supervised learning

In the field of biomedical imaging, it is often the case that while there is a large data dataset available for a given task, only a small subset is annotated. Approaches which simultaneously allow the model to learn from a small amount of labelled data and typically a larger amount of unlabelled data from the same source are called semi-supervised learning. Semi-supervised methods for deep learning are dominated by two approaches: consistency regularisation and entropy minimisation. Consistency regularisation methods aim to use the unlabelled data to accelerate feature learning by training the model to predict the same output for the input under small perturbations. Prominent examples include the Π-model (Sajjadi et al., 2016) and its extensions like temporal ensembling (Laine et al., 2016) or mean-teacher models (Tarvainen et al., 2017).

The alternative approach to semi-supervised learning is entropy minimisation (Grandvalet et al., 2004) and pseudo labelling (Lee, 2013). The intuition behind these approaches is that the decision boundary of a classifier should lie in a region of low data density. While entropy minimisation encourages the model to make confident predictions regardless of the class, the pseudo labelling approach targets the model's own predictions if its confidence is higher than some threshold. However, even confident predictions might be wrong which can mislead the model and consequently degrade its performance. This prompted researchers to investigate other metrics

besides the confidence to judge the reliability of the model predictions and it was shown that incorporating the uncertainty of the predictions into the selection of pseudo labels (Cao et al., 2021; Rizve et al., 2021) allows to improve the performance to the same level as consistency regularisation approaches. Moreover, consistency regularisation and entropy minimisation have been combined (Wang et al., 2021b) and uncertainty masks have also been introduced to the consistency regularisation framework (Zheng et al., 2022). While a lot of the methods were originally developed for classification tasks, they have also been successfully adopted to segmentation problems (Li et al., 2020; Peiris et al., 2021; Wang et al., 2022).

2.3 Automated Analysis of Mitochondria

2.3.1 Analysis of Mitochondria in Light Microscopy

As Figure 2.2 showed, mitochondria are visible with conventional microscopes which have a resolution limit around 250nm (Galbraith et al., 2011). This implies that even though the inner structure of the mitochondria cannot be resolved, it is possible to observe the structure of the mitochondria on a macro level and several authors worked towards extracting corresponding quantitative parameters automatically. Zahedi et al. (2018) provide open-source software to researches for analysing fluorescent images stained with mitochondrial-targeted dye. Based on features like area, radius and perimeter they then use K-nearest neighbor and Naive Bayes algorithms to classify the present mitochondria as 'Punctated', 'Networked' or 'Swollen' and use this to predict the overall health of the cell.

Rohani et al. (2020) aim to extract information about mitochondria from high throughput 2D multi-cell fluorescence images. They segment the mitochondria using some elaborate thresholding strategy after applying various denoising strategies. They subsequently deploy their algorithm to analyse 630 cells from three different cell lines and build a database of morphological features.

A similar strategy of combining denoising and thresholding is used by Lefebvre et al. (2021) to provide automated segmentation of mitochondria in live-cell time-lapse images.

However, deep learning methods have also been used, for example by Fischer et al. (2020) who train a UNet (Ronneberger et al., 2015) on twelve 2D maximum-intensity projection images of Caenorhabditis elegans worms and find superior performance to tools like Ilastik (Kreshuk et al., 2019).

These automated analysis tools have also already been used to investigate medical questions. Frye et al. (2021) for example compare the mitochondria morphology from patients with autism spectrum disorder. They analyse several thousand 2D cells using proprietary software to quantify the mitochondria in terms of their morphology (including metrics like area, length, roundness) and network characteristics (including metrics like branching points and ending points).

In summary, light microscopy enables researchers to analyse large numbers of cells even though they often limit themselves to 2D. Moreover, due to the nature of the fluorescent labelling, the segmentation of the mitochondria can be achieved with relatively simple methods like a combination of denoising and thresholding methods. One common theme between these studies is that they focus on parameters like mitochondria area, aspect ratios, roundness and the network structure to describe the overall shape. This representation of shape is attractive because it is easily to communicate and interpret, however the selected representation of shape influences the subsequent analysis results. Picard et al. (2013), for example, investigates the morphological differences between mitochondria from muscle cells of mice which were either assigned to a 'running group' or 'sedentary group'. After a manual segmentation, the shape was captured by the parameters mentioned above. When Zhang et al. (2022) however reanalysed the data using more sophisticated shape representations like landmarks and parameterised curves, they find a significant difference between both groups whereas Picard et al. (2013) concluded that there was none. Moreover, due to the limiting resolution of light microscopy no description and analysis of the ultrastructure is possible.

2.3.2 Analysis of Mitochondria in Electron Microscopy

In order to capture the 3D shape of mitochondria and their ultrastructure in more detail, researchers traditionally had to rely on electron microscopy (EM). Technological advances increased the volume that could be scanned and researchers quickly realised that annotating and extracting the relevant biological information by hand is extremely time and labour intensive. Vincent et al. (2019) for example provide quantitative data for the difference between human and mouse mitochondria but rely on manual segmentation. To allow the analysis of a larger number of mitochondria, computer scientists started developing methods based on early machine learning approaches to segment the mitochondria within the scanned samples.

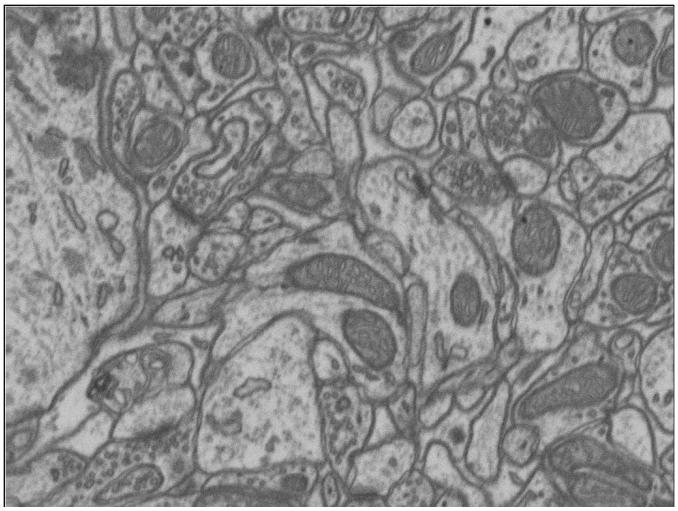


Figure 2.4: Example slice of the EPFL hippocampus dataset introduced by Lucchi et al. (2012)

.

As pointed out by Lucchi et al. (2012), it is not possible to simply run algorithms which excelled at natural images on the EM data and expect similar results for several reasons:

- Algorithms developed for natural images mostly rely on local statistics and focus less on important shape cues which are crucial to microscopy images that often feature a lot of noise and different textures.

- Moreover, the conventional assumption that strong image gradients correspond to boundaries often does not hold for microscopy data.

- And finally, especially pre-deep learning computer vision algorithms often would not be able to handle the amount of voxels generated by EM which commonly exceeds 1 billion voxels per sample.

In the beginning, researchers focused on traditional machine learning techniques. Early examples like Vitaladevuni et al. (2008) or Narasimha et al. (2009) use local texture features and compare a set of different classifiers like nearest neighbors, support vector machines (SVM) and AdaBoost (Freund et al., 1997). Moreover, they analyse the resulting segmentation in terms of several biological parameters like volume occupied by mitochondria, surface area and mean mitochondrial width. Li et al. (2016) first train a classifier following the AdaBoost algorithm for a coarse segmentation which they then fine-tune with a variational image segmentation approach. Several of these studies attempt to include strong shape priors. For example Mumcuoglu et al. (2012) try to exploit the elliptical shape of the mitochondria by employing an ellipse detection algorithm. Similar to Jorstad et al. (2014), they heavily try to take advantage of the double-membrane enclosing the mitochondria which are very distinctive in the EM images as can be see in Figure 2.4.

One major study was conducted by Lucchi et al. (2012) who introduced the EPFL hippocampus dataset which would serve as a benchmark for new methods. This dataset was acquired using a focused ion beam scanning electron microscope (FIB-SEM). An example slice is shown in Figure 2.4. Lucchi et al. (2012) base their approach on superpixels and make use of ray descriptors to capture the shape of mitochondria. They then use a graph cut approach in which they employ an SVM to predict the energy terms involved.

However, after deep learning approaches like the U-Net (Ronneberger et al., 2015) were shown to be able to achieve impressive results on medical images, researchers increasingly moved away from traditional machine learning approaches with early examples including Cheng et al. (2018), who find that a 3D CNN with factorized

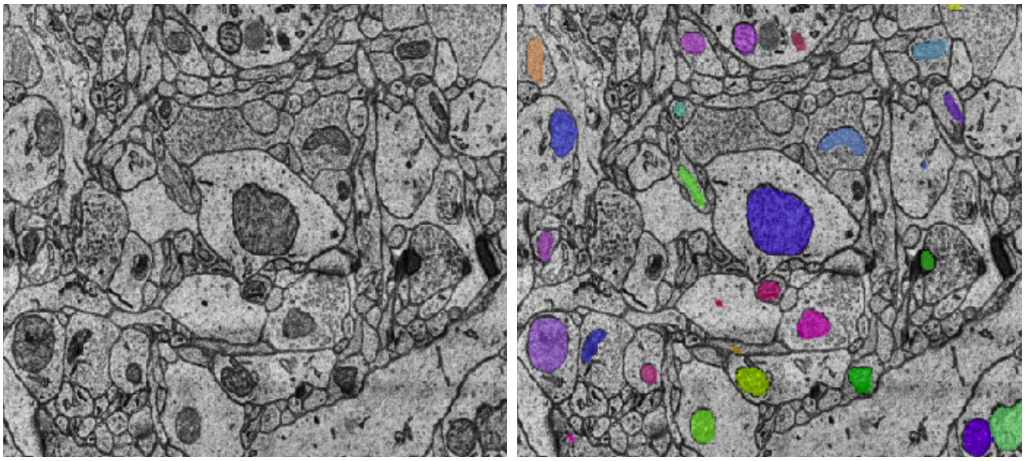


Figure 2.5: Example from the MitoEM dataset (Wei et al., 2020b) with the raw data on the left and the annotated data on the right.

convolutions outperform a 2D UNet despite having fewer parameters, or Oztel et al. (2017) who work with a 2D model but then apply median filtering in the z-dimension to account for the 3D nature of the data. However, initially these approaches cannot outperform traditional methods like a graphical model with a structured SVM as presented in Lucchi et al. (2015).

It was later observed that there are inconsistencies with respect to the boundaries and several false classifications in the ground truth of the EPFL Hippocampus data (Cheng et al., 2018) which lead to the introduction of a re-annotated version by Casser et al. (2020).

Additionally, Wei et al. (2020b) introduce a much larger dataset called MitoEM which shows $30\mu m^3$ volume samples from human and rat cortices and covers around 40K mitochondria with an example shown in Figure 2.5. Given its much larger size, it is able to show a much greater variety of the mitochondrial shapes which makes the segmentation problem much harder. However, up to this point only few studies picked up on that dataset like Nguyen et al. (2021) who integrate a long-short term memory layer as initially developed by Hochreiter et al. (1997) for modelling of sequential dependencies to incorporate the 3D information and predict the boundaries separately for instance segmentation. A similar approach of separating boundaries from the overall segmentation is followed by Nightingale et al. (2021).

It was observed by Franco-Barranco et al. (2021) that many papers described so far did not publish their code which was used to generate their results. Franco-Barranco et al. (2021) therefore reimplemented several approaches to make them publicly available and compared them against several state-of-the-art methods and different post-processing steps. Surprisingly, for the EPFL Hippocampus dataset, they obtained the best results with a 2D UNet without batch normalisation, ELU activation functions and attention gates for the decoder. Moreover, they argue that the superiority of 3D models crucially depends on the consistency among adjacent slices. Similarly, in a benchmark study Khadangi et al. (2021) still find a UNet to be among the top performing models besides their own architecture.

Recently, researchers have begun applying general deep learning techniques like contrastive learning (Li et al., 2021), semi-supervised learning (Xiao et al., 2022) and domain adaptation (Franco-Barranco et al., 2022) to the task of mitochondria segmentation. Other approaches (Han et al., 2022; Xie et al., 2023) investigate fully unsupervised methods to segment the mitochondria and other organelles within a cell. While these unsupervised methods achieve promising results without any labeled data, their accuracy and reliability are not yet sufficient for downstream biological analyses based on the resulting segmentation masks. Given the significant effort involved in sample preparation and imaging, biologists typically prioritize the highest possible segmentation quality. As such, the modest annotation effort required to boost performance often justifies the investment. Consequently, unsupervised approaches have seen limited adoption in practical settings.

2.3.3 Analysis of Mitochondrial Ultrastructure

The studies discussed in the previous section focus on segmenting the whole mitochondria but do not make any attempt yet at analysing the ultrastructure.

While mitochondria were discovered by von Kölliker in 1857, the exploration of the ultrastructure relied on the invention of electron microscopy and was pioneered by Palay et al. (1955) and Sjöstrand (1956) which lead to the standard visualisation described above and shown in Figure 2.1. As discussed by Frey et al. (2000), it was

initially thought that the cristae shows broad openings to the intermembrane space and extending across the matrix almost to the other side of the mitochondrion.

However, it was later reported (Lea et al., 1989; Mannella et al., 1994) that the cristae do not have to follow the lamellar structure but can also appear tubular, triangular (Zick et al., 2009) or show concentric circles (Gilchrist et al., 1996). Moreover, the cristae do not show broad openings but are connected to the intermembrane space through tubular structures and it was discovered that the cristae can undergo substantial rearrangements in response to stress or during the initiation of the apoptotic cell death (Frey et al., 2000; Huang et al., 2023; Mannella, 2006; Zick et al., 2009).

For example, it has been observed that the surface area of the cristae membrane is strongly correlated with mitochondrial oxygen consumption (Perkins et al., 2012) and the shape of the cristae was shown to directly regulate mitochondrial performance (Cogliati et al., 2013).

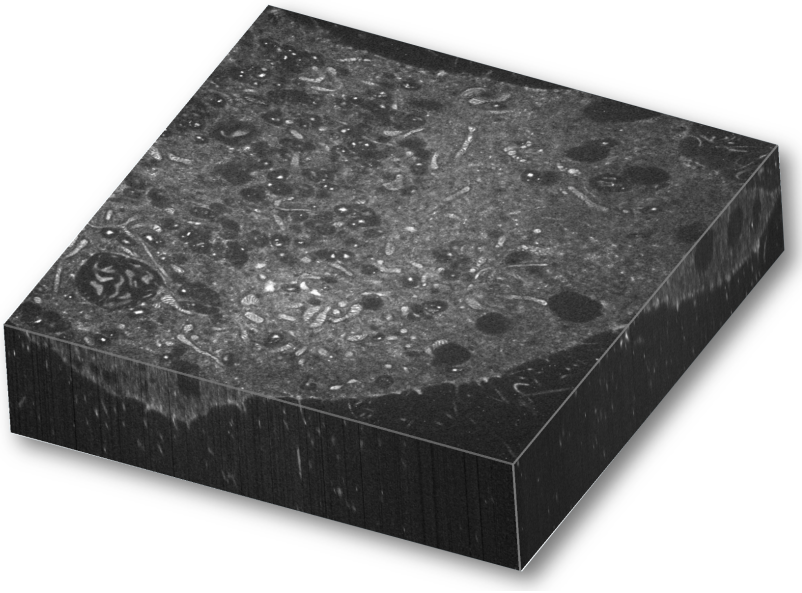
In recent year, there has also been emerging evidence shown that mitochondrial dysfunction might be connected to several diseases like Alzheimer's (Cadonic et al., 2015), diabetes (Yoon et al., 2011) or other rare diseases (Dard et al., 2020) and it has been reported that several human diseases like Parkinson's disease (Beal, 2007; Trimmer et al., 2000) or Wolf-Hirschhorn syndrome (Dimmer et al., 2008) are associated with an alternated mitochondria morphology or changes in the cristae structure (Huang et al., 2023)

However, these studies only describe the changes qualitatively and only very few studies, like Vincent et al. (2016) also provide quantifications for the changes of the inner structure. Most other studies that investigate the cristae (Bílý et al., 2021; Brandt et al., 2017; Fry et al., 2024; Stephan et al., 2020; Stoldt et al., 2018) with manual annotation looked at smaller numbers of mitochondria and cells due to laborious annotation process.

Only recently have machine learning methods been applied to the automatic segmentation of the cristae structures. Parlakgül et al. (2022) investigate mitochondria and the ER with FIB-SEM in lean and obese animals. Their 3D segmentation

includes the cristae of the mitochondria but it is not considered in their quantitative analysis. However, as discussed by Duranova et al. (2020), there is a growing interest in more quantification of the mitochondrial ultrastructure.

Suga et al. (2023) develop a human in the loop platform to automatically segment the mitochondria and their cristae which they divide into lamellar and tubular cristae. Subsequently, they conduct a quantitative analysis. Their work is based on FIB-SEM and they are investigate around 400 mitochondria from 5 cells. This growing interest in the cristae structure and demand for tools is shown also by studies like Wang et al. (2024b) who provide a new platform for segmenting EM image which also includes the segmentation of the inner structure.



3

Pan-Stained ExM and Uncertainty Analysis

Contribution

This is the first time segmentation algorithms are explicitly developed for panExM images and in this chapter, I introduce the panExM data and investigate how the data is interpreted by different annotators. I subsequently develop a new semi-supervised segmentation algorithm which is capable of utilising the additional information contained in the inter-annotator disagreement. Finally, I theoretically analyse the popular Dice loss and show its favourable properties in setting with noisy annotation labels.
This work has in parts been published as *Leveraging Inter-Annotator Disagreement for Semi-Supervised Segmentation* in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5).

Contents

3.1 Introduction 35
3.2 Expansion Microscopy with PanExM 35
3.3 Difference to EM 38
3.4 Annotation and Analysis of Uncertainty 39
3.5 Multiple Annotators for Semi-Supervised Learning . . 44
3.5.1 Method 46
3.5.2 Experiments 49
3.5.3 Results 50
3.5.4 Conclusion 52
3.6 Analysis of Dice Loss with Noisy Labels 52
3.6.1 Theory 53
3.6.2 Experiments 55
3.7 Discussion 57

3.1 Introduction

The pan-staining was only introduced recently by M'Saad et al. (2020) and this work is therefore the first to develop machine learning segmentation methods for it. In this chapter, I investigate the data, compare it to electron microscopy images and develop an annotation protocol for the segmentation of the ultrastructure of the imaged mitochondria.

Moreover, I investigate how different annotator interpret the same data. Images from the biomedical domain often suffer from a higher noise to signal ratio and greater ambiguity compared to natural images. This often leads to a higher uncertainty with respect to the true ground truth. Moreover, annotations in this domain often require experts and therefore most datasets only provide a limited amount of labelled samples. Motivated by this, I present a general approach that combines both aspects and allows to utilise the information provided by the annotator disagreement about the underlying uncertainty in the data to boost the performance of semi-supervised learning approaches.

Finally, I theoretically investigate the Dice-loss and show its performance advantage with noisy annotations.

3.2 Expansion Microscopy with PanExM

One of the main advantages of ExM is that it does not require specialised hardware. In the initial experimentation phase of the project, the data was imaged with a Leica SP8 confocal microscope. However, due to the higher imaging speed this microscope was later replaced with a spinning disk microscope (Andor Dragonfly) which resulted in a faster throughput and images of similar quality. All images in used in the analysis have been recorded by Yuan Tian at Yale University.

The voxels in the images of the expanded cells are of size $0.25\mu m \times 0.1\mu m \times 0.1\mu m$. Moreover, the expansion factor was determined by comparing the nucleus size before and after expansion by our collaborators and was on average around 15. This implies that the size of one voxel depicts around 7nm of the original cell.

The resulting image stacks have a planar size of 2048×2048 pixels with a depth that varies between 71 and 380 slices with a mean of 210.

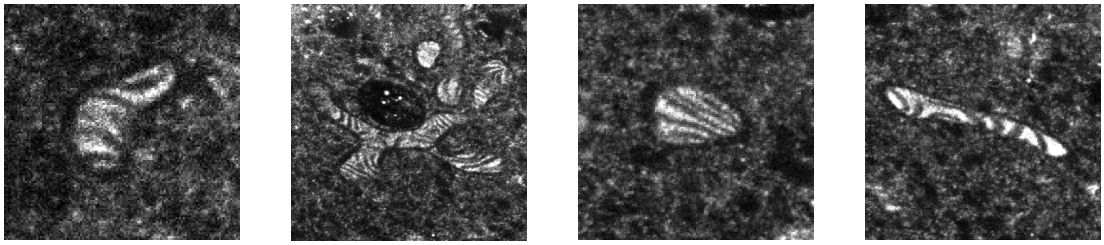
The main interest of this work lies in the image channel showing the NHS ester which allows to visualise the whole proteome (M'Saad et al., 2020) and this thesis focus on the analysis of mitochondria. As shown in Figure 3.1, mitochondria are characterised in the NHS channel by a striped inner structure showing the cristae which can vary in its orientation, intensity and thickness. Additionally, a darker halo surrounds these mitochondria indicating the intermembrane space. Moreover, for mitochondria surrounded by cytoplasm, their intensity is generally higher than the background. Finally, it can be seen that their size and shape can take a variety of forms ranging from circle-like in Figure 3.1c or tube-like in Figure 3.1d to fairly complex branching-shape in Figure 3.1b.

However, these characteristics are not always applicable and regularly the staining process lead to less distinctive features as shown in Figure 3.2. As seen in Figure 3.2a, it is possible that there is no dark shadow surrounding the mitochondria and the boundary between mitochondria and cytoplasm becomes blurred and ambiguous. Moreover, the striped structure within the mitochondria can be missing completely as shown Figures 3.2b and 3.2c. On top of that, the cytoplasm can display misleading patterns as shown in Figure 3.3 which could be easily mistaken as poorly stained mitochondria.

Finally, it is also not possible to rely on the higher pixel intensity within mitochondria as shown in Figure 3.2d as the nucleus appears as bright or brighter than the mitochondrion. The nucleus can also be seen in Figure 3.4 in the top left corner as a homogeneous area without any specific patterns in it.

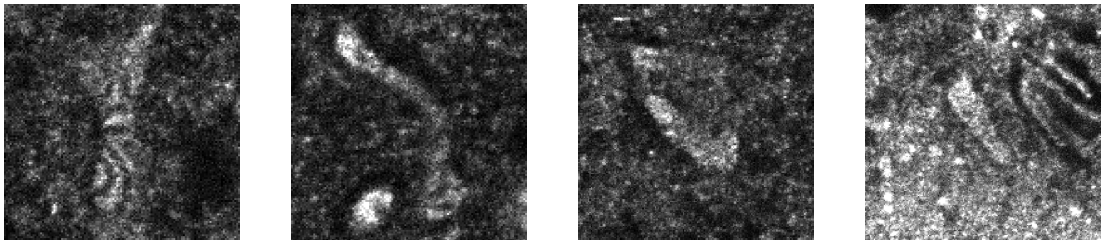
This lack of a unique feature which reliably differentiates the mitochondria from the background makes this task very challenging and leads to the conclusion that traditional machine learning model with hand-crafted features are not going to produce meaning full segmentations.

One great opportunity which panExM provides for automated segmentation algorithms is that additional dyes and immunolabeling can be applied to the sample.



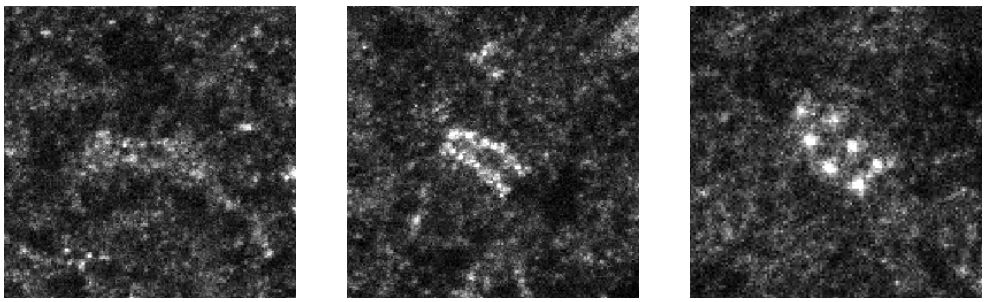
(a) (b) (c) (d)

Figure 3.1: Clearly imaged mitochondria show a dark shadow surrounding them separating them from the cytoplasm, distinctive stripes and a variety of shapes. All samples in this thesis showing the pan-staining have been imaged by Yuan Tian at Yale University.



(a) (b) (c) (d)

Figure 3.2: Some mitochondria are less clearly imaged and lack a clear boundary to the cytoplasm (a) or the stripped inner structure (b) & (c). Moreover, the surrounding area can be as bright as the mitochondria (d).

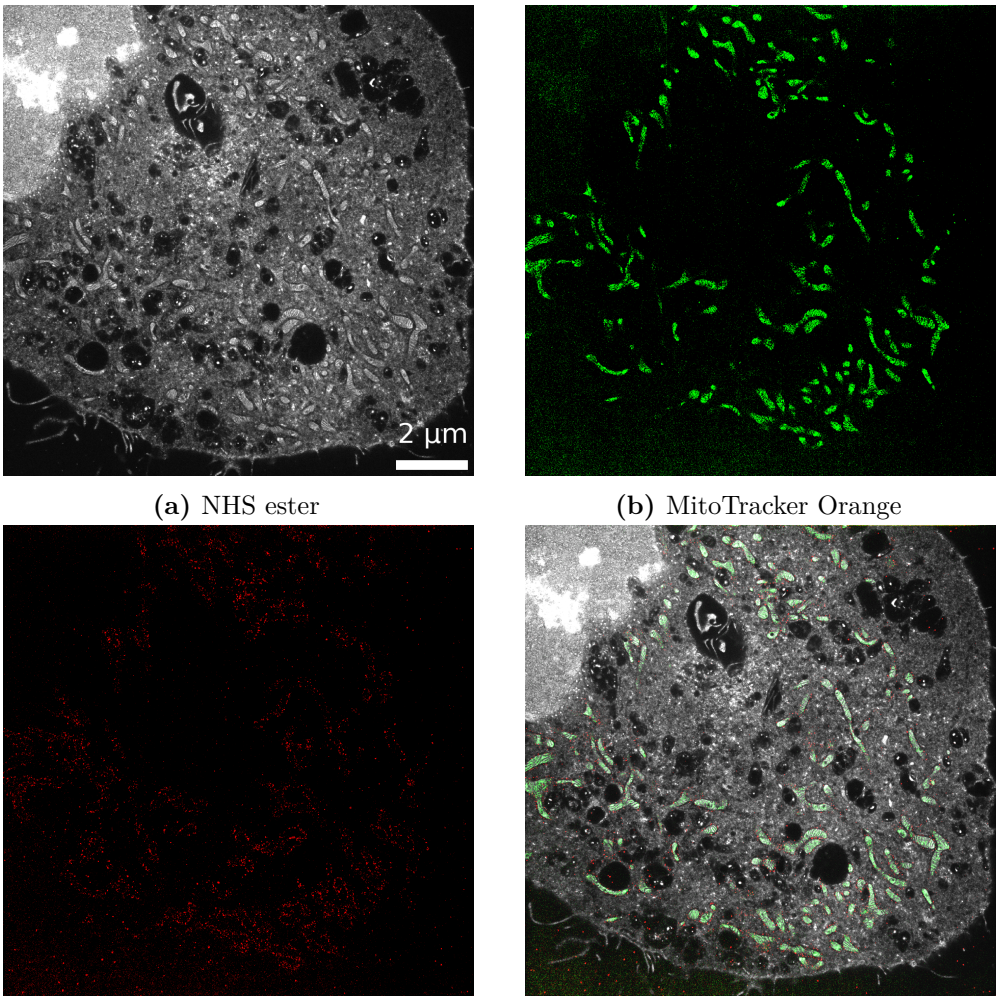


(a) (b) (c)

Figure 3.3: Cytoplasm patterns: Patterns in the cytoplasm can appear similar to poorly imaged mitochondria.

Figure 3.4 gives an overview of the different channels. As it can be seen, using the MitoTracker Orange dye (Figure 3.4b) allows to stain the mitochondria within the corresponding sample shown with the NHS ester in Figure 3.4a. On top of that, an antibody labelling with Tom20 is available. However, as it can be seen in Figure 3.4c the binding is very unspecific and noisy. However, in Figure 3.4d which overlays all channels, it can be seen that it shows a higher concentration

around the boundaries of the mitochondria. In Section 6.3, I provide a quantitative assessment of the localisation of the Tom20.



(a) NHS ester **(b)** MitoTracker Orange

(c) Tom20 antibody labelling **(d)** All channels combined

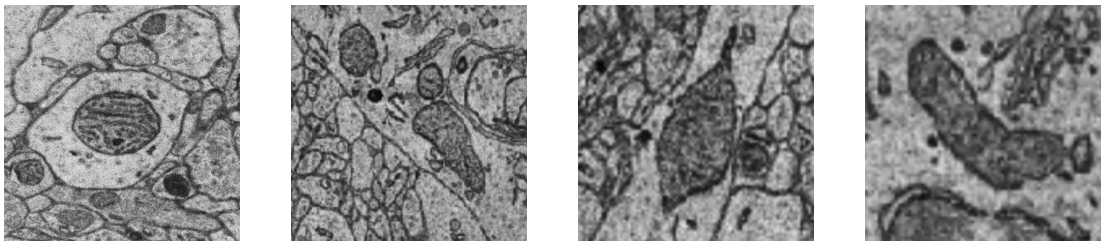
Figure 3.4: PanExm with additional dye and immunolabeling: PanExM (a) can be used with different dyes like MitoTracker (b) or Tom20 (c) to highlight specific structures within the cell. However some of these like Tom20 show substantially noise. All samples in this thesis showing the pan-staining have been imaged by Yuan Tian at Yale University.

3.3 Difference to EM

As described in Section 2.3.2, applying computer vision methods to EM data in order to extract the mitochondria is already an established field. Therefore it is important to understand how the expansion microscopy data is different in order to judge how established methods have to be adopted to this new domain.

Figure 3.5 shows several examples of mitochondria imaged with electron microscopy. It can be seen that the double membrane of the mitochondria is often clearly visible as a dark boundary and clearly separates the mitochondrion from the rest of the cell. On top of that, the cytoplasm shows a much more regular pattern which makes methods that rely on a strong image gradient to separate both segments much more likely to succeed.

Additionally, to the best of our knowledge Suga et al. (2021) and Wang et al. (2024b) are the only studies looking at the ultrastructure of the mitochondria using a machine learning approach which is therefore a very unexplored field.



(a) **(b)** **(c)** **(d)**

Figure 3.5: Visual comparison of mitochondria imaged via EM showing a clearer boundary and less noise in the cytoplasm compared to panExM. Images are examples from the MitoEM dataset (Wei et al., 2020b)

3.4 Annotation and Analysis of Uncertainty

In order to train and evaluate any future models and segment the ultracellular structure of the mitochondria, the researchers working on the project (Mariia Dmitrieva, Hongqing Han, Yuan Tian (Yale), Hannahmariam Mekbib (Yale) and myself) annotated eleven different image crops which were picked by a microbiologist such that each stack was labelled by three different annotators.

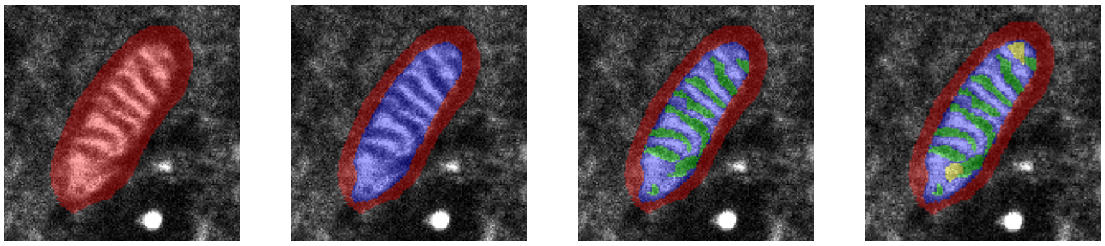
Annotation Protocol

Each stack had a depth between 20 and 35 slices and a size of 256 × 256 pixels with all of them including all three channels shown in Figure 3.4. Together with Dr. Dmitrieva, I developed and refined an annotation protocol during three annotation

rounds with the aim to minimise the annotation effort but still maximising the accuracy. The final labels are called "Intermembrane space" (abbreviated as I), "Cristae" (C), "Matrix" (M) and a final ambiguous class "Matrix or Cristae" (MoC).

The inclusion of the ambiguous class "MoC" was necessary to give annotators the option to express their uncertainty in the presence of unclear stainings. Figure 3.2 includes several examples of mitochondria in which annotators felt that making a decision between matrix or cristae was not possible. In this way, we allow annotators to label the mitochondrion as different from the background but without forcing them to make an arbitrary decision between matrix and cristae.

For the annotation, we used the open-source software ITK-SNAP[1] which was developed for 3D medical images. It allows annotators to overlay all three channels in form of an RGB image and view the image stack from different angles. The annotation itself works by drawing the boundary points which then get connected and filled by the software to form a polygon. This implies that the annotation protocol should aim to minimize the boundary length of the required annotations to save time.



(a) 1. Inter. space **(b)** 2. Matrix **(c)** 3. Cristae **(d)** 4. Ambiguous

Figure 3.6: Annotation protocol: Visualisation of the different steps.

The final annotation protocol is illustrated in Figure 3.6 and involves the following steps:

1. The outline of the mitochondrion is labelled as "Intermembrane space"

2. All of the inner mitochondrion is overwritten and labelled as "Matrix".

3. The "Cristae" is added to the inner mitochondrion.

[1] http://www.itksnap.org/pmwiki/pmwiki.php

4. Finally, any area which cannot be differentiated into either “Matrix” or “Cristae” is overpainted by the ambiguous class “Matrix or Cristae”.

This procedure was used to ensure that there are no gaps between intermembrane space and the inner mitochondrion. Moreover, the order of the annotation steps minimizes the annotation effort as cristae usually occupies less volume and has a smaller surface than the matrix as it only shares short boundaries with the intermembrane space.

One of the challenges of panExM and super-resolution microscopy in general is the relatively low contrast and unclear boundaries stemming from the fact that the microscope operates on the edge of its feasible resolution. For example, the background can closely reassemble the boundary area of the mitochondria making a definite localisation of the boundary difficult, even for experienced annotators. This problem is especially prevalent when it comes to the 3D nature of the data and the fact that the top or bottom of a mitochondria might be hardly visible in the NHS channel due to the lower depth resolution.

Analysis of Annotations

To investigate this further, three annotators labelled each image stack. Overall I find that 83% of the pixels were annotated as background, 8.6% as intermembrane space, 3.2% as matrix, 1.2% as cristae and 3.1% as ambiguous between matrix or cristae.

Upon closer examination of the final annotations, I observed a substantial degree of inter-annotator variability. This observation aligns with previous findings, such as those reported by Rohani et al. (2020), who also noted significant disagreement among annotators when labeling mitochondria in fluorescence microscopy images.

The disagreement between different annotators can be captured by Fleiss’s kappa (Fleiss et al., 2003) which extends Cohen’s kappa (Cohen, 1960) to cases with more than two raters (McHugh, 2012). Fleiss’s kappa calculates the agreement between annotator while adjusting for any agreement that could occur by chance. In the current case, I find a value of 0.66 with a standard deviation of 0.03.

However, the analysis is complicated by the spatial dependence between neighbouring pixels. I therefore follow another approach in which I consider for each annotated crop all possible pairings of annotators and compute the pairwise agreement. This approach also allows us to judge what performance can reasonably be expected from the ML model later. I evaluate the agreement between two annotators in terms of the intersection over union (IoU) and the Dice score.[2].

For two binary segmentations A and B these metrics are defined as

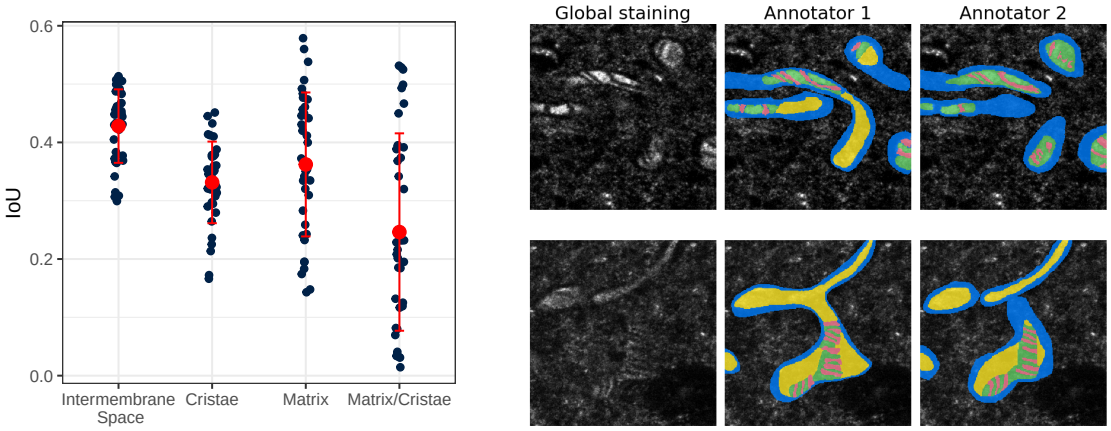
$$\mathrm{IoU}(A,B) := \frac{|A \cap B|}{|A \cup B|} \qquad \text{and} \qquad \mathrm{Dice}(A,B) := \frac{2|A \cap B|}{|A| + |B|}. \tag{3.1}$$

Only the result for the IoU shown in Figure 3.7(a) as the Dice score shows the same trend. As it can be seen, the human agreement in terms of the IoU lies around 0.39 for the intermembrane space, matrix and cristae. As expected, a much larger variety is seen for the ambiguous class as some annotators were more cautious than others.

Compared to other segmentation benchmarks these scores seem very low and Figure 3.7(b) offers some insight into the underlying reasons: On top of the different tendencies to mark some areas as ambiguous, the 3D nature of the data leads some annotators to annotate one additional slice when a mitochondria is just about to be out of sight. These annotations can however potentially lead to many disagreeing pixels and therefore low IoU and Dice scores.

To investigate this further, Figure 3.8 plots the agreement between annotators in terms of the distance from the outer membrane. Therefore, I measure the distance in voxels from the union of the two segmentations. The distance is binned into integer values and for each bin, the accuracy is calculated as the proportions of pixel with the same label over the number of pixels present at this distance. It can be seen that especially the boundaries cause the disagreement between annotators.

[2]A more detailed discussion of the model evaluation can be found in Section 4.2.4.



(a) Quantitative evaluation **(b)** Visualisation for two samples

Figure 3.7: Annotator agreement across categories: The relatively low IoU scores in (a) can be explained by the different interpretation of the 3D data as seen in (b).

Conclusion

The analysis of the inter-annotator disagreement shows that there are often different plausible annotations possible based on the global staining and therefore there is not one ground truth annotation. While the annotations do not necessary agree on every pixel, each annotation itself is biologically consistent. This observation motivates the algorithm I develop in Section 5.2 which automatically finds areas of poor contrast staining and corrects the segmentation in a biologically plausible way.

Additionally, in the next section I demonstrate how the information contained in conflicting annotations can be exploited and in Section 3.6, I investigate how the Dice loss responds conflicting labels.

Finally, in order to train the main segmentation model in Chapter 4, I conduct another round of annotations after discussing additional examples with the annotation team to establish a stronger, shared standard for the different categories. Moreover, I manually review every annotated crop and corrected areas that deviated from this standard before they are used for the model training. For this correction, I use the napari software[3] which provides better support for 3D visualisation and therefore allowed for a more consistent annotation of adjacent slices.

[3]https://napari.org/stable/

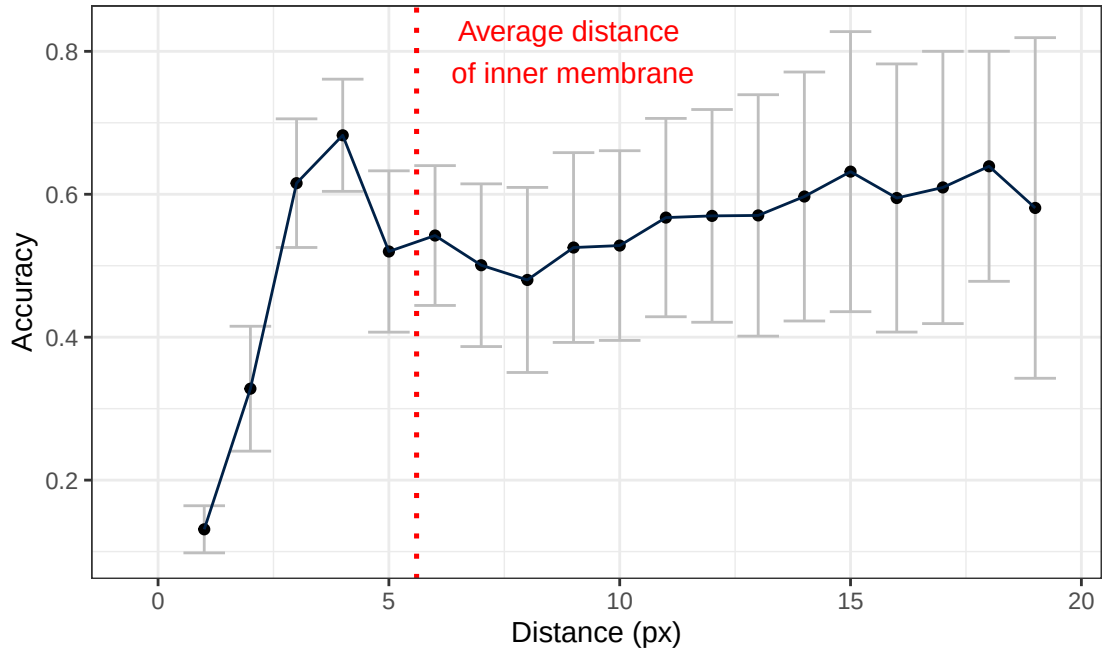


Figure 3.8: Disagreement at boundaries: Plotting the accuracy against the distance from the outer membrane shows that the disagreement between annotators is especially high close to the boundaries. The dashed red line shows the average distance between the outer and inner membrane.

3.5 Multiple Annotators for Semi-Supervised Learning

The challenge of high inter-annotator variability described in Section 3.4 is common in biomedical settings and have recently been investigated more intensively (Boers et al., 2022; Sudre et al., 2019; Zeng et al., 2021; Zhang et al., 2020a).[4] This variability is driven by greater underlying semantic ambiguity compared to natural images, often due to factors such as low contrast, overlapping structures, heterogeneous tissue morphology, and limited image resolution. Additionally, biological images are frequently affected by inconsistencies in the imaging process, including poor staining quality, variations in microscope settings, and sample artefacts.

Moreover, biomedical images often requires experts for the annotation process which increases the costs of collecting labels and leads to dataset with a smaller amount of available labelled samples compared to natural images. This lead

[4]The results of this section have been published in Sauer et al. (2023) © 2023 IEEE and parts have been quoted verbatim from it with permission.

to the development of semi-supervised methods, which try to efficiently exploit available unlabelled data (Laine et al., 2016; Sajjadi et al., 2016; Tarvainen et al., 2017; Xia et al., 2020).

In this section, I present a method which combines a setting of high inter-annotator variability and semi-supervised learning. Instead of focusing on the challenges associated with high inter-annotator disagreement (Boers et al., 2022; Sudre et al., 2019; Zeng et al., 2021; Zhang et al., 2020a), I point out the advantages that having access to different annotations of the same image can offer. I argue that disagreeing annotators provide us with the opportunity to estimate the aleatoric uncertainty (Hüllermeier et al., 2021) directly from the data and show how this uncertainty estimate can efficiently improve pseudo-labelling approaches (Lee, 2013). In the experimental section, I demonstrate the effectiveness of my approach on two public data sets, one containing dermoscopy images and one containing prostate MRI images as shown in Figure 3.10. These datasets were chosen because they offer multiple and partly conflicting annotations per image. Moreover, I focus on publicly available data to make the approach more accessible to the biomedical imaging community as the panExM data was not available or well known in the community at the point of publishing.

Uncertainty estimation

The uncertainty in a model's predictions can be categorized into epistemic and aleatoric uncertainty (Hüllermeier et al., 2021). Epistemic uncertainty arises from uncertainty in the model parameters due to limited data and can, in principle, be reduced with more observations. In contrast, aleatoric uncertainty reflects the inherent randomness in the data-generating process and cannot be eliminated.

This distinction can be illustrated using the example of a potentially biased coin that lands heads with probability $p \in (0, 1)$. Suppose we observe n past outcomes and aim to predict the result of the next coin flip. In this case, the aleatoric uncertainty corresponds to the inherent randomness of the coin flip, captured by the

probability p, while the epistemic uncertainty stems from the discrepancy between our estimate $\hat{p}_n$, based on the finite sample, and the true but unknown value of p.

Several approaches have been proposed to estimate these uncertainties: MC dropout (Gal et al., 2016) uses several forward passes through the model with different drop-out conditions to estimate the model uncertainty by considering the variation in the different predictions. Other methods include combining the predictions from an ensemble of models (Lakshminarayanan et al., 2017) trained with different initial states or using models trained with different weights for classification errors (Shi et al., 2022). The aleatoric uncertainty has been tried to be captured by combining a U-Net with a variational autoencoder for a generative model (Kohl et al., 2018) or by placing a Gaussian prior on the data distribution (Xiang et al., 2022).

Multiple annotators

Due to the lower resolution, contrast and varying imaging quality commonly found in biomedical images, conflicting annotations even by experts are common and pose a challenge to segmentation methods, as there is no clear optimal way to combine different annotations into one ground-truth in the training process (Boers et al., 2022). Several approaches (Sudre et al., 2019; Warfield et al., 2004; Zhang et al., 2020a) try to model each annotator and their reliability separately and base their ground-truth estimation on this. Another approach is to account for conflicting areas by adopting the loss function to allow for ambiguous regions (Zeng et al., 2021).

However, to the best of our knowledge, no work used conflicting annotators as a source for uncertainty estimation to improve segmentation algorithms so far.

3.5.1 Method

For every labelled image $X_i \in \mathcal{X} = \mathbb{R}^{W \times H}$ with $i = 1, \ldots, N$, there are k_i different annotations $Y_i \in \mathcal{Y}^{W \times H}$ available where $\mathcal{Y}$ is restricted to be the case of binary labels $\{0, 1\}$ in this work. Moreover, in the following I denote the unlabelled images by Z_j with $j = 1, \ldots, M$ and the segmentation network by $f_\theta : \mathcal{X} \to [0, 1]^{W \times H}$ which predicts a probability of belonging to the positive class for every pixel.

Basic pseudo labelling approaches make a prediction $\tilde{y}_j = f_\theta(Z_j)$ for every unlabelled image and restrict themselves to the pixels predicted with a high confidence in either the positive or negative class with a mask given by $\mathcal{M}_j = \mathbb{1}[\tilde{y}_j > \tau_{pos}] + \mathbb{1}[\tilde{y}_j < \tau_{neg}]$ based on some thresholds τ_{pos} and τ_{neg}. The restricted predictions are denoted by by $f_\theta^{\mathcal{M}}(Z_j) = \mathcal{M} \odot f_\theta(Z_j)$ and the restricted pseudo labels by $\hat{Y}_j^{\mathcal{M}} = \mathcal{M} \odot \mathbb{1}[\tilde{y}_j > 0.5]$ where $\odot$ denotes the element-wise multiplication. Given some loss function l (e.g. the binary cross entropy), the training procedure in iteration t now takes the form of minimizing

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_i l(f_\theta(X_i), \bar{Y}_i) + \frac{\lambda(t)}{M} \cdot \sum_j l(f_\theta^{\mathcal{M}}(Z_j), \hat{Y}_j^{\mathcal{M}})$$

where $\bar{Y}_i = s(\{Y_i^j\}_{j=1}^{k_i})$ is some ground-truth aggregation like majority voting and $\lambda(t)$ is a scaling factor that depends on the current training iteration t.

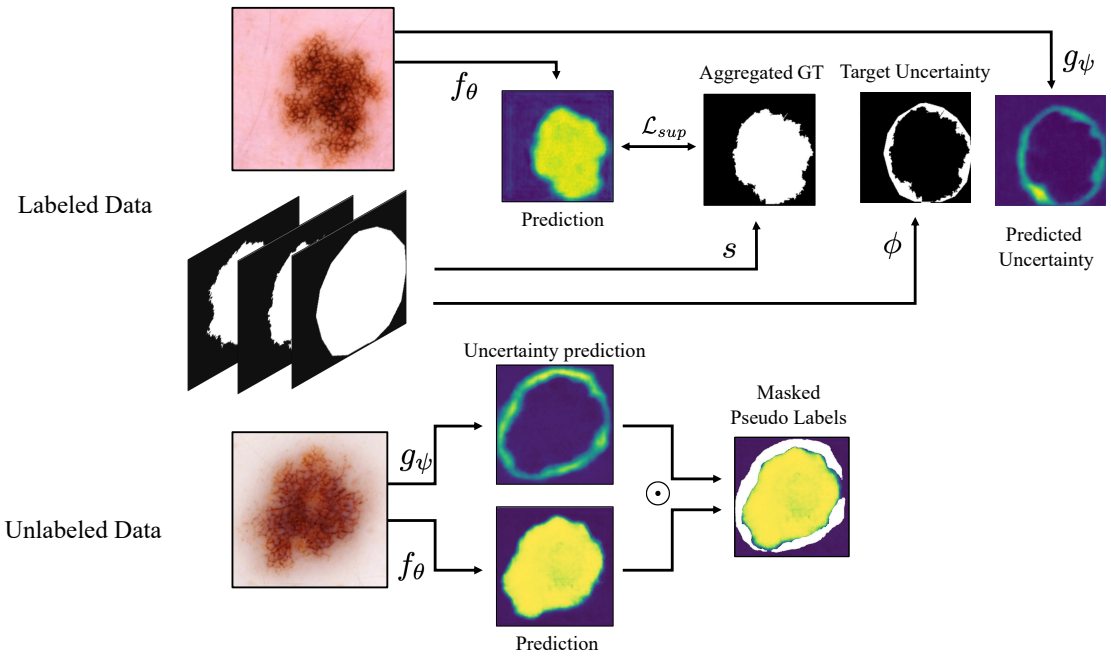


Figure 3.9: Method Overview: The labelled data is used both to train the uncertainty detection network g_ψ and for the supervised part of the loss function of the training of the segmentation network f_θ. Moreover, for the unlabelled data we create both a segmentation and uncertainty prediction but only use the areas of low uncertainty for the pseudo labelling step. © 2023 IEEE

Similar to the approach developed by Rizve and colleagues (Rizve et al., 2021), I propose to base the pseudo-label mask $\mathcal{M}_j$ not only on the confidence of the model

but also on the uncertainty of the prediction. In order to estimate the uncertainty, the dominant approach is to use MC-Dropout or Bayesian networks and it has been shown that this allows to improve the segmentation performance. However, while Bayesian networks are in general computationally challenging to train, MC-Dropout requires numerous forward passes for every sample to estimate the uncertainty which induces a high computational overhead. Alternatives like ensemble methods on the other hand require both training a large number of networks as well as forward passes through each of them for every prediction. Additionally, there is still debate how well these methods capture the true underlying uncertainty (Caldeira et al., 2020).

I therefore propose to estimate the aleatoric uncertainty directly from the annotator disagreement, which allows to use all the available information contained in the multiple annotations efficiently. Moreover, it avoids the computational overhead of the alternative methods described above.

The proposed approach is summarised in Figure 3.9. In order to estimate the aleatoric uncertainty, I propose to combine the available annotations for each image into an uncertainty measure $\phi : \mathcal{Y} \times ... \times \mathcal{Y} \rightarrow [0, 1]^{W \times H}$. Examples for ϕ include a binary indicator for annotator disagreement per pixels or the normalised standard error of annotations per pixel. Since even experts cannot agree on the label for these pixels, it is reasonable to assume that these regions are characterised by a high aleatoric uncertainty. Subsequently, I train an additional network $g_\psi : \mathcal{X} \rightarrow [0, 1]^{W \times H}$ on the data with multiple annotations available to predict the value of the aleatoric uncertainty at the pixel level using standard supervised techniques with the target values ϕ derived as described above.

This trained uncertainty prediction network g_ψ can now be utilised in the semi-supervised training process of the actual segmentation network f_θ based on pseudo-labels. In order to define these pseudo-masks in the training process, the current version of the segmentation network f_θ is used to make a prediction on the unlabelled data as described above. Additionally, the trained uncertainty network g_ψ makes predictions as well which areas are likely to cause annotators to disagree and therefore indicate a high level of aleatoric uncertainty. This uncertainty

estimate is subsequently used to further refine the pseudo label mask from the prediction confidence by excluding areas with a high predicted uncertainty. The final pseudo-masks are therefore set to $\overline{\mathcal{M}} = \mathbb{1}[g_\psi(Z_i) < \tau_{cert}] \odot \mathcal{M}$ based on some threshold value τ_{cert}. In this way, both the estimates of the epistemic uncertainty in the form of $\mathcal{M}$ and the aleatoric uncertainty in the form of g_ψ influence the definition of the pseudo masks for the unlabelled data. Finally, $\mathcal{L}(\theta)$ is minimized and the training procedure is continued until convergence.

3.5.2 Experiments

Data

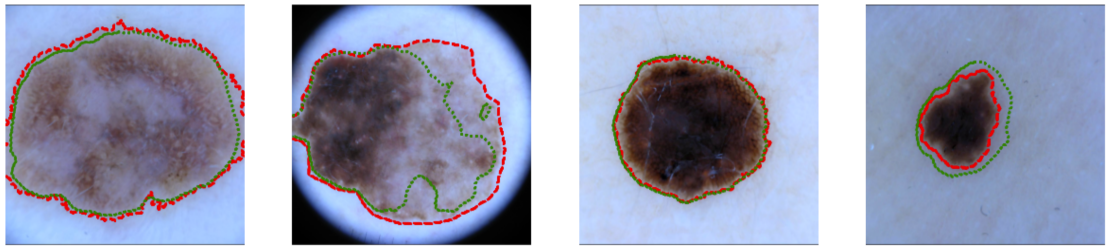
The presented experiments are based on two public datasets which both include multiple annotations per image. The first is the 2018 challenge of the International Skin Imaging Collaboration (ISIC) (Codella et al., 2019) containing colour dermoscopy images coming with up to five annotations per image. For this dataset, a considerable inter-annotator variability has been reported (Ribeiro et al., 2019). The ISIC dataset contains 2594 images in total, only 1125 come with multiple annotations which I use as training data while I split the remaining images into a validation set (20%) and test set (80%).

Secondly, the MRI prostate segmentation data from the 2021 QUIBIQ challenge (Menze et al., 2022) is explored, which provides six annotations per image with 43 images used for training, 5 for validation and 7 for testing (which corresponds to their provided validation set).

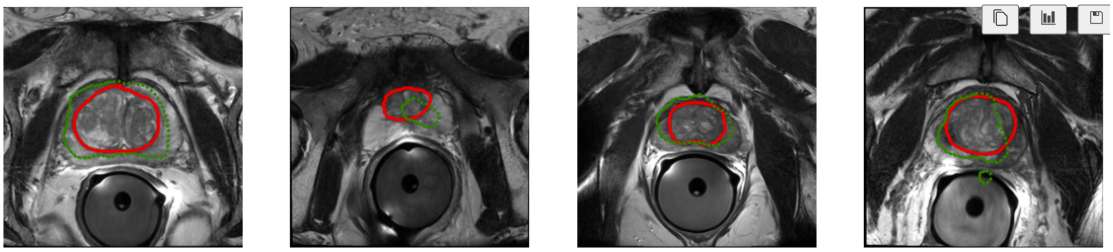
All data is resized to 256×256 pixels and normalised to $[0, 1]$. To simulate the semi-supervised setting, I keep 10% from the training data set as labelled and consider the remaining images as unlabelled by disregarding their annotations.

Implementation Details

In order to focus my experiments on the different sources of uncertainty, I decided to use a simple U-Net architecture for both the segmentation network f_θ as well as the uncertainty detection network g_ψ and I incorporate dropout layers for



(a) Dermoscopy



(b) Prostate MRI

Figure 3.10: Example dermoscopy taken from ISIC (Codella et al., 2019) (a) and prostate MRI form QUBIQ (Menze et al., 2022) prostate (b) dataset with the majority vote in red and the prediction in green. © 2023 IEEE

the MC-Dropout experiments. In order to aggregate the different annotations for the supervised training and evaluation, I consider the majority vote between annotators as the aggregation function s. Moreover, I use the standard deviation for the uncertainty function ψ and the binary cross entropy for the supervised loss l. For the function scaling the influence of the semi-supervised loss I choose $\lambda(t) = \exp(-(1 - t/3000)/0.2)$. Moreover, the Adam optimizer is used with a learning rate of 10^{-5} for both data sets. Finally, I train every model with early-stopping until there is no further decrease in the validation loss for 50 epochs. For the evaluation, I compute the Intersection-over-Union (IoU) and the Dice coefficient.[5]

3.5.3 Results

To judge the effectiveness of using the inter-annotator disagreements as an indication for the aleatoric uncertainty, I first establish a baseline of a fully supervised approach and ignore the unlabelled data. Moreover, I include the results of a simple pseudo-labelling strategy by solely basing the selection of the pseudo labels on the model

[5]The code is available at github.com/AlexSauer/MultiAnnotatorSSL.

Table 3.1: Comparison of different pseudo-labeling approaches showing that my approach outperforms several alternatives on the ISIC (Codella et al., 2019) and QUBIQ (Menze et al., 2022) dataset. © 2023 IEEE

	ISIC		QUBIQ	
Model	IoU	Dice	IoU	Dice
Purely Supervised	0.67	0.77	0.34	0.45
Pseudo Labels (Lee, 2013)	0.69	0.78	0.46	0.57
MC-Dropout (Rizve et al., 2021)	0.70	0.79	0.46	0.58
Ensemble (Lakshminarayanan et al., 2017)	0.72	0.81	0.52	0.63
Ours	**0.74**	**0.83**	**0.54**	**0.66**

confidence similar to (Lee, 2013). Additionally, I compare against the approach suggested by (Rizve et al., 2021) by including the uncertainty measure through eight separate forward passes of a drop-out U-Net and an ensemble of networks. For the ensemble, I first use the available labelled data to train six separate networks with different initial conditions and data orderings as suggested by (Lakshminarayanan et al., 2017). Subsequently, a final model as described in Subsection 3.5.1 is trained with the uncertainty based on the disagreement of the ensemble models.

As it can be seen in Table 3.1, using my approach allows the pseudo labelling approach to make the largest increase in performance for both data sets and some example predictions are shown in Figure 3.10. Additionally, training it is roughly three times faster compared to the MC-Dropout method per training iteration and it does not require the additional training of a full ensemble.

Moreover, I investigate the claim that my model provides more reliable pseudo labels compared to the confidence approach alone. Since this is in an experimental setting, I actually have access to the annotations of the data which I consider as unlabelled and am able to compare the masked pseudo labels to the annotations. For the ISIC data, I find that the precision of the predictions on the unlabelled data rose from 0.64 for the confidence based pseudo labels to 0.73 if we include my uncertainty predictions. For the QUBIQ data, I observe an increase from 0.71 to 0.75.

3.5.4 Conclusion

In this section, I have shown that multiple annotators per sample provide valuable information about the uncertainty in the data and I demonstrated how this uncertainty estimate can be used to improve semi-supervised segmentation techniques by excluding unreliable pseudo-labels. Additionally, my experimental results confirm that the proposed approach improves the resulting segmentation compared to competing approaches and that it reduces the computational complexity of the uncertainty estimation.

In a previous study, Czolbe et al. (2021) find that uncertainty estimations are not useful for semi-supervised or learning when traditional uncertainty estimates are used. However, my work extends this finding by showing that it indeed can be useful if there is access to an estimate of the aleatoric uncertainty based on multiple annotators.

However, the method presented in this chapter should not be considered as an alternative to existing methods which try to extract the underlying ground-truth from multiple annotations: while this work has used a simple majority vote in my experiments to aggregate different labels for the supervised part, it is straightforward to integrate more sophisticated techniques and combine them with my approach.

Nevertheless, I found that the increase in performance due to several annotations is smaller than labelling additional samples as I describe in more detail in Chapter 4. In a new study with a fixed annotation time budget, I therefore recommend to cover as many different samples as possible instead of spending the constrained annotation time on multiple annotations of the same instance.

3.6 Analysis of Dice Loss with Noisy Labels

As discussed in the last two sections, noisy labels are ubiquitous in biomedical imaging and there is a mature literature that proposes methods to tackle noisy labels in the classification setting (Song et al., 2023). Many works in particular investigate the effect of noisy labels on the very prominent cross entropy loss (Wang et al., 2019; Zhang et al., 2018). However, in biomedical imaging tasks the Dice loss (Milletari

et al., 2016) enjoys a growing popularity (Ma et al., 2021) and can show superior performance. To my best knowledge and surprise, no similar investigation of the Dice loss in the presence of noisy labels has been published. In this section, I show that the Dice loss can have favourable characteristics in the presence of noisy labels.

3.6.1 Theory

The Dice loss is defined as one minus the Dice metric which is also known as F1 score or Sørensen index. For a ground truth segmentation X and a predicted segmentation Y, the Dice metric is given by

$$D(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2TP}{T+P} = \frac{2TP}{2TP + FN + FP}$$

where T, P, TP, FN, FP refer to the truely positive cases, the positively predicted, the true positive, false negative and false positive cases respectively. However, in this form, the metric is based on a binary segmentation, so in order to use it as a loss function it needs to be transformed in its soft form based on the true segmentation target X and the outputted probabilities of our network P:

$$SD(X,Y) = \frac{2\sum_i x_i p_i}{\sum_i x_i + \sum_i p_i} = \frac{2\langle X, P\rangle}{\langle 1, X\rangle + \langle 1, P\rangle}.$$

In order to gain a better understanding of this loss function, its gradient with respect to P which are used in the backpropagation algorithm to train the model are considered. Introducing $M := \sum_i x_i + \sum_i P_i$ to shorten the notation for the denominator, first consider the gradient with respect to the predicted probabilities

$$\frac{\partial SD}{\partial P} = 2\frac{XM - \langle X, P\rangle}{M^2} = \frac{1}{M}(2X - SD) \in \mathbb{R}^n.$$

However, for the training of the network the more relevant quantity are the gradients with respect to the logits.[6] The derivative of a sigmoid $\sigma(x) = \frac{1}{1+e^{-x}}$ is given by $d\sigma/dx = \sigma(x)(1-\sigma(x))$. Together with the chain rule, this implies that the gradients with respect to the logits are given by:

[6] This point is missed by Kervadec et al. (2023) and Tilborghs et al. (2022) which prevents them from reaching the relevant conclusion with regard to noisy settings.

$$\frac{\partial SD}{\partial P} = \frac{1}{M}\left(2X - SD\right) P(1-P).$$

It can be seen that $P(1-P) \in [0, 25]$ and the gradients will be larger for uncertain pixels ($p \approx 0.5$) and scaled down for pixels that are already settled on a specific direction (either $p \approx 0$ or $p \approx 1$).

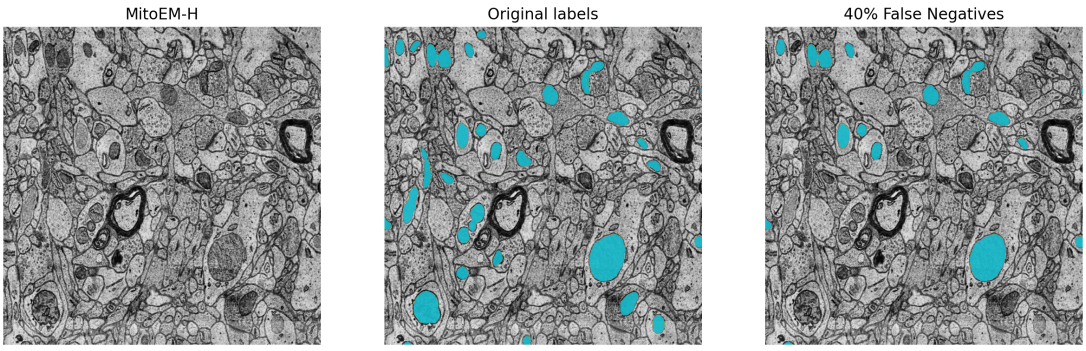
For comparison, the binary cross entropy (bCE) based on the log-likelihood $LL = x_i \log x_i + (1 - x_i) \log(1 - x_i)$ has gradients given by:

$$\frac{\partial LL}{\partial P} = x_i(1 - p_i) - (1 - x_i)p_i.$$

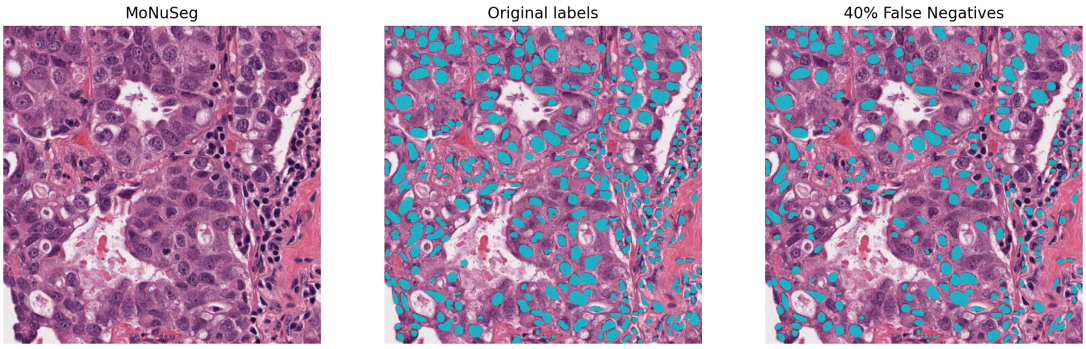
For noisy samples, this result has interesting consequences: If the model has a strong prediction on a pixel but the label is flipped, the dice loss will scale down the gradients. However, for the binary cross entropy loss, the gradient is in full force in the wrong direction.

Moreover, the bCE loss is symmetric in positive and negatives labels whereas the Dice loss differentiates between false positives and false negatives. Investigating the gradients, it can be seen that the gradients are proportional to $(2X - SD)$ and are always bigger for a positive pixel than for a negative one as $SD \in [0, 1]$ and therefore $|2 - SD| > | - SD|$. Therefore, the loss naturally puts a stronger focus on the foreground. This has implications for noisy annotations: Because of stronger gradients for positive pixels, the dice loss is better suitable for a scenario in which the annotator is likely to miss annotating some regions of interest. However, it is more vulnerable to additional annotations that do not indicate the real ground truth of interest.

Additionally, it should be noted that true negatives ($x = 0$ and $p \approx 0$) almost go unnoticed in the dice score which stays basically unchanged. On the other hand, a true positive ($x = 1$ and $p \approx 1$) increases the dice score. This point can also be seen from the fact that TN is missing in the dice score definition.



(a) MitoEM-H



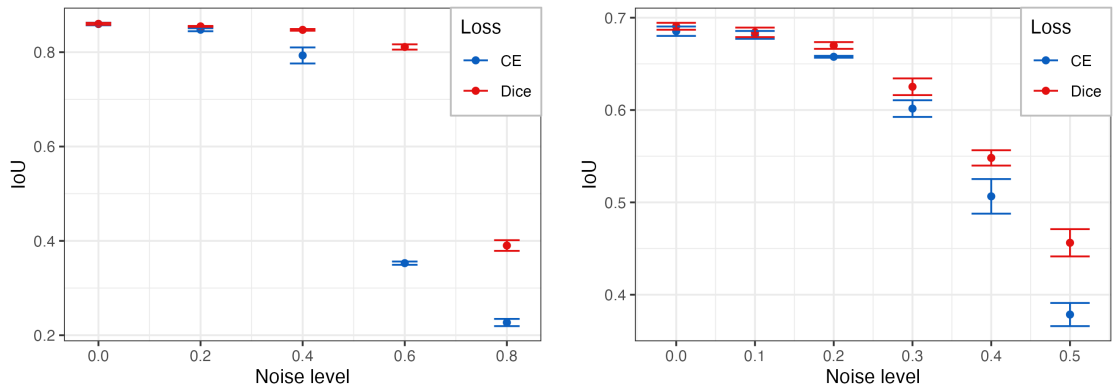
(b) MoNuSeg

Figure 3.11: Visualisation of the (a) MitoEM dataset and (b) MoNuSeg dataset. The first column shows the original data which is overlayed in the second column with the ground truth segmentation mask. The third column illustrates the case in which a fraction of $\eta = 0.4$ of the mitochondria or nuclei instances are eliminated to simulated missing annotations. For visualisation, only a slice of 1024×1024 pixels is shown for the MitoEM-H dataset.

3.6.2 Experiments

In order to validate these findings empirically, I compare the effect of noisy annotations for mitochondria segmentation in EM data and a nucleus segmentation task. For the mitochondria, I use the MitoEM-H dataset (Wei et al., 2020a) which was already mentioned in section 2.3.2 and shows mitochondria in 3D human tissue with a resolution of $30 \times 8 \times 8$nm. I use a block of $100 \times 4096 \times 4096$ for testing and a separate block of size $400 \times 4096 \times 4096$ for training of which 20% are used as a validation set.

The nuclei segmentation dataset is taken from the MoNuSeg challenge (Kumar et al., 2017, 2020) and consists of H&E stained tissue from the cancer genome atlas archive (Weinstein et al., 2013). The data includes 38 RGB training images with



(a) MitoEM-H **(b)** MoNuSeg

Figure 3.12: Comparison of the test IoU of models trained with Dice and CE with different noise levels shows the robustness of the Dice loss to false negatives. The error bars show the calculated standard deviation over 3 and 5 runs for (a) and (b) respectively.

a size of 1000×1000 pixel and roughly 22,000 annotated nuclei and a separate test set of 15 images of the same size.

I train a standard U-Net (Ronneberger et al., 2015) in its 3D version for the mitochondria segmentation and in 2D for the H&E images to predict the segmentation mask using the nnUNet framework (Isensee et al., 2024) and compare the Dice loss to the cross-entropy loss. In order to model an annotator missing some nuclei or mitochondria, I delete a certain noise level η of the individual instances on each image in the training set and investigate the effects on the model prediction on the unchanged test images when using the different loss functions. This implies that I only introduce false negatives completely at random in the annotation but no false positives. Figure 3.11 illustrates the datasets and the modified labels for the example of $\eta = 0.4$.

Figure 3.12 shows the model performance in terms of IoU against a varying fraction η of corrupted labels for different loss functions for the MitoEM-H dataset (a) and MoNuSeg (b). For the MitoEM-H data three and for MoNuSeg five independent models are trained and the mean and the estimated standard deviation are shown. Figure 3.12 shows that the Dice loss is surprisingly robust even under a high fraction of false negatives as even when deleting 60% of the original mitochondria instances, the performance on an unseen test set suffers only very mildly. For the MoNuSeg

dataset, the decrease in performance is slightly higher probably due to a larger variability in the background of the slices but the Dice loss still always performs better than the CE loss. This confirms the theoretical arguments above.

Conclusion

In this section, I explained that the Dice loss performs superior to the CE loss in segmentation settings in which the model is confronted with incomplete annotations, missing some of the objects of interest. I theoretically explained this behaviour by analysing the gradients that the Dice loss induces for the logits and showed that they are less susceptible to misleading labels. Moreover, I empirically validated these results on mitochondria segmentation in EM data and nuclei segmentation in H&E images. These insights provide valuable guidance on the loss function for the training of model in the next sections, especially for the processing of noisy immunolabeling in Section 4.4.

One limitation of the presented experiments are that I eliminate mitochondria and nuclei instances completely at random. In a real life situation, in which an annotator would be missing some annotation, the missing instances would probably be correlated with some features or morphological characteristics that lead the annotator to miss them. However, this limitation might not be present in the case of sparse anti-body labelling.

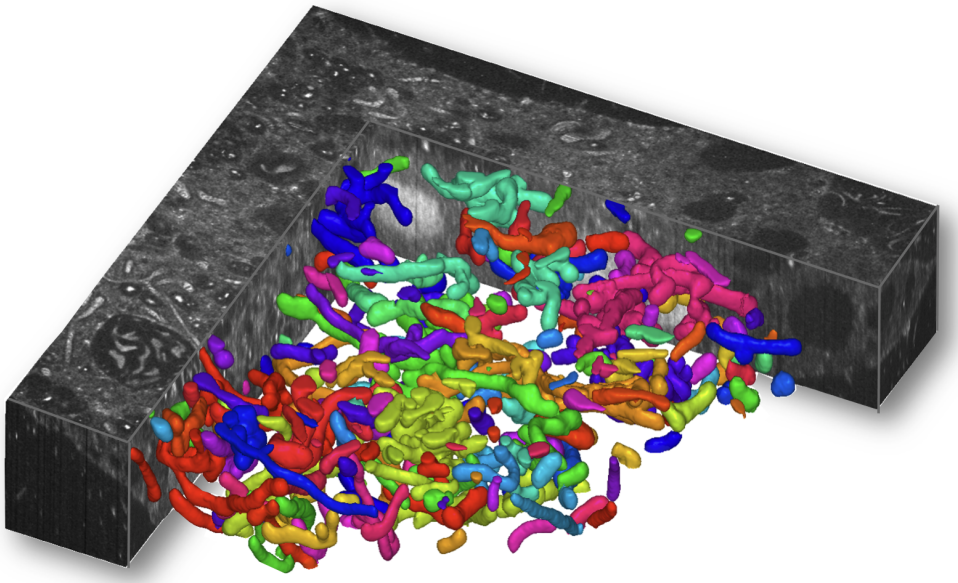
3.7 Discussion

In this chapter, I discussed the pan-stained ExM data and showed that special attention has to be paid to multiple annotators interpreting the data in different ways. To alleviate the problem in the following analysis, I reviewed all annotations and adjusted them to follow a more unified interpretation and annotation style. However, while this mitigated some issues around varying interpretations of the ambiguous class, several segmentation masks are usually acceptable for the same image as long as the underlying biological constraints are respected. Consequently, some differences in the annotation style remain. I therefore account for the ambiguity

of the segmentation in Chapter 4 when evaluating the model and develop a new reconstruction in Section 5.2 to improve the biological consistency of the segmentations.

Moreover in this chapter, I developed an approach that allows to boost the performance of semi-supervised segmentation approaches in the presence of multiple annotations. While most previous work is focused on estimating the one single ground truth, my work tries to extract the additional information about the underlying data uncertainty from conflicting annotations and uses this to guide a pseudo-labelling approach. However, with a fixed annotation budget I still find in Chapter 4 that it is more efficient to spread the annotations over multiple mitochondria compared to re-annotating the same instance. Therefore, the following chapters focuses on samples with a single annotation.

Finally, I extended the theoretical analysis of the dice loss which was not previously available in the literature. Specifically, I show that the induced gradients are favourable compared to the CE loss in the case of annotations with a higher fraction of false negatives. This insight is used in Section 4.4 when basing the segmentation on noisy immunolabeling.



4

Mitochondria Segmentation

Contribution

This chapter demonstrates how additional dyes or immunolabeling can be utilised to decrease the annotation effort required to segment mitochondrial and their ultrastructure based on the pan-staining. Firstly, I present a mask based approach based on the MitoTracker dye in HeLa cells and describe the benefit of appropriate data augmentation and ensembling on the performance. Moreover, I demonstrate how the model can be adopted to investigate the effect of several drug treatments on the mitochondria morphology and present several metrics for the comparison. Finally, I extend the framework to kidney tissue and provide an novel algorithm which allows to train a mitochondria outline model based purely on sparse and noisy antibodies before this model is fine-tuned to segment the inner structure and detect lamellar cristae.

Contents

4.1 Introduction 60
4.2 MitoTracker as a Shape Prior 61
4.2.1 Mask Based Approach 63
4.2.2 Architecture 65
4.2.3 Pre-processing and Data Augmentation 66
4.2.4 Results 68
4.3 Drug Treated Cells 74
4.3.1 Drug Inspection 74
4.3.2 Drug-Conditional Model 77
4.3.3 Experimental Results 80
4.4 Mitochondria in Tissue 83
4.4.1 Kidney Tissue Data 85
4.4.2 Training with Sparse Antibodies 87
4.4.3 Results 93
4.5 Antibody Requirements for Future Applications 93
4.6 Discussion 95

4.1 Introduction

While modern microscopy techniques like panExM allow to image a large numbers of cells, analysing these large amounts of data poses a challenge as manual annotation of 3D data is extremely time-consuming. With the advent of deep learning (DL), computer vision algorithms have achieved or surpassed human level performance on several benchmarks over the last few years (Alzubaidi et al., 2021b).

One of the challenges associated with deep learning is that its success is based on large amounts of labelled data. While traditional machine learning methods are based on hand crafted features, DL is able to utilise its training data to construct more expressive and robust intermediate features itself during the training process which leads to superior performance as discussed in Section 2.2.

However, these web-scale sized datasets used for pre-training by current state-of-the-art models are only available for natural 2D images which induces a bias for an image distribution which is very different compared to microscopy images as microscopy images often exhibit a substantial higher noise to signal ratio than natural images and lack clear boundaries and colour contrasts. Therefore, it has been repeatedly shown that the pre-training on natural images does not necessarily transfer favourably to microscopy or other images from the medical domain (Alzubaidi et al., 2021a). On top of this, the 3D structure of the microscopy images poses challenge for re-using 2D pre-trained natural images and finally, it has been shown that pre-training for image classification often cannot provide the same impressive boost for downstream task like pixel-wise segmentation (He et al., 2019).

In order to replicate the success of deep learning on natural images for microscopy data without excessive manual annotation, it is therefore necessary to be able to pre-train a 3D model on a large amount of microscopy data without any human annotation. While adding protein-specific staining to electron microscopy in the context of CLEM is a complex task, it is natural and easily achieved to use additional dyes or antibodies in combination with the global staining provided by NHS ester in the context of panExM.

Using the example of mitochondria segmentation, I show in this chapter that having access to the additional staining allows DL models to be robustly pre-trained such that only a limited amount of annotation data allows the model to achieve human like performance. I demonstrate how the model performance depends on the number of annotated samples and investigate the optimal annotation strategy in terms of achieving a trade-off between annotated different samples and annotating dense slides.

Subsequently, I apply the model to cells treated with different drugs in order to investigate the resulting morphological changes of the mitochondria. I therefore develop a modification to the model architecture which allows to efficiently incorporate information about the drug treatment to maintain a high segmentation quality.

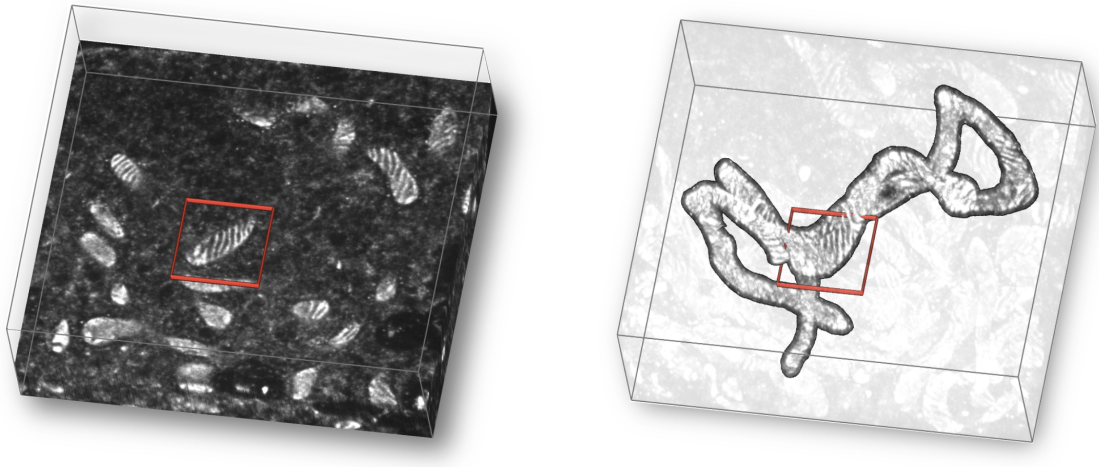
The ultimate goal is to apply panExM and the segmentation pipeline to human patient data to enable the investigation of real diseases. To progress towards this goal, I extend the analysis to mice kidney tissue and demonstrate how the outline model can be trained based on sparse and noisy biomarkers.

Finally, in order to provide some guidance on future application I investigate what quality of the immunolabeling should be sufficient to repeat the analysis in a new setting.

4.2 MitoTracker as a Shape Prior

Expansion microscopy allows to add additional dyes or anti-bodies to the global staining provided by panExM. These stainings provide visual clues about specific organelles that are informative and easily interpreted by the human observer. However, once the staining is investigated more closely on a pixel level a significant amount of ambiguity and variability is observed. This makes using these additional channels in an automated fashion a non-trivial task and I develop several approaches in order to train a neural network on these stainings.

As shown in the literature review, a substantial amount of research has been conducted into the questions of how to include shape priors into segmentation models. However, this work has mostly focused on small variations from a target shape



(a) One 2D slice **(b)** 3D with background removed

Figure 4.1: In 2D, the shape of mitochondria appears simple and textbook-like (left). However, the 3D visualisation (right) of the full mitochondrion reveals the much more complex, real shape. All samples in this thesis showing the pan-staining have been imaged by Yuan Tian at Yale University.

(like shape priors for brain segmentation for MRI scans in which it is reasonable to expect only very small derivations from the average brain (Lebrat et al., 2021)), 2D shapes (Peng et al., 2020) and dominantly for setting in which the task is to segment one single instance per image. These constraints are not satisfied in the current setting as large differences can often be observed in the size and morphology of mitochondria. Moreover, as shown in Figure 4.1 the shape of a mitochondria seen in a 2D slice might seem fairly simple as demonstrated by the mitochondrion highlighted in the red rectangle in Figure 4.1a which could be described as elliptical and convex. However, when considering the same mitochondrion in 3D, it can be seen that the true shape is actually much more complex. This makes capturing the shape by any explicit mathematical formulation very challenging.

Finally, traditional methods which only deal with one instance per image can afford to be computationally intensive and therefore rather slow (for example the algorithm proposed by Guo et al. (2021) takes around 5 minutes to process one instance). In the current case, the model is faced with hundreds of mitochondria per cell and given the goal to segment dozens of cells, an approach is required that is able to scale to thousands of mitochondria.

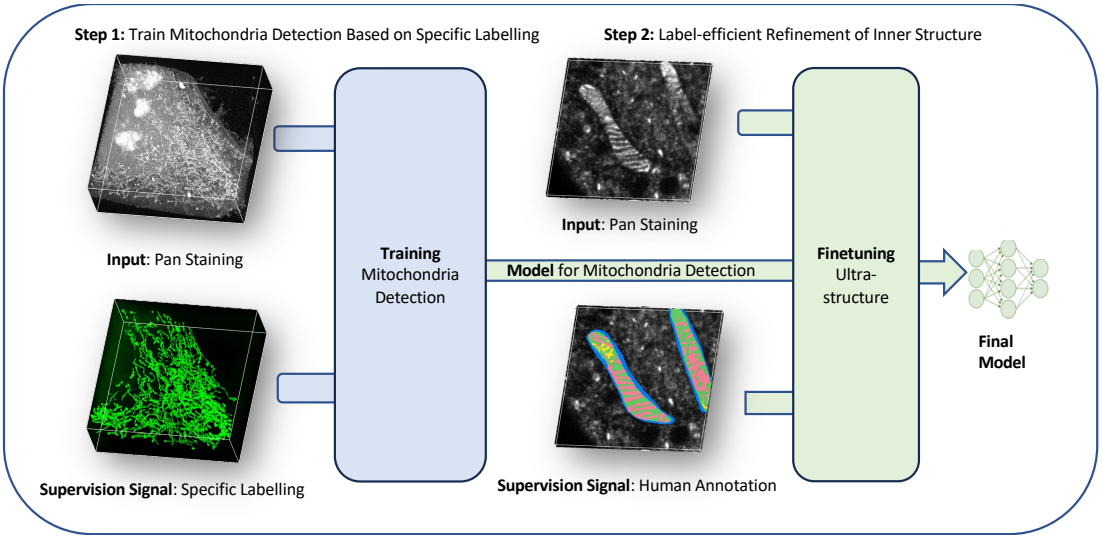


Figure 4.2: Workflow MitoTracker Based Model: In a first step, a mitochondria outline model is trained purely on the MitoTracker. Subsequently, the model is fine-tuned to fill in the ultrastructure of the mitochondria based on additional human annotations.

In the spirit of Sutton (2019), this work instead strives to learn a shape prior for the mitochondria based on the MitoTracker in a completely data-driven fashion. The idea is to use the MitoTracker channel to provide rough masks of the outline of the mitochondria and outlined in 4.2.

This approach allows the model to learn to differentiate between the mitochondria and the background cytoplasm from a large amount of data and provides an efficient 3D pre-training task. The purpose of this approach is that the output of that pre-trained model is reused in order to mask out the background. This allows that the fine-tuned model will only be required to fill in the finer inner-structure segmentation of the mitochondria including a much smaller amount of background and therefore requires a limited amount of manual annotations.

4.2.1 Mask Based Approach

In order to eliminate the noise, I apply and combine several morphological filters to the MitoTracker channel which are applied slice-wise instead of using their 3D equivalents for computational reasons. In order to remove the background noise I employ a median filter with a kernel size of 5 and the result is transformed into a binary mask by thresholding pixels above the 98% quantile. The median filter

was used as it allows to remove some of the unspecific binding while also retaining more of the fine structures and boundaries of the MitoTracker than a Gaussian filter. Finally, an opening operation is performed in which the erosion kernel has a size of 1 while the dilation kernel is of size 3.

I additionally remove all the area which is outside of the cell which is identified based on the NHS staining.

While the procedure described above gives a rough outline of the mitochondria, I also want to preserve some of the inner structure in order to allow the model to learn some of the features associated with the matrix structure in this pre-training step. I therefore apply a simple thresholding to the MitoTracker in order to cast the setting as a three-class segmentation problem between background, outline of the mitochondrion and inner part of it.

In order to find the optimal threshold I utilise the annotated data. The choice of the threshold includes an obvious trade-off: Choosing a very low threshold allows to capture all relevant area but includes a large amount of false-positives, i.e. area which is marked as mitochondrion but actually represents background. A very high threshold on the other hand avoids these false positives but leads to more false negatives, i.e. pixels which should be marked as mitochondrion but are considered background. In order to find the optimal trade-off I apply a specific threshold value t to the MitoTracker and find the number of pixels which are correctly classified as inside of a mitochondrion (as given by the annotation labels 'matrix', 'cristae', 'matrix or cristae'). This percentage is denoted by $p(t)$ and the number of pixels which are falsely identified as 'matrix' by the MitoTracker are denoted $f(t)$. The goal is now to maximise $p(t)$ while minimising $f(t)$ and by choosing to weight them equally decide to maximise $|p(t) - f(t)|$. Given the small amount of relevant values of $t \in [100, 150]$ I adopt a brute force approach.

The mean threshold for the maximal difference over the annotated crops is 125 which is used as a thresholding value in subsequent analyses. However, the optimal threshold values covered a range between 122 and 138 which shows the variability involved in the MitoTracker channel and emphasises one of the reasons why a

simple segmentation approach based on the MitoTracker for the mitochondria is not sufficient.

The final three target masks are shown in Figure 4.3.

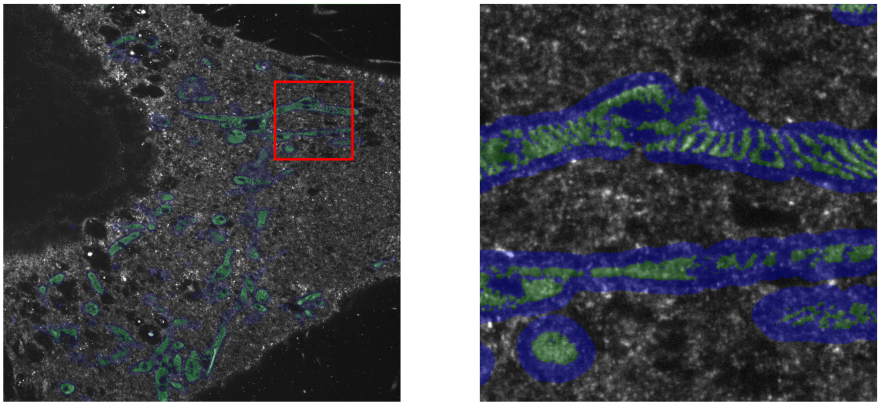


Figure 4.3: Visualisation of the three masks created from the MitoTracker staining with blue showing the larger outline mask and green indicating the matrix of the inner structure. The right Figure shows the red rectangle zoomed in.

4.2.2 Architecture

The model architecture is based on the classical UNet model (Çiçek et al., 2016; Ronneberger et al., 2015) which despite its age is still widely employed with various minor modifications in the medical imaging community (Siddique et al., 2021b) and mitochondria segmentation (Franco-Barranco et al., 2021; Wei et al., 2020b). Since I aim for a method that can be easily adopted by the microscopy community it was important to demonstrate that the proposed method does not depend on a exotic network architecture which require obscure training tricks to squeeze out a couple more percentage points of performance. Moreover, Wang et al. (2024b) find that many claims of new architectures to improve the performance over UNets are not reproducible once carefully examined. I compare the performance to Transformer based architectures like Hatamizadeh et al. (2022b) but find no substantial differences in line with other studies like Wang et al. (2024b) which might be explained as the main advantage of the Transformer based models is to be able to incorporate long range dependencies which does not seem necessary for the task at hand.

One popular feature that is added are residual connections (He et al., 2015). Moreover, I replace the batch normalisation by group normalisation (Wu et al., 2018) and weight standardisation (Qiao et al., 2020). This change was motivated by large differences between the intensity levels of the imaged cells and therefore the value for the mean of the feature activation during the training might be very different compared to a new cell that the model is deployed on. I found that the group normalisation was much more robust to these changes and I address this issue again in Section 4.2.3.

The decoder of the network is shared between the task of mitochondria identification and inner segmentation while separate encoders are used for each task.

4.2.3 Pre-processing and Data Augmentation

The data is recorded with 16 bits, taking values between 0 and 65536 and is highly skewed with the outliers driven by very few very bright pixels. Moreover, the distribution of the pixel intensities varies substantially between different imaging stacks which poses a challenge for transforming the pixel intensities into $[0, 1]$ as common in machine learning applications.

These changes in intensities are mainly driven by three factors: Firstly, the presence and size of the nucleus in the image causes the mean pixel intensity to be significantly higher. Moreover, the amount of nucleoli which appears very bright as seen in Figure 4.4 varies between cells and alters the shape of the tails of the pixel distribution. Secondly, the amount of background varies between stacks with more background leading to a large shift to the right in the corresponding histograms. Finally, the NHS staining itself varies in its brightness due to differences in the sample and imaging settings.

In order for the model to be able to deal with this variety of input distributions, several forms of data augmentation are applied. Firstly, in order to tackle the problem of large outliers, the pixel intensity are clipped at the top q percentile and subsequently normalise the values to $[0, 1]$. To increase the variety of data the

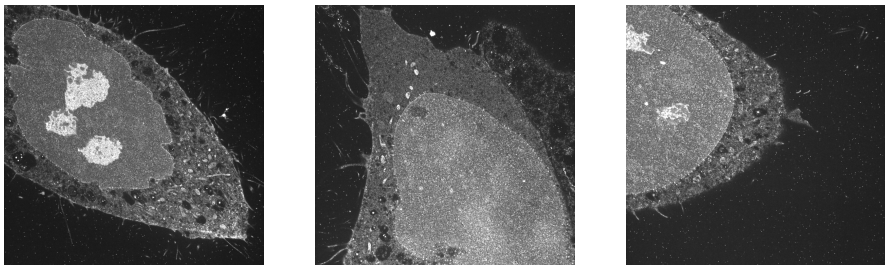


Figure 4.4: The left figure shows a nucleus with a large quantity of nucleoli which appear very bright. However, these nucleoli do not always appear as shown by the middle figure as our stack may not include the whole nucleus. Finally, the right figure shows that sometimes the amount of background is substantially larger.

model has seen, q is randomly drawn from $[0.985, 0.9875, 0.99, 0.9925, 0.995, 0.9975]$ at every training iteration to account for different amounts of nucleoli.

However, this operation only squeeze the distribution but has no effect on the shape of the histogram itself. Therefore, I additionally transform the pixel values by $h_{\alpha,\beta} : [0,1] \to [0,1], x \mapsto (x^\alpha + (1-(1-x)^\beta))/2$ where the parameters are randomly sampled from $\alpha, \beta \in \{\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, 4, 5\}$. The results are visualised in Figure 4.5. In order to prevent the augmentation to be too aggressive, augmentation attempts that would use $\alpha > 2$ or $\beta < 0.8$ are disregarded. On top of this, I apply the standard augmentation of randomly flipping the image along the x, y or z axis.

Training Procedure

The training consists of two steps: The first step uses the MitoTracker-based masks as targets and train the network until it converges using the Adam optimizer and a starting learning rate of $3 \cdot 10^{-4}$ which decreases by a factor of ten if the model reaches a plateau for two times before stopping the training.

In the second step, the decoder is kept and the weights are frozen and while a new decoder is attached to the model. This training step only involves the annotated crops which are described in Chapter 3.

I keep five crops as a validation set and five crops as a test set. In order to allow the new decoder to only focus on the inner parts of the mitochondria, all pixels in the loss function which are not part of the MitoTracker-based masks are ignored.

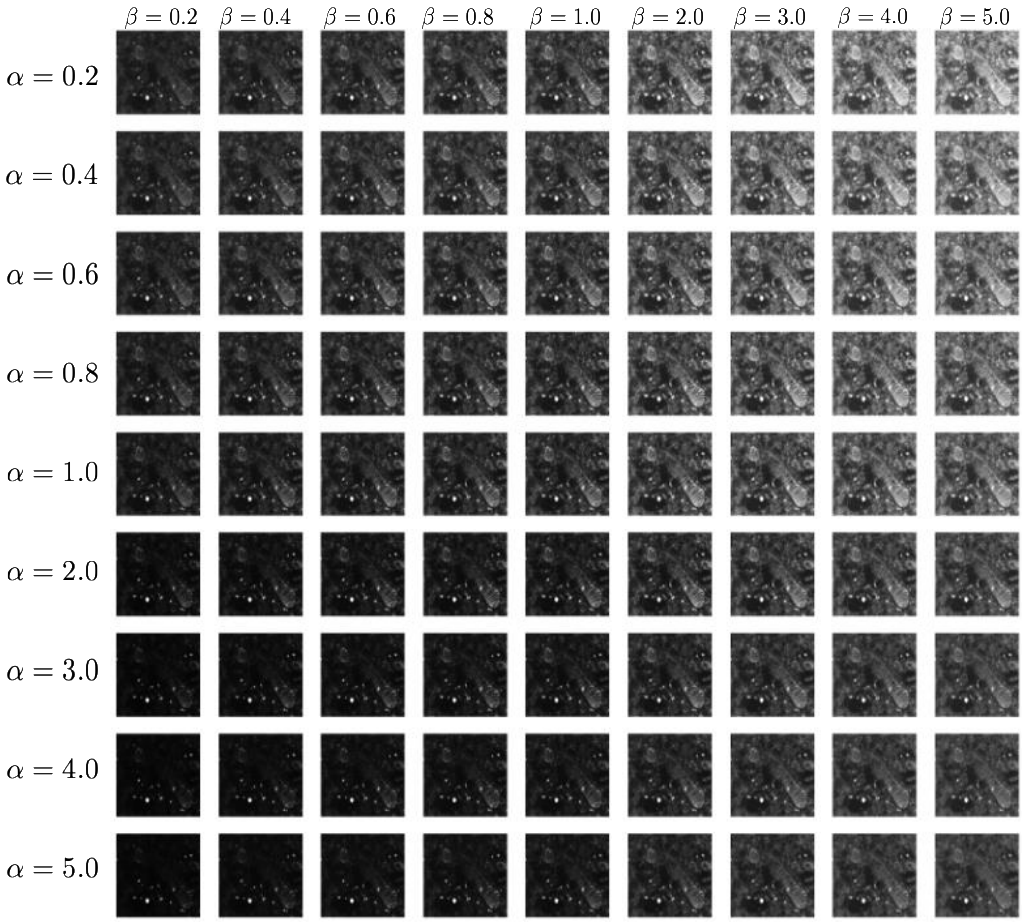


Figure 4.5: Visualising the effects of the data augmentation h for different values of the parameters α and β. The original image is shown at the centre for $\alpha = \beta = 1$. In order to prevent the augmentation to be too aggressive, $\alpha > 2$ or $\beta < 0.8$ are not used.

For both steps, I use the Dice loss as I found it to work better than cross entropy or a mix of both and since this choice is supported by the analysis in Section 3.6.

4.2.4 Results

Comparison of training data

In this section, I demonstrate the benefits of the pretraining based on the MitoTracker for the segmentation of the inner structure of the mitochondria.

It is noteworthy that when it comes to actually segmenting the full cell, these results actually underestimate the benefit of the MitoTracker pre-training as these results assume the right outline of the mitochondria is already available.

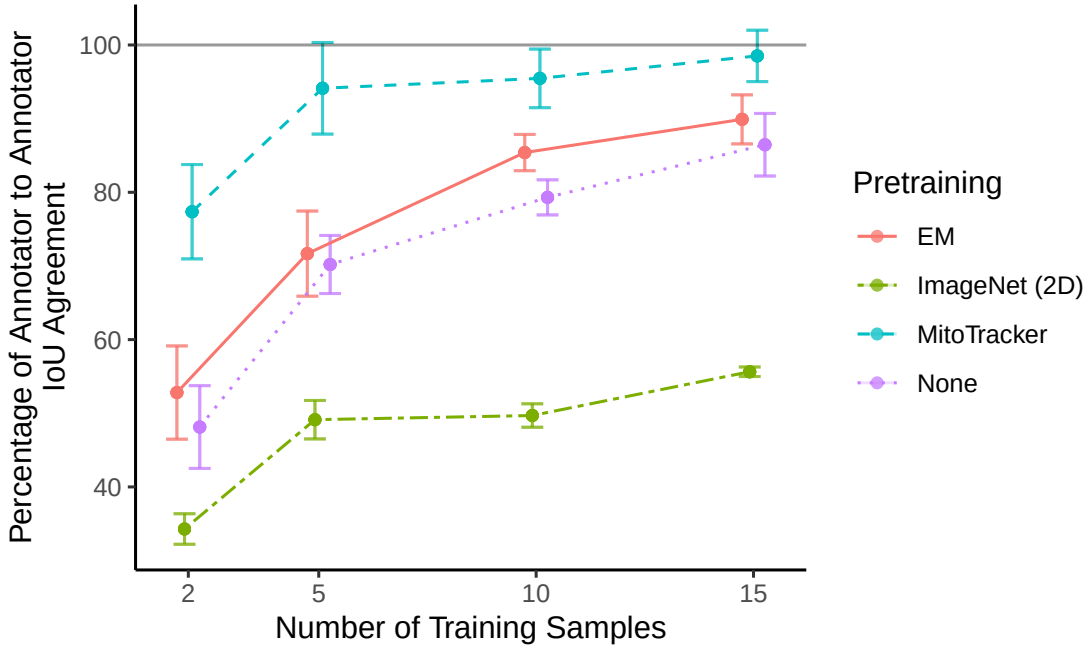
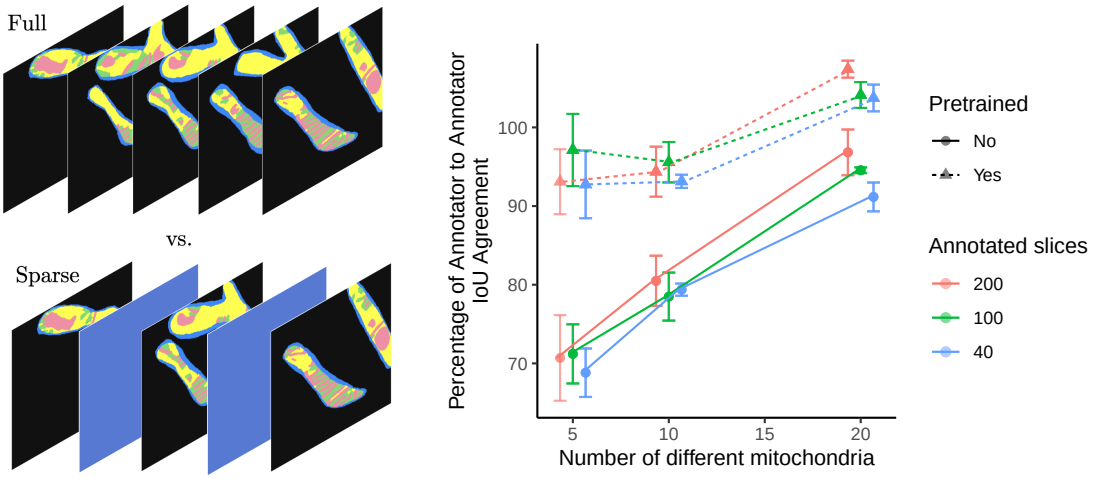


Figure 4.6: Results annotation efficiency: Using the pretraining approach based on target specific stainings like MitoTracker allows to reduce the necessary amount of annotations to achieve human like segmentation performance. Performance measured in relation to the inter-annotator agreement discussed in Chapter 3.

In Figure 4.6, I compare the results using my pretraining to several alternatives. For an applied researcher like a biologist who considers the proposed method, it is especially important to know how much annotation effort is required for the model to perform well. I therefore simulate a setting with various amounts of available annotated samples. In order to produce more robust results, I repeat this procedure three times with different training samples for each method. The individual runs are depicted as fainted out points while the mean and error bars for the standard deviation are shown in bold in Figure 4.6.

The most popular approach in classical computer vision is to use a model pre-trained on the ImageNet dataset (He et al., 2019). However, this approach has to bridge a large shift in the data distribution and cannot access the 3D nature of our data and results therefore in very poor performance. An alternative approach is to use some of the available 3D electron microscopy datasets that provide ground truth masks for the mitochondria (Lucchi et al., 2012). I first train the model on



(a) Visualisation of sparse annotations in blue.

(b) Showing the effect of sparse labelling and the MitoTracker pre-training.

Figure 4.7: Using the MitoTracker pre-training allows increases the performance on sparse slices substantially. Moreover, annotating a large variety of mitochondria is more important than dense annotations. Performance measured in relation to the inter-annotator agreement discussed in Chapter 3.

this dataset and use the resulting weights as the initialisation for the network before training on our data. While this approach allows the model learn from images of the same domain it still has to cope with the change in the appearance of the mitochondria due to the different imaging modality. However, Figure 4.6 shows only has a mild positive benefit if there is little data available but eventually even harms the performance compared to training from random initialisation.

When I finally consider the pre-training with the MitoTracker for target specific staining, a large boost in the performance for all tested amounts of available annotations can be observed. Most importantly, even with a small number of annotated samples, the model is able to reach almost human like performance.

I furthermore hypothesis that the MitoTracker pre-training allows the model to learn a suitable representation for the 3D structure of the data. If true, this would allow to save even more annotation time by not annotating a sample densely but only every kth slice as visualised in Figure 4.7(a).

In order to simulate this situation I allow a slice budget of B slices. Moreover, it is of interest to find out whether it is more efficient to spread these B slices over a few or many different annotation stacks. If considering K different samples and a

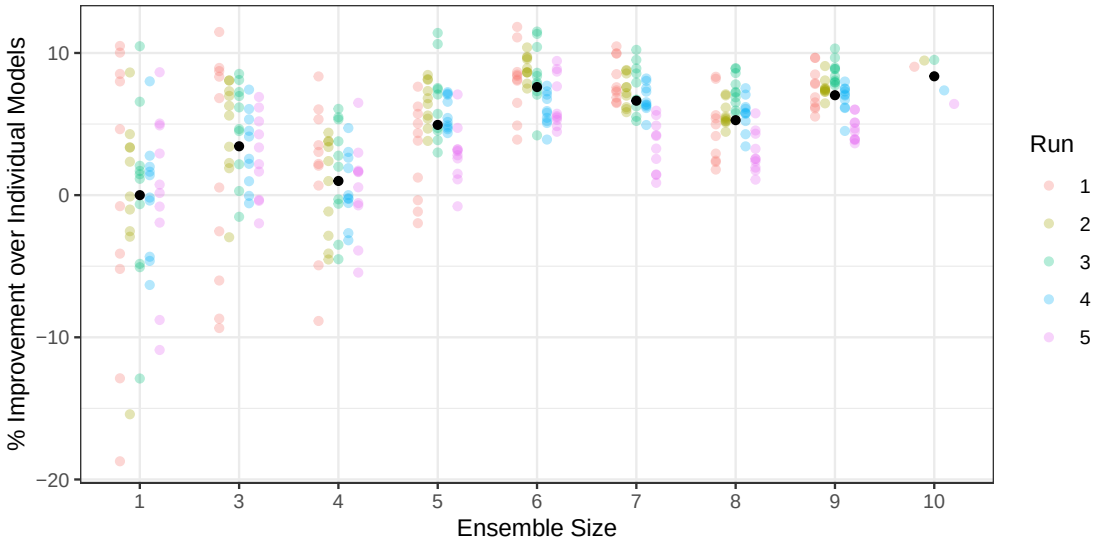


Figure 4.8: Building ensembles of models trained on different training/validation splits improves the performance over individual models with the best performance at an ensemble size of six.

slice budget of B, I evenly sample B/K annotated slices from each of the K stacks as far apart as possible and ignore the remaining slices in the loss function.

I now compare the effect of the MitoTracker pre-training and different values of $B \in \{40, 100, 200\}$ and values of $K \in \{5, 10, 20\}$ with the results are shown in Figure 4.7(b). It can be seen that the benefits observed in Figure 4.6 do not come from more annotated slices but rather from more annotated mitochondria. This results is in line with the high inter-slice dependency as an adjacent slices does not add a lot information for the model but can rather be interpreted as a very mild form of data augmentation. An additional stack on the other side potentially presents the model with a completely different mitochondria morphology that allows it to perform better on unseen test samples This insight is going to be utilised in Section 4.3.

Ensemble

During the experiments, it was noticed that the model predictions changed substantially across different test-validation dataset splits. I hypothesise that this is due to the high inter-annotator variability described in Chapter 3 and depending on the composition of the training set, the model learns different annotation styles.

It is a well known technique in machine learning that building ensembles of models often leads to better predictions than their individual components (Sagi et al., 2018). However, for this technique to fulfil its potential it is important that the individual components of the ensemble are in general unbiased but still show some variation (Dietterich, 2000). This requirement is fulfilled in the present setting.

In order to test this hypothesis and investigate the optimal size of the ensemble, I test the performance for different ensemble sizes. To establish a baseline, I train ten individual models by randomly selecting 20% of the training data as the validation set that are used for early stopping. For each possible ensemble size ranging between three and ten, I build ten ensembles by composing the training and validation set. Moreover, I repeat the procedure five times using a different splits between the training/validation and the test set with the results shown in Figure 4.8.

In order to compare the runs over different test sets, I compute the mean performance for the individual models in terms of the IoU and then normalise all further results with respect to that baseline to show the relative change. The ten random ensembles per run and ensemble size are coloured according to the run they represent and the mean is depicted in black. The results confirm that employing ensembles of models is highly effective in the current setting and an ensemble size of six seems to represent the optimal size which allows to increase the performance by around 8%. I suspect that the drop in performance for an ensemble size of four is due to inconsistencies introduced due to tie-breaking with between a lower number of models. In the following analysis, I therefore employ an ensemble of six models.

Evaluation

The large amount of ambiguity in the interpretation of the NHS data as discussed in Chapter 3 makes the evaluation of the final model challenging as several segmentations are equally acceptable. This has been seen in low values for the IoU with regard to the test set but visually very convincing segmentations as in Figure 4.9.

To evaluate the model nevertheless, I randomly pick 30 mitochondria and manually revise the model predictions, changing them if they are judged it to

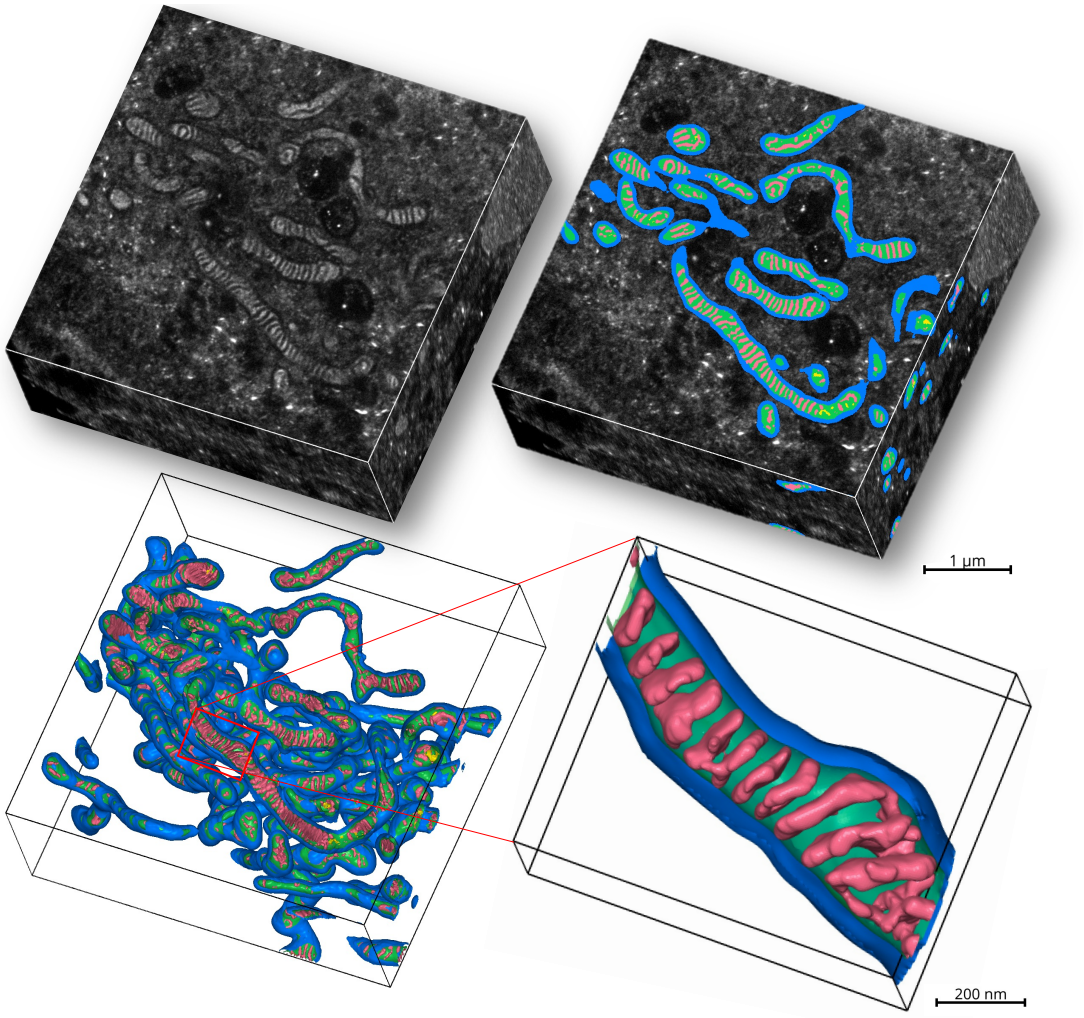


Figure 4.9: Visualisation of the 3D segmentation and reconstruction of the mitochondria ultrastructure. The first row shows a pan-stained block (left) and the overlay with the segmentation mask (right). The second row shows a 3D rendering of the segmentation of the full patch (left) and a zoomed in version (right).

r

Table 4.1: The high IoU values between the model prediction and a version revised by a human annotator for the different components of 30 randomly selected mitochondria show that the model ensemble produces very reliable predictions.

Metric	Intermem. Space	Cristae	Matrix	Ambiguous
IoU	0.981	0.953	0.988	0.968
(Standard error)	(0.036)	(0.087)	(0.026)	(0.083)
Dice	0.993	0.989	0.993	0.961
(Standard error)	(0.009)	(0.022)	(0.011)	(0.179)
Volume Change (%)	-0.1737	1.4525	0.4787	-2.2325
(Standard error)	(0.198)	(0.877)	(0.309)	(0.504)
HD95	2.858	2.924	2.328	2.705
(Standard Error)	(2.7508)	(3.3544)	(2.5848)	(3.9819)

be unacceptable. I then compare the original model prediction with the revision produced by the human annotator.

The problem of choosing an appropriate evaluation metric has recently received increased attention (Maier-Hein et al., 2024; Reinke et al., 2021, 2024) and it can be concluded that one metric is in most cases not sufficient. To evaluate the segmentation thoroughly, I therefore combine metrics that focus on different aspects: The IoU and Dice score are standard metrics to evaluate segmentations but can be misleading for complex shapes and objects of different sizes (Reinke et al., 2024). The 95th quantile Hausdorff distance on the other hand focuses on the localisation of the boundary and is measured in voxels. Finally, the proportion of different components within the mitochondria is often of interest. Therefore, I also include the change in the volume between the predicted and refined volumes to capture this aspect regardless of the precise localisation.

The results can be found in Table 4.1 separately for the different components and show that model produces very high quality segmentation masks which is also shown by the average IoU over all components of 0.98.

4.3 Drug Treated Cells

This section aims to segment a population of cells which have been treated with different drugs which are known to change the morphology and inner structure of the mitochondria. The goal of this section is to demonstrate how to adopt the presented algorithm to this new setting and continue to produce high quality segmentation masks with a limited amount of annotation effort. This enables the quantification of the change of the mitochondria on a large scale.

4.3.1 Drug Inspection

The drugs investigated are Antimycin A, HBSS (standing for Hank's Balanced Salt Solution and leads to nutrient starvation), Oligomycin and DMSO (Dimethyl Sulfoxide) which acts as the control group.

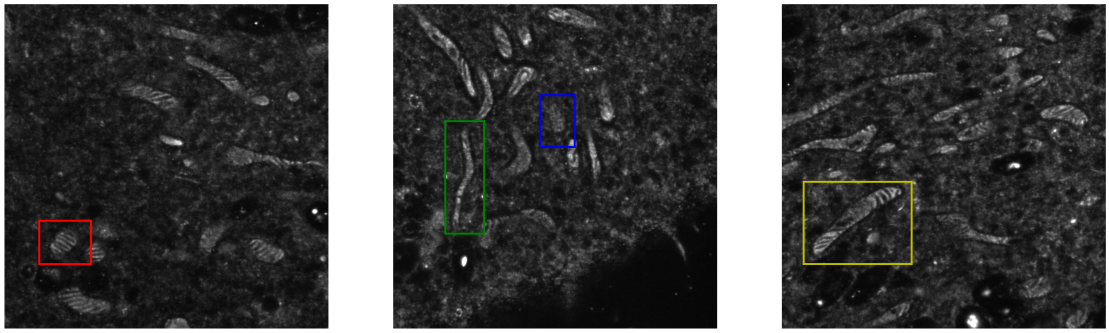


Figure 4.10: Examples of mitochondria from the control cells treated with DMSO. The mitochondria show a health range of shapes and inner texture which are highlighted by rectangles of different colour. All samples in this thesis showing the pan-staining have been imaged by Yuan Tian at Yale University.

Inspecting the mitochondria from the control DMSO cells in Figure 4.10, a variety in shapes can be observed in the inner structure of the mitochondria. Some mitochondria (as dominant in the leftmost figure), have an elliptical shape and show a clearly visible stripped pattern inside like in the red rectangle. However, for some mitochondria the inner structure is much less clear and they blend in more with the background as shown by the blue example. This variability can also appear within a mitochondrion as shown by the yellow example in which we see both the clear cristae pattern and area which a much more homogeneous texture. Moreover, the yellow example shows a dark hole within the mitochondrion. Preliminary experiments by our collaborator indicate that this area is occupied by the mitochondria DNA and the topic is discussed in more detail in Section 4.6. Finally, the green example shows that some mitochondria appear slimmer and more elongated.

In the subsequent paragraphs, I compare the visual appearance of the mitochondria after the treatment by different drugs.

Antimycin A is an inhibitor of the mitochondrial electron transport chain (ETC). This inhibitation impairs mitochondrial respiration, disrupts the proton gradient and interferes with cellular energy production and leads to a decrease in ATP production (Ma et al., 2011). This can lead to oxidative stress and cellular damage.

Figure 4.11 shows several crops of mitochondria from cells treated with Antimycin A. However, no consistent change in the appearance of the mitochondria compared

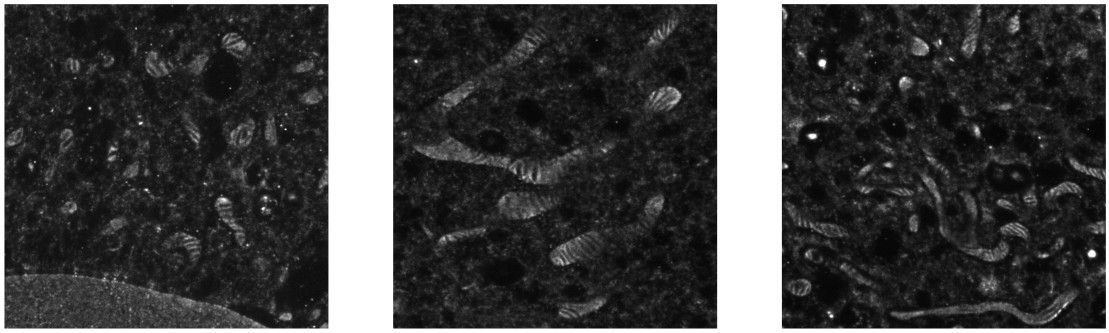


Figure 4.11: Examples of mitochondria treated with Antimycin A. No clear change in the mitochondria is detectable by visual inspection.

to the control group could be detected by eye.

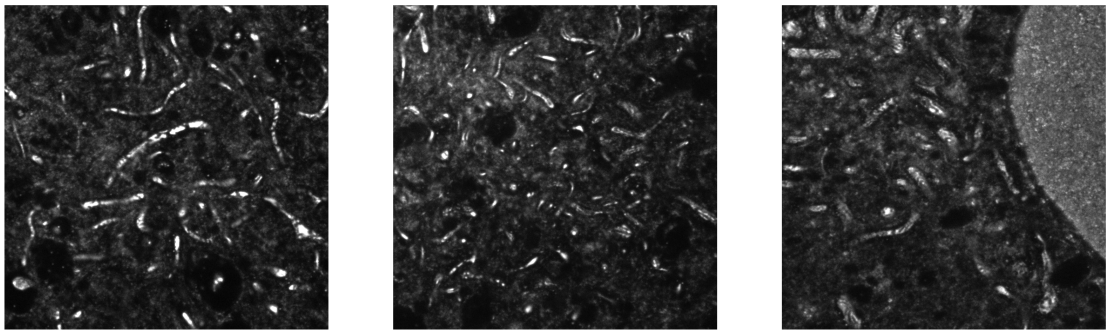


Figure 4.12: Examples of mitochondria treated with HBSS. A large proportion of the mitochondria shrink and appear a lot thinner with less inner structure visible. However, not all cells and mitochondria are effected by this change as shown by the rightmost image.

Figure 4.12 shows different examples of mitochondria from cells treated with HBSS which is used to induce stress and nutrient starvation in the cells (Desideri et al., 2012; Wu et al., 2013).

For this treatment the effect on the appearance of the mitochondria is visually striking: A large proportion of the mitochondria appear much thinner and elongated. This reduction in the diameter seems to be associated with a reduction in the cristae appearance. For some mitochondria the change is rather extreme as shown by the centre figure. However, not all mitochondria and cells are effected equality and the rightmost figure shows an example where the change is rather mild.

Finally, Figure 4.13 shows examples of the changes in mitochondria observed in cells treated with Oligomycin which is supposed to introduce mitochondria fission

and blocks the ATP synthase (Cover et al., 1993). Some of the mitochondria often appear larger than in DMSO as shown in the rightmost figure. However, the inner structure is often not clearly visible. Again, these changes are highly variable as shown by the centre figure which does not indicate any enlargement of the mitochondria. Moreover, the rightmost figure also shows that the disappearance of the inner structure not necessarily occurs.

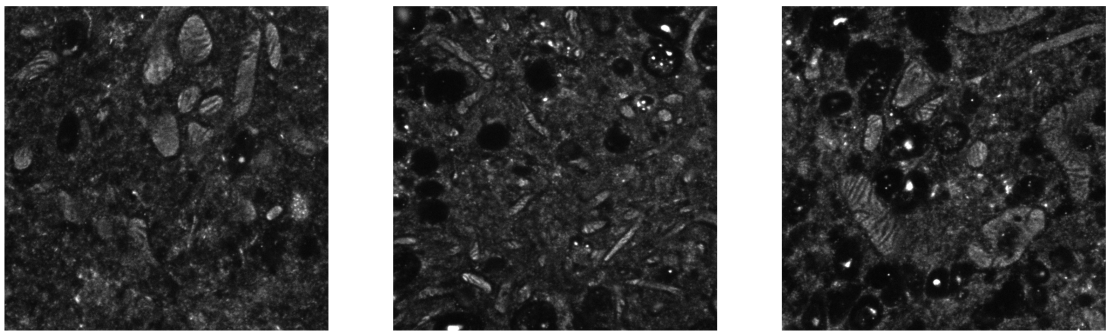


Figure 4.13: Examples of mitochondria treated with Oligomycin. Visually the mitochondria seems to be more separated.

This variability both between mitochondria within the same cell but also between different cells makes it very challenging to draw reliable conclusions from a purely visual inspection of the cells. However, in order to derive quantitative metrics of the changes, I first need to segment the mitochondria and their inner structure. In the next section, I develop an approach to adopt the model to the new data in an annotation efficient way.

4.3.2 Drug-Conditional Model

The changes in the appearance of the mitochondria described in the previous section pose a challenge for the segmentation problem since the altered mitochondria represent a substantial shift in the data distribution.

While there exists a large literature on domain adaptation (Franco-Barranco et al., 2022; Toldo et al., 2020), these methods provide a model increase in model performance but fall short of reaching supervised results. However, the subsequent analysis depends on a high quality segmentation to reliably extract the parameter of interest and compare them to the baseline.

Building on the insight from the previous Chapter 4 that the pretraining with the MitoTracker allows us to efficiently train the model from sparse annotations, we annotate a 36 Antimycin slices, 28 Oligomycin slices and 65 HBSS slices distributed over 5 Antimycin crops, 4 Oligomycin crops and 9 HBSS crops. The larger number of HBSS annotations was motivated by the more substantial change in morphology of the mitochondria.

In the next section, I develop a method inspired by hierarchical model to efficiently learn to segment the new data.

Method

As discussed above, the drug treatment does not introduce completely new shapes but rather redistributes the proportion of mitochondria with different shapes. For example while Figure 4.12 showed that HBSS leads the mitochondria to be much thinner, the example marked by the green rectangle in Figure 4.10 shows that these type of mitochondria are not unseen in the control group. Motivated by this observation, I hypothesise that is it beneficial to jointly learn from all available data to build the most relevant internal features but also enable the model at the same time to reweigh these internal features according to the drug treatment.

In order to implement the intuition outlined above, I argue that the model should not only take the NHS images as an input, but also have access to information about the drug that the current sample was treated with.

As discussed in section 2.2, a convolutional network works by transforming the current internal state $X \in \mathbb{R}^{C \times H \times W}$ where H and W represent the current spatial resolution and C denotes the number of channels which are interpreted here as the current features. The standard block B of a convnet starts by convoluting X with a kernel $W \in \mathbb{R}^{k_x \times k_y \times C \times C'}$ into a new representation with C' features. Subsequently, a normalisation layer is applied to stabilise the training procedure and an activation function introduces some non-linearity as shown in Figure 4.14.

In order to allow the model access to the drug treatment information, I propose to include a learnable embedding z_d for each drug d in the model. Every block

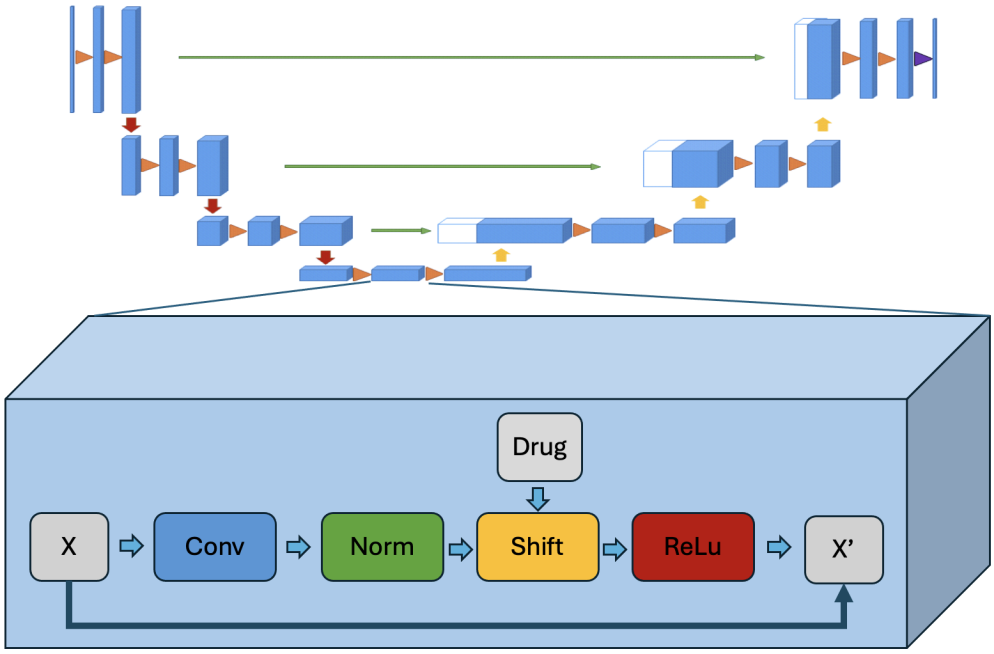


Figure 4.14: Drug Embedding: Each layer in the U-Net is modified by introducing a drug specific feature reweighting module that takes the current drug treatment as an input. The illustration of the 3D U-Net is adopted from Çiçek et al. (2016).

B in the UNet-like architecture (both in the encoder branch or in the decoder in which the convolution is replaced with a transposed convolution) is augmented by a shift module $S(x, d)$ between the normalisation and the activation function as shown in Figure 4.14.

This block $S(x, d)$ applies a transformation in form of a one-layer MLP f_θ parametrised by some learned θ to the drug embedding z resulting in two arrays of size C that match the number of dimension C of the current feature map X at that layer. The first array $f_\theta^{scale}(z_d) \in \mathbb{R}^C$ is used to scale the features, while the second array $f_\theta^{shift}(z_d) \in \mathbb{R}^C$ allows for an additive shift, resulting in an affine transformation. The motivation behind this transformation is to allow the module to learn a drug-specific reweighting and boosting or dampening of the features to adjust for different morphologies. Therefore an intermediate feature $X_c \in \mathbb{R}^{H \times W}$ is transformed into

$$X'_c = S(X_c, d) = \left(1 + f_\theta^{scale}(z_d)_c\right) \cdot X_c + f_\theta^{shift}(z_d)_c.$$

At the beginning of the training, I initialise $f_\theta^{scale} = f_\theta^{shift} = 0$ so that $S(x, d)$ starts out as the identity. Therefore, the model starts out as unconditional and but is subsequently able to learn a drug-specific reweighting of the features similar to hierarchical models in the statistics literature (Hox et al., 2017). ci Following the interpretation of Sandler et al. (2018) that the role of the ReLu activation function is to shut down the information in certain features, it is natural to place the shift module before the non-activation to allow the architecture to boost the influence of some features while minimising others.

4.3.3 Experimental Results

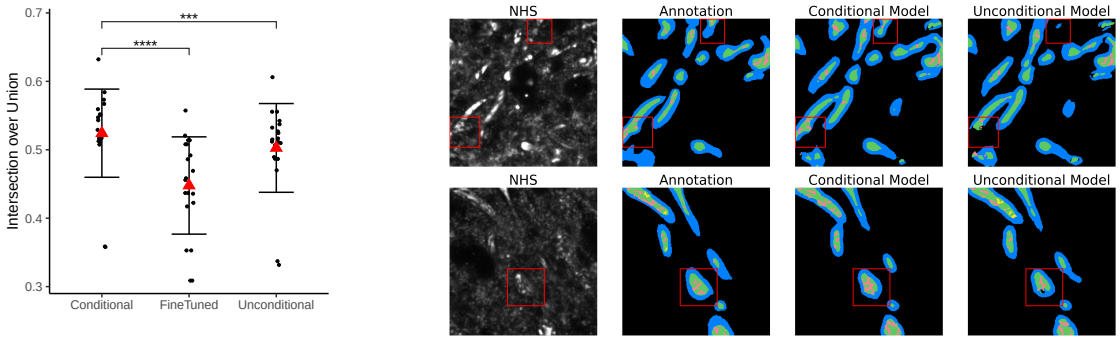
We evaluate the model using a five fold cross-validation technique and compare it to simply fine-tuning the model trained on the control group with the drug data and retraining the whole model from scratch using all data but no conditional information about the drug that the sample was treated with.

The results are shown in Figure 4.15a in which we compare the results in terms of the IoU. As it can be seen, just fine-tuning the model to the new data performs the worst. Judging from Figure 4.15, the benefit of the conditional embedding is small but significant with respect to a paired Wilcoxon Test. However, as already discussed in Chapter 4 the standard metrics like the IoU often fail to capture the true quality of a segmentation mask. Therefore, I also provide a visual comparison in Figure 4.15b in which in becomes clear that the conditional model can often avoid some of the artefacts or missed segments highlighted in the red rectangles.

Moreover, in Figure 4.16 I show how using the wrong drug embedding as a model input results in an increased number of segmentation artefacts and segmentation masks of lower quality.

Model validation

Figure 4.17 shows the segmentation masks of the crops presented in Figures 4.10, 4.11, 4.12, 4.13 and demonstrates qualitatively that the model accurately captures



(a) Quantitative comparison **(b)** Qualitative comparison

Figure 4.15: Evaluation of Conditional model: The conditional model increases the performance quantitatively and avoids some of the segmentation artefacts shown in red. Significance results based on a paired, Wilcoxon Test. ***P<0.001, ****P<0.0001

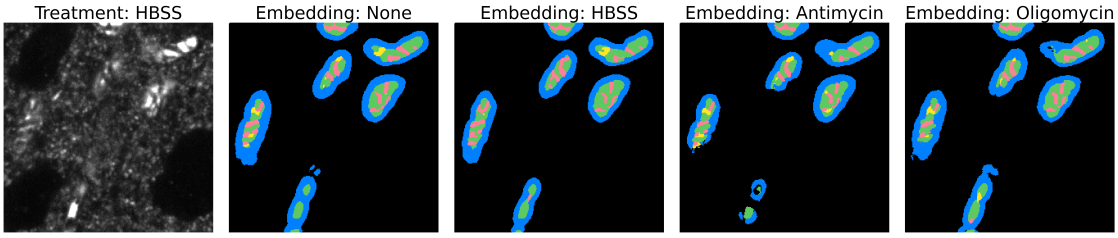
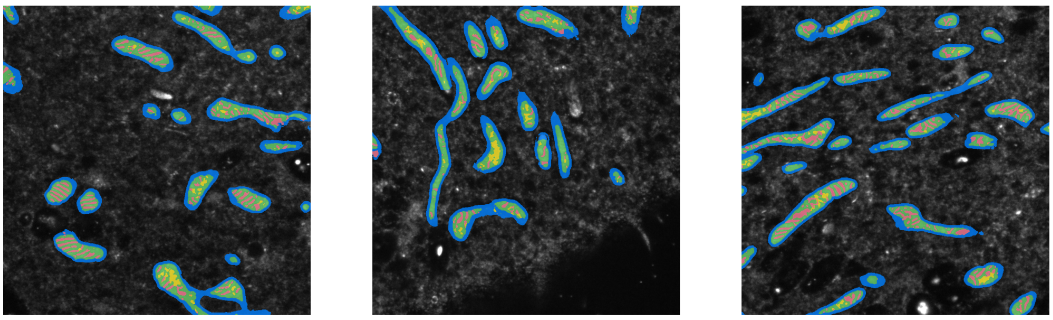


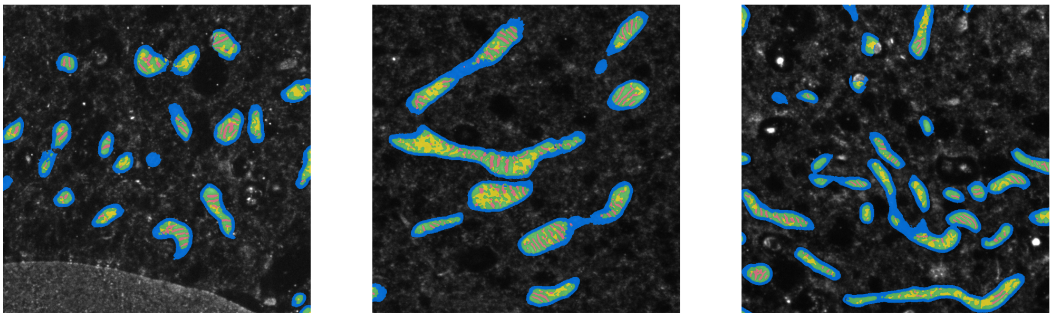
Figure 4.16: Effect of perturbing embedding: Using the wrong drug embedding results in worse segmentation results.

the mitochondria. Moreover, Figure 4.18 presents a 3D rendered visualisation of the data and segmentation.

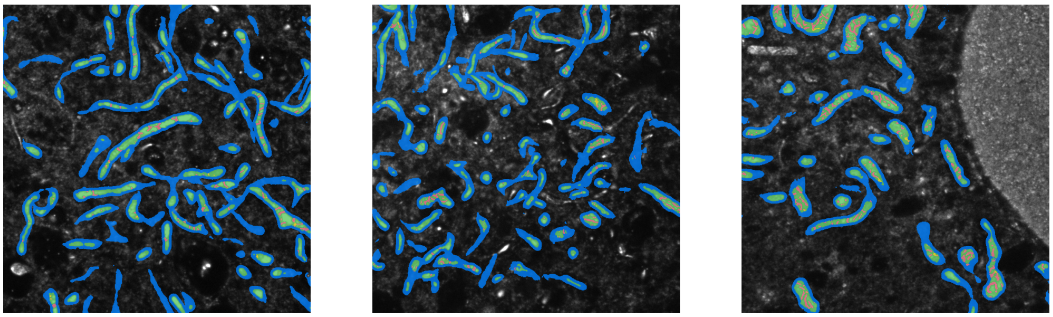
To evaluate the model quantitatively, I follow the same protocol as in the previous section: I randomly pick 25 mitochondria for each condition and have a human annotator revise and possibly correct the prediction made by the model. Subsequently, I compare the original prediction with the revised one both in terms of different segmentation metrics like IoU, Dice, volume change and Hausdorff distance with the results shown in Table 4.2. As it can be seen, the model produces very reliable results and very few corrections are made by the human annotator. The biggest change is seen with regard to the volume of the annotated cristae which is due to the small overall volume which implies that small changes result in relative large percentage differences. This is seen in Table 4.3 which shows proportions of each component and it is seen that the change in the cristae only results in a very small change of the overall proportion. However, the model sometimes



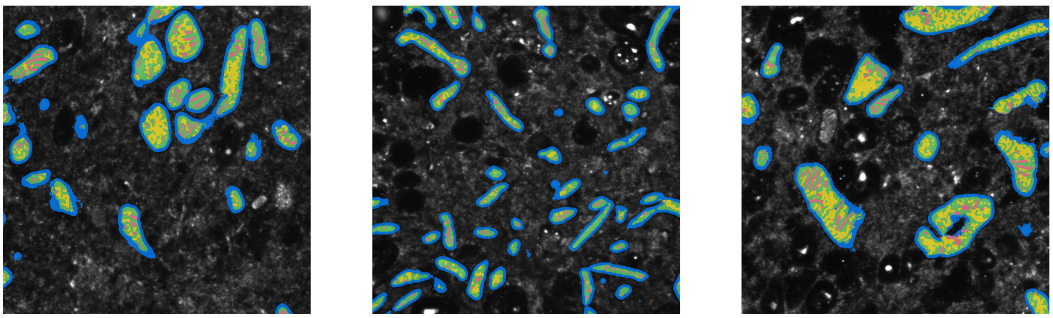
(a) Predictions for DMSO.



(b) Predictions for Antimycin A.



(c) Predictions for HBSS.



(d) Predictions for Oligomycin A.

Figure 4.17: Segmentation masks of the crops presented in Figures 4.10, 4.11, 4.12, 4.13 confirm that the model is able to accurately capture the inner structure of mitochondria across all drugs.

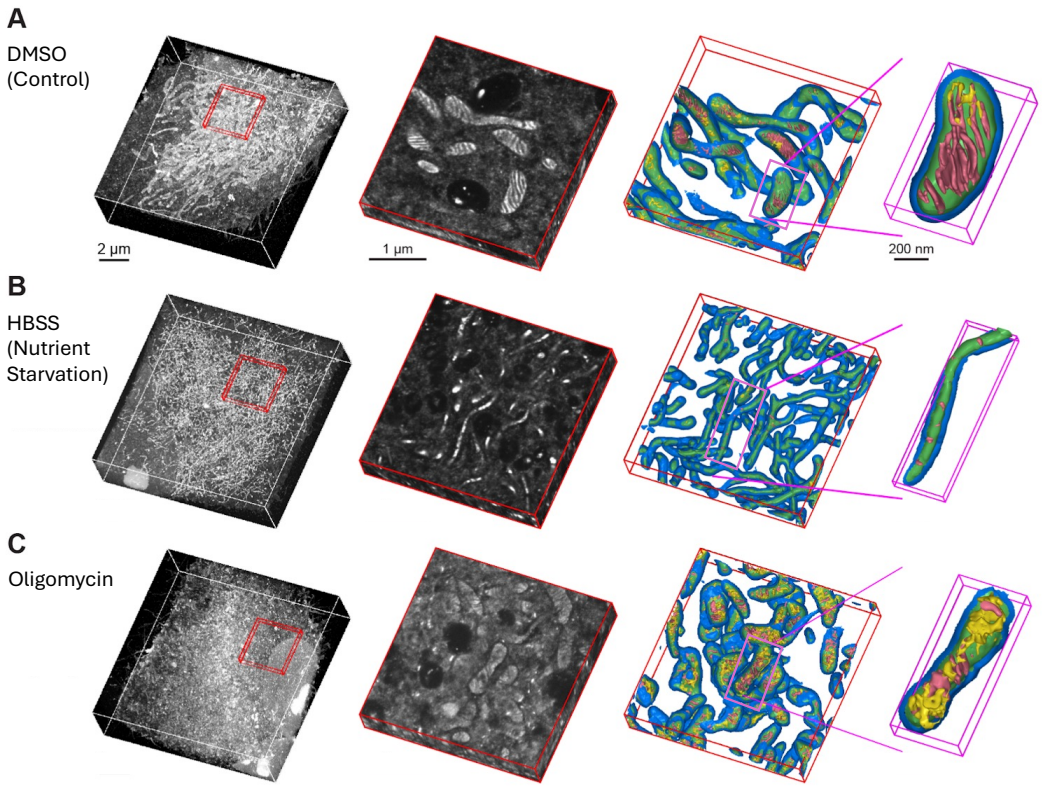


Figure 4.18: 3D Rendering of the pan-staining and segmentation of the drug treated cell. The first column shows a maximum intensity projection of the full field of view. The red rectangle is shown with a higher zoom factor in the second column. Moreover, the contained 3D rendered mitochondria are shown in the third column and one specific instance is displayed in the fourth column. HBSS treated cells (B) shows significantly thinner mitochondria compared to control group in (A). Moreover, we see significantly more ambiguous areas in the cells treated with Oligomycin (C).

misses some of the very thin mitochondria for the HBSS treated samples which is discussed in more detail in Section 4.6.

4.4 Mitochondria in Tissue

The ultimate goal of panExM is to enable researchers across many domains to investigate cells at the ultra-structural level. In this section, I therefore demonstrate how the segmentation pipeline can be adopted and applied to analyse real tissue samples instead of cell cultures like HeLa cells.

The kidney consumes around 10% of the body's oxygen and has, after the heart, the second highest mitochondrial count in the body (Duann et al., 2017) and is

Table 4.2: Evaluation metrics like IoU, Dice coefficient, Volume change and Hausdorff distance for different drug treatments show the reliability of the segmentation model.

Metric/Drug	Intermem. Space	Cristae	Matrix	Ambiguous
DMSO				
IoU	0.971	0.983	0.982	0.941
(Standard Error)	(0.0092)	(0.0066)	(0.0077)	(0.0415)
Dice	0.985	0.991	0.990	0.948
(Standard Error)	(0.005)	(0.0036)	(0.0042)	(0.0399)
Volume	-0.1268	0.5127	0.1926	2.2246
(Standard Error)	(0.1298)	(1.1871)	(0.2506)	(2.2996)
HD	3.435	2.681	3.418	1.974
(Standard Error)	(0.5755)	(0.6638)	(0.8069)	(0.6271)
HBSS				
IoU	0.992	0.835	0.973	0.766
(Standard Error)	(0.0032)	(0.0256)	(0.0045)	(0.082)
Dice	0.996	0.904	0.986	0.774
(Standard Error)	(0.0017)	(0.0166)	(0.0024)	(0.0809)
Volume	-0.1620	16.5632	-1.9364	9.3559
(Standard Error)	(0.0989)	(3.3738)	(0.429)	(7.7706)
HD	1.742	1.019	1.494	0.442
(Standard Error)	(0.3922)	(0.106)	(0.2198)	(0.181)
Oligomycin				
IoU	0.983	0.964	0.978	0.995
(Standard Error)	(0.0067)	(0.0124)	(0.01)	(0.0027)
Dice	0.991	0.981	0.988	0.998
(Standard Error)	(0.0035)	(0.0071)	(0.0055)	(0.0014)
Volume	-0.4669	-0.2073	1.0151	-0.4342
(Standard Error)	(0.2988)	(1.717)	(0.7566)	(0.6456)
HD	2.622	2.566	1.953	1.069
(Standard Error)	(0.4348)	(1.0154)	(0.4376)	(0.4444)
Antimycin				
IoU	0.981	0.953	0.988	0.968
(Standard Error)	(0.0072)	(0.0175)	(0.0053)	(0.0167)
Dice	0.990	0.974	0.994	0.981
(Standard Error)	(0.0039)	(0.0103)	(0.0028)	(0.0103)
Volume	-0.1996	1.9841	0.4775	-1.6155
(Standard Error)	(0.181)	(2.0842)	(0.4037)	(1.8607)
HD	3.919	2.859	2.339	1.944
(Standard Error)	(0.5658)	(0.8567)	(0.5895)	(0.5239)

Table 4.3: Proportion change after revision: For each drug treatment, the change in the proportion of each component between the initial model prediction and after the human revision is shown. Since the changes are mostly small, it can be concluded that the model produces very reliable segmentations.

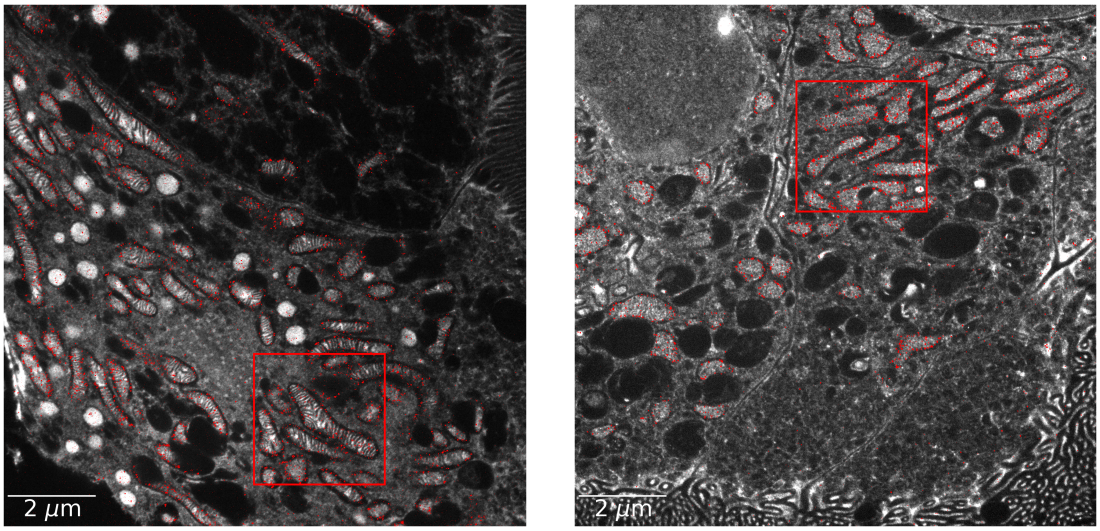
Segment	DMSO	HBSS	Oligo	Antimycin
Intermembrane	60.25 → 60.18	69.42 → 69.31	57.83 → 57.54	61.64 → 61.53
Cristae	7.59 → 7.58	3.32 → 3.92	7.23 → 7.23	6.20 → 6.30
Matrix	30.55 → 30.61	26.99 → 26.44	29.96 → 30.26	27.02 → 27.15
Ambiguous	1.61 → 1.63	0.26 → 0.32	4.98 → 4.98	5.14 → 5.03

therefore an interesting candidate to apply our method to. Acute kidney injury (AKI) is defined as a sudden decline in the kidney function and it has been a global health concern with a worldwide prevalence of 13 million people (Mehta et al., 2015). AKI have been associated with changes in the mitochondria but there currently no therapies available (Ishimoto et al., 2016; Ralto et al., 2016).

In this section I extend our model developed in Section 4.2 to segment the mitochondria in mouse kidney samples. I develop an algorithm which allows to learn the mitochondria outline purely from noisy and sparse immunolabeling. Subsequently, I present two ways to fine-tune the model to analyse the mitochondria in the wild type and compare the inner structure of the wild type to samples with AKI. Finally, I investigate conditions for the immunolabeling for the presented method to work as a guideline for future adoptions to different settings.

4.4.1 Kidney Tissue Data

In order to investigate the mitochondrial changes of AKI cell, mouse cells of healthy animals and mice suffering from AKI are imaged by our collaborators at Yale University and provided to us. Specifically, we look at cortical proximal tubular segments (S1/S2) (Schuh et al., 2018). The AKI was induced through the treatment with cisplatin which is a cancer drug that causes nephrotoxicity as a side effect (Guo et al., 2023). The sample preparation was conducted by Xiaojia Guo and the expansion and imaging was conducted by Yuan Tian, both at Yale University.



(a) WT **(b)** AKI

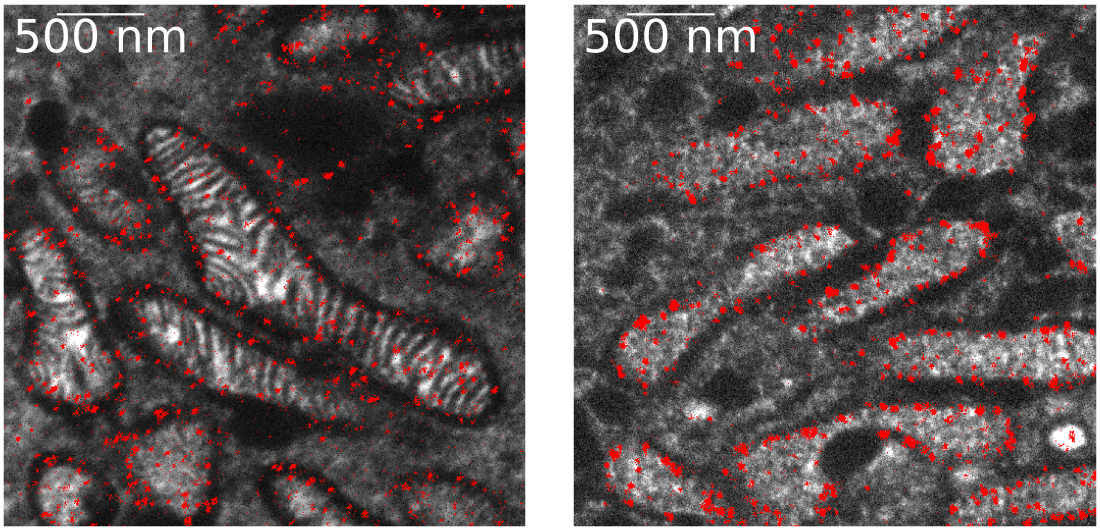
Figure 4.19: Kidney WT and AKI pan-staining: Acute Kidney Injury (AKI) is associated with a drastic change in the morphology of the mitochondria compared to the wild type. The binding of the anti-body mix is shown in red. The red rectangles are shown in Figure 4.20. All samples in this thesis showing the pan-staining have been imaged by Yuan Tian at Yale University.

The samples go through the same expansion process as the HeLa cells and are stained with NHS ester to generate a global pan staining.

Figures 4.19 and 4.20 show examples of two kidney cells imaged with panExM. It can be seen how the inner structure of the mitochondria in the AKI cell does no longer show any lamellar shaped cristae.

The method presented in Section 4.2 relied on the MitoTracker dye to give us the outline of the mitochondria. However, for tissue samples the MitoTracker dye does not work as it needs to be applied to living cells. The dye relies on the electrochemical gradient across the mitochondrial membrane to accumulate and this membrane potential collapses if the cellular respiration stops. As an alternative our collaborators provided us with a mix of two anti-bodies Mic60 and COX IV which is shown in red in Figure 4.19 and in greater detail in Figure 4.20. The antibodies are described in more detail in Section 6.3.

The presence of new structures unseen in the HeLa cells like the microvilli and the substantial changes in the appearance of the mitochondria implies that our



(a) WT **(b)** AKI

Figure 4.20: Zoomed in Kidney WT and AKI pan-staining and anti-bodies: The antibody labelling (red) for the kidney samples is a lot sparser than the MitoTracker dye and more susceptible to false positives.

previous models are not directly applicable. Moreover, was also cannot simple retrain the model based on the approach outlined in Section 4.2 as the anti-body staining does not offer the same sensitivity and precision as the MitoTracker.

The next section develops a training procedure which allows to identify the mitochondria based on the available sparse labeling. Subsequently, I fine-tune the model from the HeLa cells to identify the inner structure of the mitochondria.

4.4.2 Training with Sparse Antibodies

Due to the substantially higher noise and resulting false positives and many false negatives, it is not possible to just apply morphological filtering to the immunolabeling in order to derive segmentation targets but a more sophisticated methodology is required.

Using anti-bodies in microscopy images to derive training labels for machine learning algorithms is a very unexplored field. One example in histology is Stenman et al. (2021) follows a similar approach to Section 4.2 as the authors are able to derive binary masks using simple morphological operations. The method I propose combines ideas from sparse annotations and noisy labels to generate a

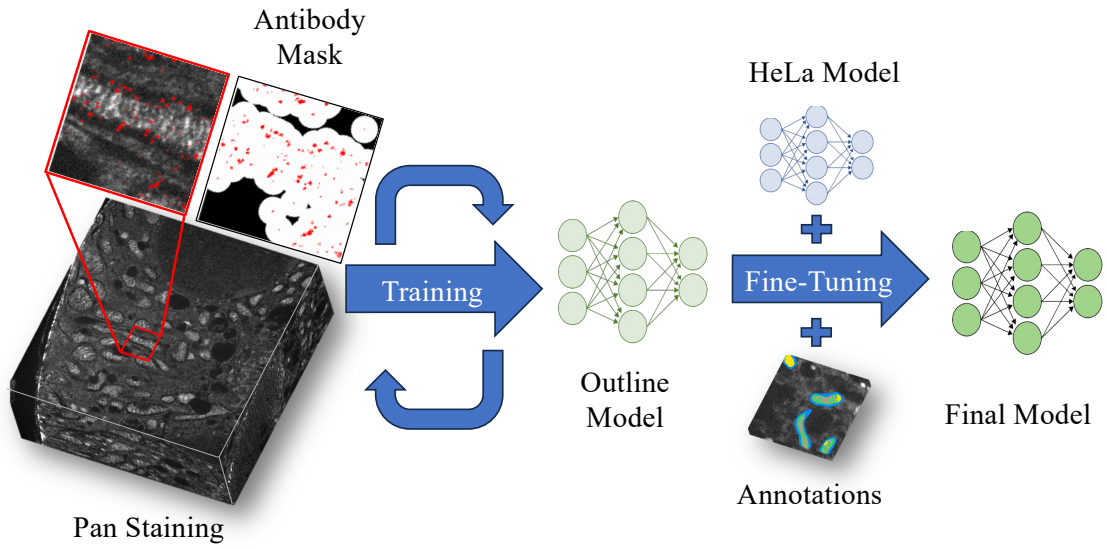


Figure 4.21: Antibody based Training: Based on the antibody labelling (red) the neighbourhoods R_τ (white) are derived and the background mask B (black). These serve as input for the semi-supervised training loop and produce the initial outline model which then gets further fine-tuned based on the HeLa model ultrastructure predictions and additional annotations.

semi-supervised learning problem. The literature regarding noisy labels and semi-supervised approaches have already been reviewed in Chapter 2. Sparse annotations have already received attention before the change to deep learning as they can substantially lower the annotation time (Sommer et al., 2011). Examples include annotations based on scribbles (Can et al., 2018; Gao et al., 2022) and points (Bearman et al., 2016) which have been shown to good performance. However, these sparse annotations refer to human annotations with a very different character from the sparse annotations created by the immunolabeling.

Method

To formalise the proposed method, denote the presence of a mitochondrion at pixel i by $Z \in \{0, 1\}$ and the pan-stained image by $X \in \mathbb{R}^{Z \times H \times W}$. Additionally, let us denote the set of all pixels in the image cube by $\mathcal{P}$. Moreover, the anti-body information $y_i \in \{0, 1\}$ at pixel i is correlated with the presence of a mitochondrion. However, as shown in Figure 4.20 this relationship is noisy with very many false negatives (i.e. pixels within the mitochondrion that are not labelled) and some false

positives (i.e. anti-body markings occurring far from any mitochondrion).

To address the large number of false negatives, it is assumed that an antibody marking y_i not only tells us something about the presence of a mitochondrion at the specific pixel i but also about all pixels in some neighbourhood $R_\tau(i) \subset \mathcal{P}$ around i with the size of the neighbourhood parametrised by τ. Therefore, I define the background mask as the complement of the intersection of all neighbourhoods of labelled pixel

$$B := \mathcal{P} \setminus \bigcup_{i \in \mathcal{P}: y_i = 1} R_\tau(i)$$

and assume that no mitochondrion is present and set the prediction target value to 0. This antibody masked is visualised in Figure 4.21.

For pixels within some R_τ which are not labelled by the antibodies

$$I := \{i \in \mathcal{P} : y_i = 0 \text{ and } i \in \bigcup_{i \in \mathcal{P}: y_i = 1} R_\tau(i)\}$$

I assume a prior of ignorance and make no judgement about the underlying mitochondria z there. This implies that these pixels are ignored in the loss function initially. Moreover, at this point there is have no way of judging the reliability of the anti-body markings and I start by treating all of them as examples of underlying mitochondria

$$M := \{i \in \mathcal{P} : y_i = 1\}.$$

This assumption is justified even though the number of false positives is high due to the large amount of background space since the false positives seem to be random and not structured.

I now cast this problem as a semi-supervised learning problem and train a network to predict $\hat{p}(z|x)$ with the positive class given by pixels M and the negative background class given by B. Pixels in I are considered unlabelled and ignored initially. By ignoring the neighbourhoods I which are not marked initially, the method is able to eliminate the large amount of false negatives and build on the hypothesis that the number of true positives dominates the number of false negatives.

In the initial phase of training, this model for $\hat{p}(z|x)$ does not provide an accurate estimate of the true underlying probability but over the course of training the model will be of increasing quality and therefore can be used to identify false positive and negative markings of the antibodies. This allows to follow a pseudo-labelling approach (Lee, 2013). More specifically, if for some pixel i with positive anti-body labelling the model finds that its associated image features are very different from what it learned to expect for mitochondria and therefore it predicts a very low probability $\hat{p}(z|x)$, this pixel is judged this y_i to be a false positive and set it to zero. On the other hand, if it is found that a neighbourhood around a positive anti-body marking is very close to the prototype representation of a mitochondrion, it is judged them to be false negatives and set the target information to 1. This procedure is followed until the training converges.

Fine-Tuning with the HeLa Cell Model

The training algorithm just described is necessary as the differences in the cell structure between the HeLa cells and the kidney tissue are too big to just apply our model from Section 4.2.

However, it turns out that if the HeLa model is applied to just to the mitochondria masks and restricted to predict matrix, cristae and ambiguous, it are able to produce acceptable initial segmentation masks. In a first step, I therefore train our outline model to reproduce the combination of its outline predictions filled with the predictions of the HeLa model.

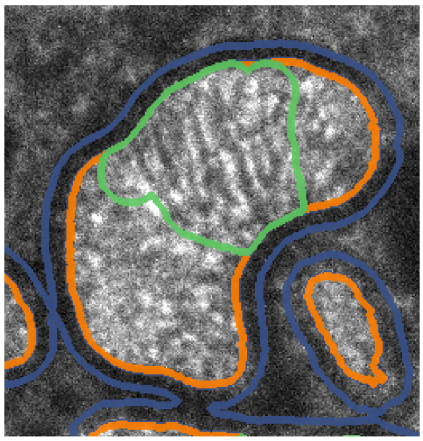


Figure 4.22: Lamellar Cristae: Lamellar cristae is circled in green, remaining inner structure in orange.

In order to further improve the segmentation quality, I follow an active learning approach (Ren et al., 2021) and select crops in which the model produces erroneous

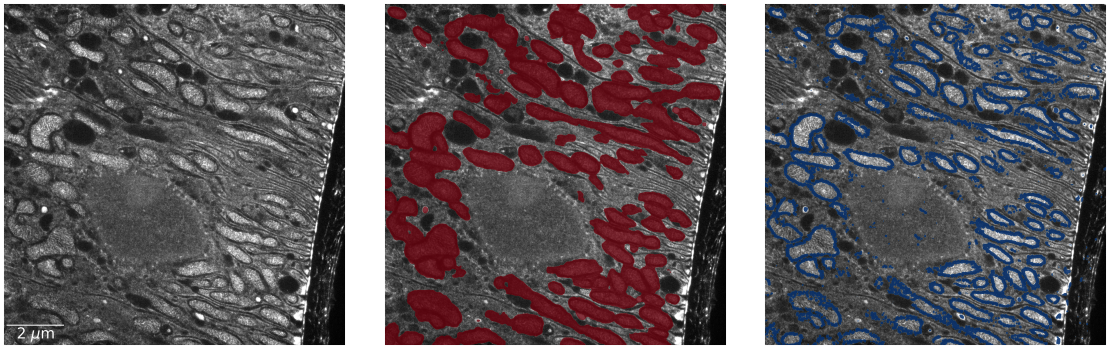
segmentations and correct these manually
before refitting the model. As noted in Section 4.2 it is sufficient to sparsely correct only a few slices due to the high spatial dependency along the z-axis and it is important to note that correcting the model segmentations is substantially faster than creating the annotations from scratch.

The visual impression in Figure 4.19 shows that most of the AKI mitochondria do no show any lamellar cristae. In order investigate this claim more thoroughly, I train a second, independent fine-tuned model. This model is trained to highlight lamellar cristae and assign a new class to the remaining inner area of the mitochondria. Figure 4.22 provides an example of the resulting distinction. These annotations can be added very efficiently as the fine structure is already provided by the initial model and only the broad indication of non-lamellar cristae is needed.

Implementation Details

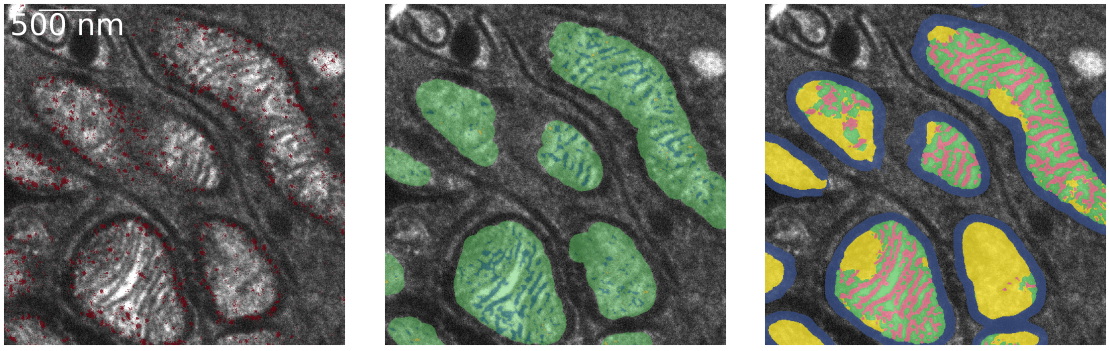
Regarding the implementation details, I use the same modified UNet architecture and data augmentation as in Section 4.2. I manually set the neighbourhood parameter to $\tau = 18$ which was visually judged to be sufficient so that the resulting neighbourhoods mask most of the mitochondria as can be seen in Figure 4.21. Moreover, to reduce additional noise isolated, single pixels that are labelled by the anti-bodies, are removed.

The outline model is trained with all available stacks. Because the field of views (FoVs) were imaged with three different samples, I adjust the anti-body threshold manually for each sample to account for differences in the intensity levels. For the general fine-tuning, 20 additional sparsely annotated mitochondria crops of size $32 \times 512 \times 512$ are manually corrected. Because of the much faster annotation for the differentiation between lamellar and non-lamellar structure, I provide 65 sparsely annotated mitochondria crops in this case. In both cases, the time required for the annotation was about four hours.

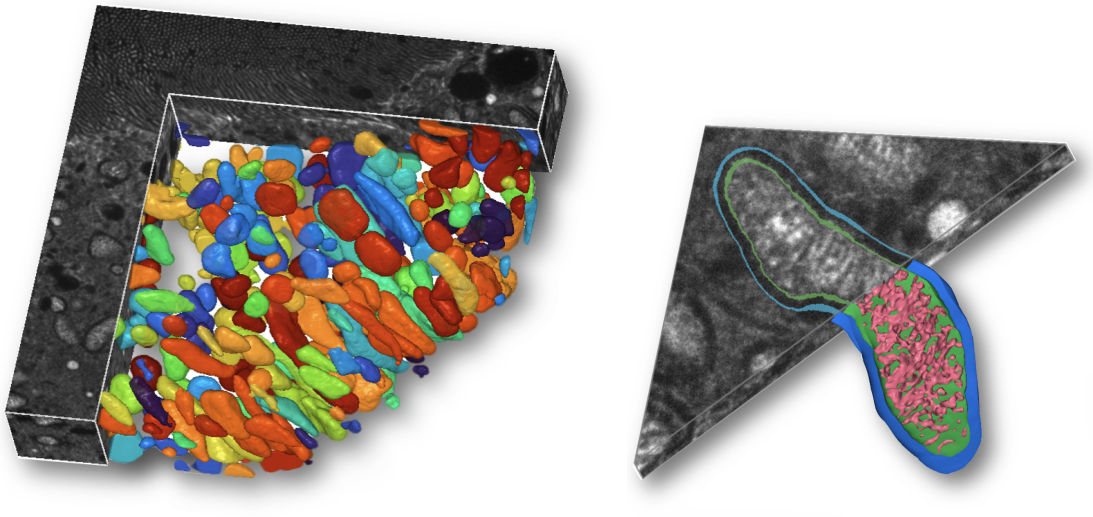


(a) Pan Staining **(b)** Predicted Outline **(c)** Outline with smaller $\tilde{\tau}$

Figure 4.23: Fully unsupervised outline prediction: The outline and the boundary of the mitochondria can be segmented by a model trained purely on the anti-bodies. Combining both predictions allows to identify individual mitochondria instances.



(a) Pan staining **(b)** HeLa + Outline model **(c)** Final Model



(d) 3D Field of View **(e)** 3D Mitochondrion

Figure 4.24: Finetuning of Kidney Predictions with HeLa model: After filling in the outline prediction with the prediction of the inner structure by the HeLa-model (b), the rough outline cristae structure is present. After finetuning (c), the model is able to predict the whole ultrastructure.

4.4.3 Results

Figure 4.23b shows the results of the mitochondria outline prediction based purely on the anti-bodies. It can be seen that the model provides surprisingly high quality segmentations given the noisy nature of the training data provided by the anti-bodies. However, as for example shown in the bottom right corner, the model is sometimes not able to differentiate single instances of mitochondria if they are in close proximity.

A closer inspection of the anti-body labelling reveals that the staining is more prominent on the outer membrane of the mitochondria and in the inner area. This allows to apply the training algorithm again with a smaller neighbourhoods $\tilde{R}_{\tau'}(i) \subset R_\tau(i)$ and additionally ignoring any pixels outside of the masks found in the previous step. This procedure enables to learn a more precise outline of the mitochondria boundary as shown in Figure 4.23c. Combining these two masks and a watershed algorithm (Neubert et al., 2014) generates the instance segmentation and provides the full mitochondria segmentation without any human annotation.

Figure 4.24(b) shows an example of the initial segmentation which combines the outline from the model above with the inner structure predicted by the HeLa-cell model from Section 4.2. Moreover, the final prediction after the fine-tuning on the manual annotations is shown in Figure 4.24(c). Figure 4.24(d) and (e) show 3D rendered versions of the whole FoV and one zoomed in mitochondrion. The high accuracy of these segmentation masks was confirmed by our collaborators and form the basis of the subsequent analysis presented in Section 6.2.

4.5 Antibody Requirements for Future Applications

One interesting aspect for future applications of the presented algorithm to other cell types or organelles is the question how precise the anti-body labelling needs to be in order to allow the identification of the organelles of interest.

Treating our fine-tuned segmentation masks as ground truth labels for the mitochondria, it is possible to evaluate how much of the anti-bodies localise within

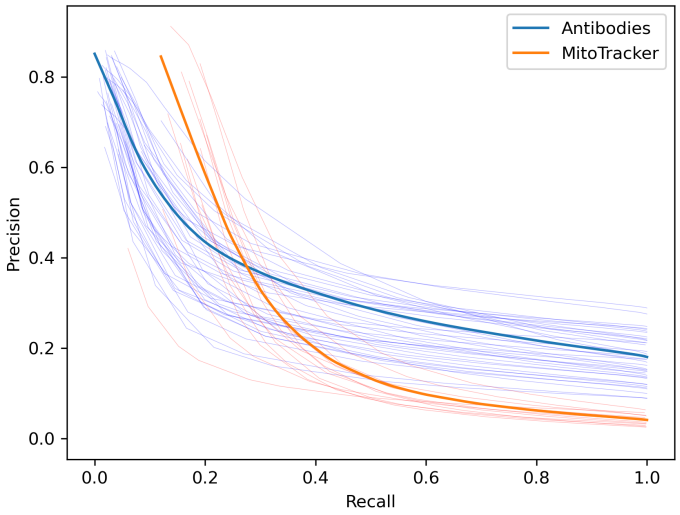


Figure 4.25: Kidney Antibody Localisation: The precision and recall curves show that MitoTracker localises preciser for the same recall. Individual cells are shown with thin lines, the average is shown in bold. I hypothesize that in a new application, any anti-body that performs at least as well as the Mic60-COXIV mix is sufficient to apply our method for training an outline-segmentation model.

them and what proportion of the mitochondria gets labelled. However, the anti-bodies do not come as a binary mask but with continuous intensity levels. I therefore treat the anti-bodies as a 'classifier' and calculate the precision (proportion of true positive predictions to number of all positive prediction) and recall (proportion of true positive prediction to the number of true positives) when the immunolabeling is thresholded at increasing intensity levels.

The resulting precision-recall (PR) curves are shown in Figure 4.25 and compared to the MitoTracker used in Section 4.2. The best results with our method are achieved when setting the threshold to achieve a very high precision and account for the low recall with a larger neighbourhood radius τ. Specifically, the threshold chosen implied a mean recall of 0.12 and a mean precision of 0.6. In terms of Figure 4.25, this implies that the crucial part is the top left corner and I hypothesis that in a new application scenario, any antibodies with a PR curve lying between the blue and orange line curve should work. The PR curve in the new setting could be estimated based on a small annotated crop.

4.6 Discussion

When applying deep learning methods to a new domain or a new dataset one of the biggest challenges is to obtain a large amount of labelled samples. These annotations are especially laborious in a 3D setting. In this chapter, I demonstrated how deep learning algorithms can exploit the capability of panExM to add additional target specific stainings to the pan staining to alleviate the required amount of human annotations.

Section 4.2 introduces the idea to split the segmentation task into the two sub-problems of finding the mitochondria outline and subsequently filling in the inner structure. In this section, I work with HeLa cells for which the MitoTracker dye provides a fairly accurate outline of the mitochondria. Based on the MitoTracker I derive outline masks which serve as segmentation targets to train an initial outline model capable of segmenting the mitochondria based on the pan staining. Subsequently, I fine-tune the model with additional human annotations to also segment the inner structure and I provide evidence how the two step approach allows to minimise the annotation requirement. Moreover, I validate the segmentation masks to provide accurate results.

After demonstrating the feasibility of the framework on control HeLa cell, I apply the pipeline in Section 4.3 to HeLa cells which are treated with drugs to induce various morphological changes and develop a modification to the model to efficiently adopt to this new setting.

In Section 4.4, I extend the model to tissue samples and generalise the two-step training procedure to generate a mitochondria outline model without any human annotation purely based on a very noisy antibody labelling instead of the MitoTracker. I demonstrated that with a very limited amount of additional annotation time and the readily available baseline model from Section 4.2, the model can again be finetuned to segment the complete mitochondrial ultrastructure. This allows the quantitative analysis of the mitochondria and their ultrastructure for mouse kidney samples. Moreover, using the example of AKI, I showed how

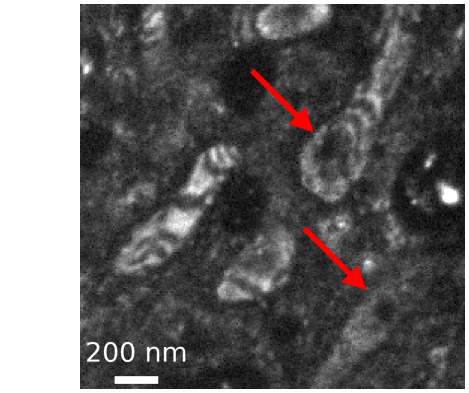
the model can be adopted to answer context specific questions like fine-tuning the model to differentiate between lamellar and non-lamellar cristae.

Section 4.4 significantly improves the applicability of the presented ML approach as it proves that the segmentation pipeline is not limited to cell cultures and settings in which a very accurate dye for the organelles of interest is available. We are confident that this work paves the way for more machine learning applications to pan stained ExM data and hope to facilitate the adoption process by having provided some guidelines for the selection of the antibodies.

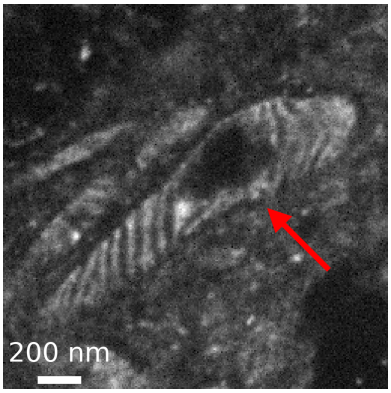
Recently, other self-supervised pre-training methods like pre-training masked autoencoders (Kraus et al., 2024; Xie et al., 2023) have also shown promising results. However, these approaches rely on very large amounts of data and compute that might not be available feasible for average biologist aiming to segment their panExM data. Moreover, MAE are specific to the ViT architecture whereas the presented approach can be used independent of any architecture. Finally, in the case of the sparse immunolabeling shown in Section 4.4 for the kidney, the presented approach not only offers efficient pre-training for the ultra-structure segmentation but also enables to segment the mitochondria themselves in the first place without any human annotation required.

Limitations

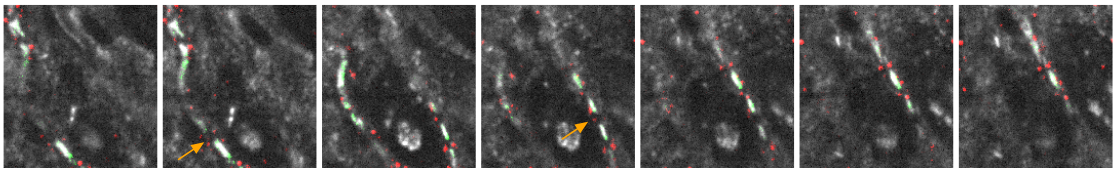
As discussed in Chapter 2, mitochondria have their own DNA. This DNA manifests itself as dark holes within the mitochondrion and for most mitochondria, these holes are of small or modest size as seen in Figure 4.26(a). However, for some mitochondria they can occupy a substantial volume as seen in Figure 4.26(b). The presence of DNA within these dark holes was confirmed by our collaborators through DNA targeting immunolabeling. This target specific staining would however only label a small fraction of the volume of these holes and at the time of writing, it could not be said with certainty whether the DNA fills the complete holes. However, as these holes are visually very similar to the representation of cristae, they have



(a) Modest hole



(b) Large hole



(c) Large gaps in the mitochondria on adjacent slices in HBSS cells.

Figure 4.26: Potential biases: (a) shows how the mitochondria DNA is represented by the pan staining as dark circles (red arrows). This might bias the quantification of the cristae area as for a few mitochondria like in (b) the area can be significant. Moreover, for HBSS treated mitochondria it is sometimes not clear to identify connected instances as they are separated by large gaps (orange arrow) as seen in (c).

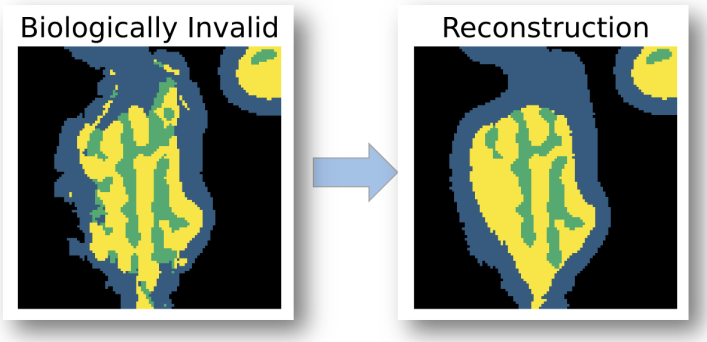
the potential to bias the analysis. Future work therefore could aim to segment the DNA as a separate components in the segmentation mask.

Additionally, some cristae structures approach the limit of the currently achievable resolution and are just a few pixels apart. It is therefore possible that the analysis misses some cristae structures which could account for some of the ambiguous labels. This is also true if the cristae is orientated dominantly in the z-stack direction as the resolution is about 2.5 times smaller there.

Finally, it is observed that the mitochondria can become very thin under nutrient starvation (HBSS) and the segmentation model therefore might sometimes fail to discover all connections between different mitochondria instances. Figure 4.26(c) for example shows seven adjacent slices of the panstaining, MitoTracker (green) and Tom20 (red). As it can be seen, the mitochondrion is unusually thin and interrupted by gaps larger than its diameter. These situations pose a significant challenge to the segmentation model. As only very few connections have to be missed to

influence the subsequent network structure and average length, it is possible that the presented analysis underestimates the resulting length and networks formed by the mitochondria. I reviewed the segmentation masks extensively and while some thin mitochondria are missed there seems no indication that these missing mitochondria masks are responsible for the fragmentation of long networks and I am therefore confident in our result.

With regard to the guidance on the immunolabeling quality, the current guidance is based for a future segmentation task of similar difficulty which means organelles that have similar degree if differentiability from the background. While I aim to provide some initial advise here, only additional future applications can truely show what biomarker quality is sufficient.



5

Post-Processing and Parameter Extraction

Contribution

This chapter extends the previous work by providing post-processing steps for the initial segmentation masks and to extract parameters of interest.
I develop a new algorithm building on ensembling methods and masked autoencoders to learn a prior for biologically plausible segmentation masks purely from the data and apply it automatically and selectively to areas of high uncertainty. This section has been published with minor modifications as *Refining Biologically Inconsistent Segmentation Masks with Masked Autoencoders* in 2024 in the *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6904-6912).
Subsequently, I describe the extraction of parameters of interest like the skeletons, length and diameter based on the final segmentation masks.

Contents

5.1 Introduction 100
5.2 Refining Biologically Inconsistent Segmentation Masks 100
5.2.1 Uncertainty Detection 103
5.2.2 Shape Constraints 104
5.2.3 Method 104
5.2.4 Experiments 109
5.2.5 Postprocessing 115
5.3 Parameter Extraction 116
5.3.1 Instance Segmentation 116
5.3.2 Skeletonisation 117
5.3.3 Length and Diameter Calculation 119
5.4 Discussion 121

5.1 Introduction

While the automated segmentation masks are the first step towards a quantitative analysis, they still lack any measurements which are easily interpreted by biologists. Moreover, small segmentation errors can easily bias the subsequent results. This chapter therefore first present a new reconstruction algorithm that allows to correct the segmentation masks in areas with high uncertainty. Subsequently, I present additional postprocessing steps like the skeletonisation of the mitochondria which are necessary to compute measurements like the lengths and diameter which form the basis for the quantitative analysis in the next chapter.

5.2 Refining Biologically Inconsistent Segmentation Masks

Similar to all microscopy images, the pan stained ExM images are sometimes affected by uneven signal quality.[1] This can be due to various reasons like the biological sample itself and its preparation, the staining process, inadequate exposure times or photon-induced bleaching. Such degradation in image quality not only compromises the visual examination of biological specimens but also impedes subsequent analysis steps. For example, the resulting regions of low SNR pose a challenge for the computer generated segmentation masks as they offer substantially more room for interpretation and this ambiguity results in the lack of a well defined corresponding segmentation ground truth.

Nevertheless, the inherent self-similarity of biological structures often provide constraints to perform effective signal amplification and refinement. In this section, I demonstrate on the concrete example of mitochondria imaging a new post-processing framework for resolving highly ambiguous regions of poor signal quality to enhance the phenotypical analysis of cellular organelles.

[1] This section have been published as Sauer et al. (2024) © 2023 IEEE and parts have been quoted verbatim from it with permission.

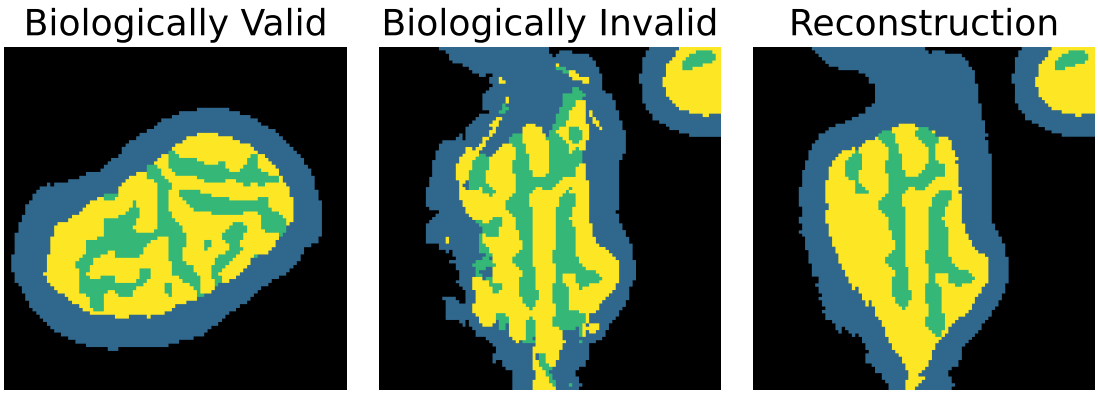


Figure 5.1: Comparison of a biologically valid segmentation mask, a biologically invalid segmentation masks and the proposed reconstruction Left: Example showing a segmentation mask which is biologically plausible showing *cristae* invaginations (green) surrounded by *matrix* (yellow), the inner membrane being surrounded by *intermembrane space* (blue) and overall smooth boundaries. Middle: A segmentation mask that violates the biological constraints with separated *cristae* areas and fragmented boundaries. Right: The same example adhering to biological constraints again after applying my MAE refinement. ©2024 IEEE

As depicted in Figure 5.1 and as described in Chapter 2, mitochondria are formed by a double membrane structure (Scheffler, 2008) and these membranes typically form smooth structures without any fragmentation. Moreover, the inner membrane forms invaginations, called *cristae*, which therefore should mostly be surrounded by area of the inner membrane, called *matrix*, and cannot appear independently.

The typical approach for image segmentation is to predict a probability distribution over the possible classes for each pixel and output the class with the highest likelihood. If the underlying image offers a lot of ambiguity, however, the segmentation model can produce segmentation masks which are not coherent with the biological prior knowledge as dependencies among labels are only indirectly accounted for through the shared context.

These segmentation mistakes are problematic as they can introduce artefacts in subsequent analysis steps. The example in Figure 5.2 shows how the resulting skeletonisation of the inner structure of the mitochondria shows several loops and branches which would bias any resulting length measurement without additional post-processing steps.

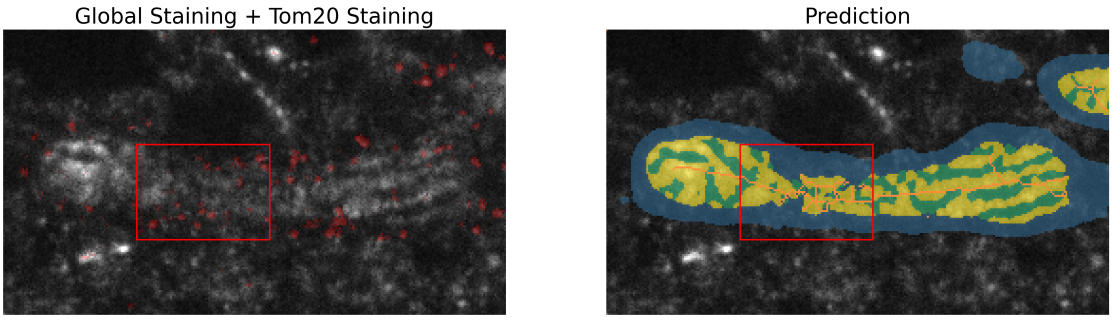


Figure 5.2: Poor segmentation as a consequence of low SNR: The red rectangle highlights an area in which the global staining (M'Saad et al., 2020) is unclear and shows how this leads a segmentation mask that is biologically inconsistent. This introduces artefacts in subsequent analysis steps like the skeletonisation (Lee et al., 1994) which is shown in orange. The Tom20 staining in red (which is known to bind to the outer membrane of the mitochondria) shows that the highlighted area is still part of the inner membrane and not outside or on the boundary. This Tom20 staining is not used in any subsequent analysis. ©2024 IEEE

Addressing this challenge can be divided into two sub-problems: how is it possible to identify areas of possibly poor staining quality and subsequently, what is the appropriate strategy to deal with them? Sticking with the example of mitochondria, one naive approach would be to exclude all mitochondria which are somewhere affected by some poor imaging. However, this introduces a significant bias into any subsequent analysis: As longer mitochondria occupy more area, they are more likely to be affected by some imaging issue and therefore more likely to be removed.

Previous attempts have been made to impose shape priors into segmentation models (Bohlender et al., 2023). However, formulating these priors is challenging as they have to be fine-tuned manually and the resulting methods suffer from long run times unsuitable for the large volumes of data produced by modern microscopes. On the other hand, recent advances in machine learning suggest that learning such priors directly from the data offers a superior approach.

When human experts are confronted with a small area of high ambiguity, they annotate the sample to be consistent with the well-imaged surrounding area and their knowledge of the underlying biological structure. This ability has been gained by having seen a large number of coherent samples and is made possible by a large degree of local self-similarity between the mitochondria samples.

My method aims to imitate the process followed by a biologist and I demonstrate how it improves the segmentation quality of mitochondria imaged with a global 'pan'-staining using expansion microscopy (Chen et al., 2015; M'Saad et al., 2020).

In this section, I first show that the disagreement between an ensemble of models provides a good approximation for the underlying image quality. Once areas of low SNR are identified, a masked auto-encoder (MAE) is trained to reconstruct biologically plausible segmentation masks from a corrupted segmentation masks. For inference, the method specifically masks the area of high ambiguity and uses the trained MAE to reconstruct the segmentation output consistently with the surrounding area to form biologically coherent predictions. I demonstrate that my method produces more sound segmentation masks than competing post-processing methods like conditional random fields (CRF) and leads to fewer artifacts in the subsequent analysis.

5.2.1 Uncertainty Detection

Common methods to quantify prediction uncertainty include Bayesian neural networks (BNN), Monte Carlo drop-out methods and deep ensembles (Abdar et al., 2021; Ganaie et al., 2022). While BNN (Graves, 2011) offer a theoretically elegant solution to capture the uncertainty in the model weights, they suffer from high computational complexity. Monte Carlo drop-out methods (Gal et al., 2016) approximate the posterior predictive distribution by multiple forward passes of the same input through the network while omitting random subsets of neurons, but might be poorly calibrated and sensitive to hyper-parameter choices (Caldeira et al., 2020).

Deep ensembles (Lakshminarayanan et al., 2017) are conceptually simple as they only require retraining the same model with different initialisations and the prediction uncertainty is computed by investigating the difference in the individual predictions.

5.2.2 Shape Constraints

Popular approaches to include shape priors in segmentation models are conditional random fields (CRF) and active contour models (ACM) (Bohlender et al., 2023). CRFs are typically used as a post processing step and build a neighbourhood graph that allows to incorporate information about adjacent pixels into the prediction. However, the approach is limited by high computational cost for large neighbourhoods except when additional constraints are imposed on the neighbourhood influence (Krähenbühl et al., 2011).

ACMs on the other hand are based on the idea of iteratively transforming an initial outline to minimise some energy function. This energy function again needs to be fine-tuned to the application at hand and the resulting models have a limited adaptability to datasets showing a large variety of shapes and object sizes as typically encountered with the inner structure of mitochondria.

Another approach attempts to capture the shape in a lower dimensional space through an autoencoder. Larrazabal et al. (2019) pass the initial segmentation mask through an autoencoder to smooth out low frequency information to regularise the final shape. However, compared to my method, their approach works globally without focusing on ambiguous areas and requires handcrafting a deterioration function to train the autoencoder.

5.2.3 Method

In order to formalise the proposed method, denote the globally stained image by $X \in \mathbb{R}^{H \times W}$ where H, W denote the height and width of the image. Moreover, denote the set of possible segmentation labels by $\mathcal{C}$ and the associated segmentation mask by $Y \in \mathcal{C}^{H \times W}$. For notational simplicity, the method is presented for the two dimensional case as the extension to 3D follows naturally.

The standard practice for image segmentation in the deep learning framework is to train a network f_θ parametrised by θ to predict a probability distribution $\mathbb{P}_{f_\theta}[Y_i|X]$ over $\mathcal{C}$ for each pixel i given the input X. The final prediction is subsequently chosen as the label which maximises this probability $\hat{y}_i = \arg\max_{c \in \mathcal{C}} \mathbb{P}_{f_\theta}[Y_i = c|X]$.

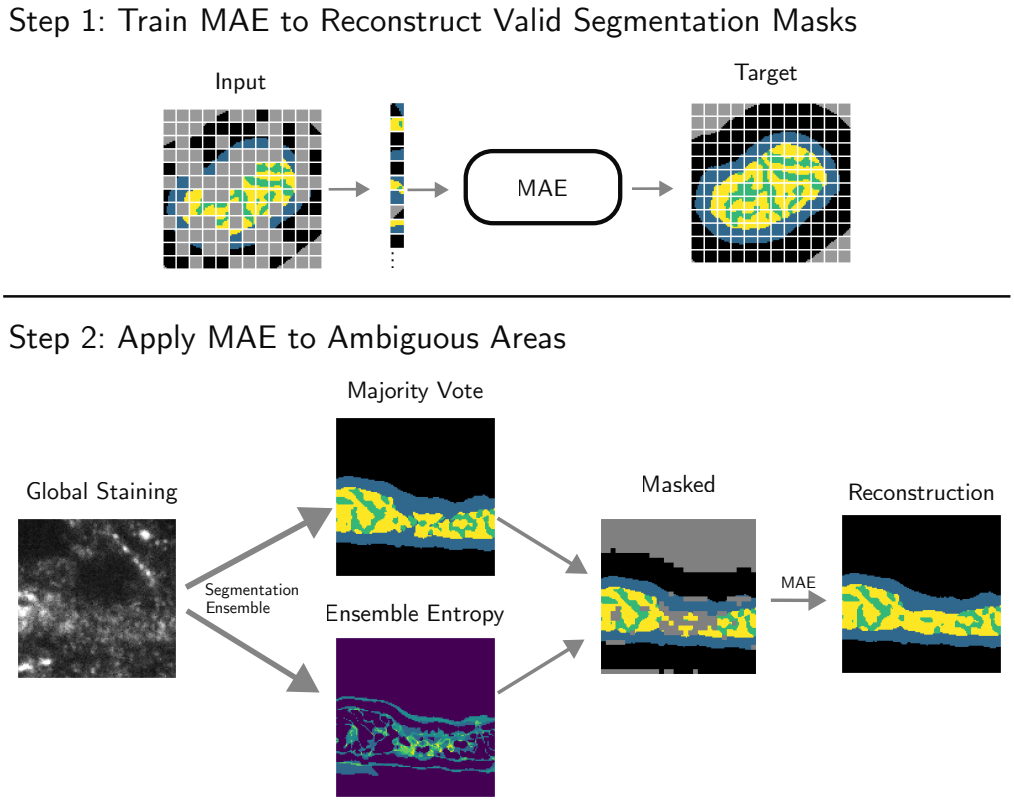


Figure 5.3: Reconstruction method overview: In the first step a Masked-Autoencoder (MAE) is trained to construct biologically coherent segmentation masks from partial segmentations while ignoring random patches (greyed out). The second step applies an ensemble of models to the global staining to compute the majority vote and ensemble entropy. Subsequently, patches with a high ensemble entropy are removed and reconstruct them using the MAE. Background patches far from the segmentation mask are ignored as well. The shown example patch corresponds to the red rectangle from Figure 5.2. ©2024 IEEE

This approach is based on the assumption that conditional on the input X, the pixel labels are independent and the overall probability can be factorised over individual pixels $\mathbb{P}[Y|X] = \prod_i \mathbb{P}[Y_i|X]$.

However, let us consider pixel i in more detail and rewrite the global probability accurately as

$$\mathbb{P}[Y|X] = \mathbb{P}[Y_i|X, Y_{-i}]\mathbb{P}[Y_{-i}|X] \tag{5.1}$$

where Y_{-i} denotes all remaining pixels. If pixel i is poorly stained and shows a low SNR, X provides little guidance for the associated segmentation label of this pixel. Therefore, in this work I explore the assumption that it is therefore better

predicted by its surrounding pixels, i.e. $\mathbb{P}[Y_i|X, Y_{-i}]$ might be better approximated by $\mathbb{P}[Y_i|Y_{-i}]$ than $\mathbb{P}[Y_i|X]$ for poorly stained pixels. When denoting the area of unclear staining by U and the remaining, reliable pixels by R, I therefore propose to predict the final segmentation according to the factorisation

$$\mathbb{P}[Y|X] = \mathbb{P}[Y_U|Y_R] \cdot \prod_{i \in R} \mathbb{P}[Y_i|X] \tag{5.2}$$

in which the conditional independence assumption only is only applied for reliable pixels in R.

This approach leaves us with two tasks: Firstly, finding the pixels with potentially ambiguous staining and therefore possibly inconsistent segmentations and secondly, learning the unconditional distribution $P(Y_U|Y_R)$ for applying a prior of consistent annotations to this area. Our proposed pipeline is summarised in Figure 5.3.

Detecting Regions with High Uncertainty

I propose to define an area of unclear staining as a set of pixels $U \subset \{1, ..., H \cdot W\}$ in which X does not carry enough information by itself to produce an unambiguous segmentation mask Y_U. This implies that in a well calibrated model, one would expect the predicted probability $\mathbb{P}[Y_U|X]$ to be fairly uniform and therefore have a high entropy. However, deep learning based methods are known to be overly confident (Guo et al., 2017).

I therefore use the prediction entropy of an ensemble of models trained on different subsets of the annotated data as deep ensembles have been shown to be able of identifying uncertain predictions more accurately (Abdar et al., 2021; Caldeira et al., 2020; Czolbe et al., 2021). I expect such capability to be even stronger in the setting at hand as it can be exploited that the annotated training data also includes ambiguous areas. This allows to create a diverse set of models within the ensemble as individual models settle on different associations between ambiguous areas and the segmentation labels during their training process.

Moreover, this method aligns well with the design of the method as a post-processing step since it does not require to modify the original segmentation algorithm.

To capture the disagreement among models in the ensemble the algorithm computes the average prediction $\hat{p}_i^c$ over the models for class c and pixel i. Subsequently, the Shannon entropy $H_i = -\sum_{c=1}^{C} \hat{p}_i^c \log(\hat{p}_i^c)$ is computed and normalised into $[0, 1]$ as our indicator of the associated uncertainty.

Segmentation Prior

Once the ambiguous areas in the overall image with potential inconsistent segmentations are identified, these pixels are predicted based on their surroundings following $\mathbb{P}(Y_i|Y_{-i})$ instead of $\mathbb{P}(Y_i|X)$.

The aim is to learn an approximation to $\mathbb{P}(Y_i|Y_{-i}) = \int \mathbb{P}(Y_i|x, Y_{-i})\mathbb{P}_V(x)dx$ as a prior for what biologically valid segmentation masks look like by integrating $\mathbb{P}(Y_i|x, Y_{-i})$ over the distribution of biologically plausible image patches $\mathbb{P}_V(x)$. The intuition behind this approach is that the segmentation masks of the mitochondria show a high degree of local self-similarity. Therefore, given a large pool of mitochondria, it is possible to learn a model that is able to reconstruct small missing areas coherently from its surroundings.

Inference: While MAEs were originally designed as a self-supervised pre-learning technique, I use it as a generative model and apply it directly to the segmentation mask of an initial segmentation model. Specially, after the MAE is trained, I use the predictions of the initial segmentation model to generate a first segmentation mask and identify areas of possible unclear staining based on the ensemble entropy. The initial segmentation of size $W \times H$ is then divided into L smaller patches of size $w \times h$ and filter out patches with a mean uncertainty above some hyperparameter τ_{thres}. The remaining patches are forwarded through the MAE to reconstruct the uncertain patches and filled in to produce the final segmentation as shown in Figure 5.3.

This approach is motivated by the observation that MAEs are able to act as very efficient generative models (Wei et al., 2023). Moreover, while implementing this method with a convolutional neural network needs careful consideration in order to avoid information leakage (Liu et al., 2018) from the masked areas, the

architectures of a vision transformers (ViT) (Dosovitskiy et al., 2021) and masked autoencoders (MAE) (He et al., 2022) align very closely with our requirement of ignoring specific areas in the input and allow for a straight forward implementation.

Training: The strategy to train the MAE for $\mathbb{P}(Y_i|Y_{-i})$ is to first randomly mask out areas and reconstruct the clearly stained areas from the remaining pixels. Given a masking ratio ρ, a random subset of $\rho \cdot L$ of all patches are removed and only the randomly selected patches are augmented by a positional encoding and then forwarded through an encoder model h. Moreover, I want to encourage the model to be able to work with a variety of masking ratios as the size of the entropy mask in the inference step varies and therefore sample $\rho \sim \mathcal{U}[0.1, 0.9]$ uniformly.

The encoder is a ViT which first transforms the patches into a latent feature $z_0 \in \mathbb{R}^{d_z}$ which subsequently is passed through a series of m multiheaded self-attention, MLP and normalisation blocks to produce the final latent representation z_m. These features are subsequently concatenated with a learnable mask token for every disregarded patch, the positional encoder is added and forwarded through a decoder build on the same architecture as the encoder.

The output of the decoder is passed through a final linear layer that transforms the output into prediction probabilities $[0, 1]^{|\mathcal{C}|}$. The MAE is then trained to match the original segmentation as closely as possible in terms of the cross-entropy loss. In order to ensure that the model only learns coherent segmentations, the loss function only includes pixels i which have an entropy value H_i smaller than some threshold τ_{train} while the labels of uncertain pixels are ignored. Additionally, pixels which are too far from any mitochondrion are ignored in order to prevent the loss from being dominated by background pixels.

Comparing our method to graphical models like CRFs, the encoder based on the self-attention blocks can be interpreted as constructing a global graph between all present patches. The random masking scheme therefore corresponds to sampling random subgraphs and learning how to complete them. However, in contrast to CRFs in which the interactions between different vertices needs to be handcrafted,

the MAE allows to learn those dependencies automatically from the data in order to be able to reconstruct the missing patches.

5.2.4 Experiments

Data

I apply the method to panExM images (M'Saad et al., 2020) described in Chapter 3 aiming to segment the double membrane structure of the mitochondria. In the segmentation figures, the *intermembrane space* between the outer and inner membrane is depicted in blue and presents itself as a dark halo around the mitochondria in the global staining. Moreover, the inner membrane forms invaginations to increase its surface area which are called *cristae* and shown in green in the figures. Finally, the space enclosed by the inner membrane is called *matrix*, shown in yellow in the figures and can be recognised by its higher intensity in the global staining. For the sake of clarity, I ignore the ambiguous class in this section.

For the experiments I use twelve image stacks of size 2048×2048 pixel with a z-stack depth between 74 and 234. Each stack shows parts of one HeLa cell and contains on average 150 mitochondria. While the vast majority of the mitochondria are clearly stained as shown in Chapter 4, this work focuses on a sub-sample of mitochondria which are partly effected by some ambiguous staining. I therefore use all of the available data for training but focus the evaluation of the reconstruction algorithm on a separate set of 100 patches of size 128×128 which visually show clear anomalies. Moreover, I select an additional set of 100 control patches which show coherent segmentation masks which allows to compute baseline values for different metrics to compare the reconstructed samples with.

The initial segmentation and entropy masks are provided by an ensemble of 3D U-Nets from Chapter 4.

Implementation Details

The overall images considered are of size 128×128 and the sub-patches have size of 4×4 pixels. The encoder of the MAE has a latent dimension of 128 and a depth of

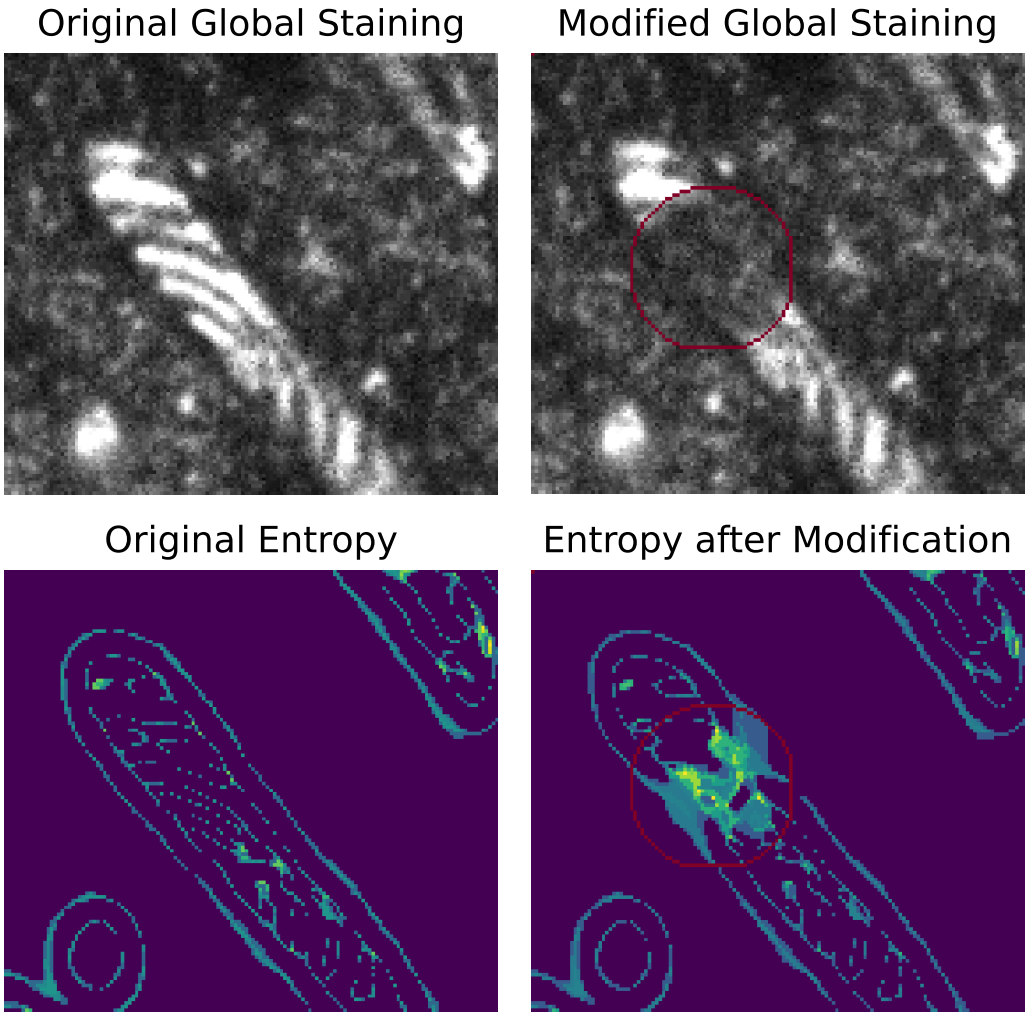


Figure 5.4: Simulation of unclear staining: The top row shows a global staining patch before and after degrading the imaging quality in the area highlighted by the red circle. The bottom row shows the corresponding ensemble entropy with brighter colours corresponding to higher values. The quality degradation leads to a substantial increase in the predicted uncertainty in the modified area. ©2024 IEEE

14. Following the original MAE implementation (He et al., 2022), I use a lighter decoder with a latent dimension of 64 and a depth of 7. Both models use a twice as many latent dimensions for the hidden layer of the MLP.

For the identification of reliable areas in the training a threshold of $\tau_{train} = 0.1$ is used while the threshold for the inference step is set at $\tau_{thres} = 0.4$.[2]

[2]An implementation of the method is available in a separate repository at Github `https://github.com/AlexSauer/BioMaeRef`.

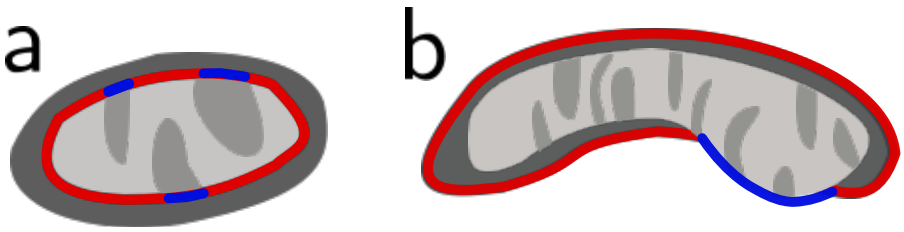


Figure 5.5: Validity metrics: a: The 'Cristae Boundary Ratio' $\frac{\text{blue}}{\text{blue + red}}$ is the proportion of *cristae* (blue) in the boundary between the inner structure (*cristae* + *matrix*) and the *intermembrane space* (blue + red). b: The 'Inner Membrane Background Ratio' $\frac{\text{blue}}{\text{blue + red}}$ refers to the proportion of the boundary between the segmentation mask and the background (blue + red) which does not consist of *intermembrane space* (blue)). ©2024 IEEE

Results

Identifying Uncertain Area: In order to test the assumption that poorly imaged areas in the global staining correspond to an increase in the ensemble entropy, I conduct experiments in which I artificially degrade the imaging quality for some areas and measure the change in the prediction consistency as shown by Figure 5.4.

Specifically, I create a mask along the skeleton of the mitochondrion which is indicated by a red circle in Figure 5.4. I then take a patch consisting purely of background and compute the linear interpolation between the original mitochondrion, some Gaussian noise and the background to simulate a situation with unclear global staining. This interpolation is weighted by the distance map of the mask in order to ensure a smooth transition between the original image and the added background. The induced change can be seen in the top row of Figure 5.4.

The lower row of Figure 5.4 shows a substantial increase in the entropy in the corresponding area. To quantify the effect in more detail, I repeat the above procedure for 100 mitochondria and find an increase of 160% in the entropy associated with these areas and it can be seen that areas that were in the lowest 10% of entropy original are within the top 60% of entropy after the modification. Additionally, Figure 5.6 shows how areas of poor signal in the global staining are related to areas of inconsistent initial segmentations and are reliably picked up by the entropy mask.

Table 5.1: Reconstruction accuracy: Dice scores of the reconstruction for clearly stained regions using different ratios of randomly masked patches. ©2024 IEEE

Masking ratio	0.1	0.25	0.5	0.75	0.9
Dice Score	0.97	0.97	0.95	0.92	0.86

Table 5.2: Segmentation evaluation with respect to biological constraints: The 'Cristae Boundary Ratio' and 'Inner Membrane Background Ratio' are explained in Figure 5.5. The different methods are evaluated on a subset of 100 mitochondria with visually implausible segmentation masks. Additional a control group with plausible segmentation masks was evaluated to provide baseline values. Standard errors are reported in parenthesis. ©2024 IEEE

Method	Skeleton End Points	Skeleton Cycles	Boundary to Area Ratio (%)	Cristae Boundary Ratio (%)	Inner Membrane Background Ratio (%)
Init. Segmentation	10.90 (0.60)	2.93 (0.37)	14.21 (0.59)	17.50 (1.19)	0.00 (0.00)
Difference to Control	7.35	2.67	2.07	7.41	0.00
CNN AE	5.51 (0.30)	0.08 (0.04)	11.39 (0.22)	2.98 (0.34)	0.76 (0.28)
Difference to Control	1.96	-0.18	-0.75	-7.11	0.76
CRF	6.14 (0.34)	0.81 (0.15)	12.95 (0.49)	13.46 (1.06)	8.46 (1.48)
Difference to Control	2.59	0.55	0.81	3.37	8.46
MAE+X	8.64 (0.50)	1.02 (0.16)	13.13 (0.64)	15.31 (0.87)	0.65 (0.18)
Difference to Control	5.09	0.76	0.99	5.22	0.65
MAE (ours)	**5.05 (0.29)**	**0.18 (0.07)**	**11.41 (0.39)**	**8.71 (0.58)**	**0.36 (0.03)**
Difference to Control	**1.50**	**-0.08**	**-0.73**	**-1.38**	**0.36**
Control Group	3.55 (0.30)	0.26 (0.14)	12.14 (0.94)	10.09 (0.88)	0.00 (0.00)

Reconstruction of plausible areas: While the previous experiment suggests that the approach has a good performance of correctly identifying areas of unclear staining and therefore possible implausible segmentations, it might also accidentally create false positives and flag areas which have been segmented consistently. I therefore test the reconstruction quality of my method for plausible segmentation masks by randomly dropping patches from the control group of plausible mitochondria masks and compute the resulting reconstruction accuracy. Table 5.1 shows that even under high masking ratios the model is able to reconstruct accurate segmentation mask due to the high self-similarities of the biologically valid structures.

Evaluation: In order to evaluate the entire post-processing pipeline shown in Figure 5.3, I focus on several metrics that approximate whether the segmentation mask adheres to the biological constraints: As previously discussed, a biologically valid segmentation mask should show most of the *cristae* area surrounded by *matrix* area and should include some *intermembrane space* between the inner

structure and the background. To evaluate the plausibility of the *cristae*, I therefore compute the length of the boundary that the *cristae* forms with the *intermembrane space*. To account for mitochondria of different sizes I compute the ratio of this value to the length of the overall boundary that the inner structure forms with the *intermembrane space* and call this metric **cristae boundary ratio** (Figure 5.5a). This metric is higher if the segmentation exhibits independent areas of *cristae*. Additionally, I compute the length of the boundary between any inner structure and the background and compute to ratio to the length of the overall boundary length between the mitochondria mask and the background. I call this metric **inner membrane background ratio** and it captures areas of missing *intermembrane space* (Figure 5.5b).

Moreover, in order to capture the smoothness of the mask, I calculate the ratio between the boundary length over the area and look at the number of branches that a skeleton of the inner part of mitochondrion forms.

In order to interpret these values, I additionally compute these metrics for a control group of biologically consistent segmentation masks that serve as a baseline for the method evaluation.

Comparison to other methods: I compare the presented methods against CRFs that are based on mean field approximations (Krähenbühl et al., 2011) to allow for long range dependencies and global convolutional autoencoders (CNN AE) similar to Larrazabal et al. (2019). The results are shown in Table 5.2 and show that the segmentation masks by my method bring the described metrics closest to the metrics of the control group of biologically valid segmentation masks. CRFs are able to smooth out the segmentation boundaries but fail to reproduce biological constraints as seen by the high inner membrane background ratio which indicates that it often misses some *intermembrane space*. The CNN AE on the other hand fails to reconstruct the cristae accurately. Figure 5.6 shows a visual comparison of the methods.

Test the validity of $\mathbb{P}(Y_i|Y_{-i})$: Section 5.2.3 factorised $\mathbb{P}[Y|X]$ into the product $\mathbb{P}[Y_i|X, Y_{-i}]\mathbb{P}[Y_{-i}|X]$ and argued that for poorly stained regions $\mathbb{P}[Y_i|Y_{-i}]$ might be a

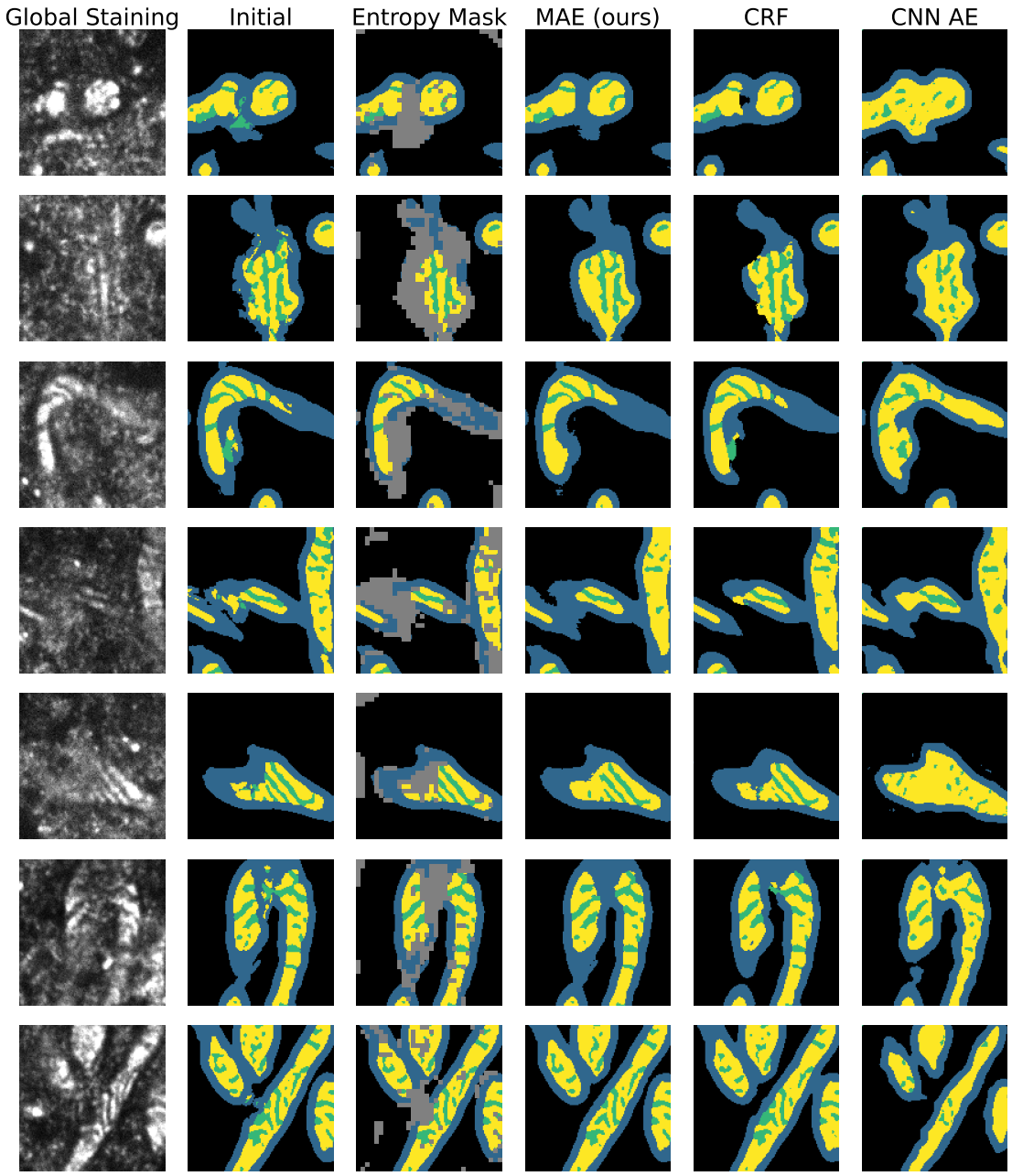


Figure 5.6: Reconstruction results: Visual comparison of the initial segmentation, the patches masked based on high entropy levels, the reconstructed segmentation masks and the global staining that the initial prediction is based on. It can be seen that the MAE reconstructed masks better respect the biological constraints like no *cristae* area (green) that is detached from *matrix* area (yellow) and less fragmented boundaries. ©2024 IEEE

better target than $\mathbb{P}[Y_i|X]$, which is given by the initial segmentation method. In this section I compare how the estimate of $\mathbb{P}[Y_i|Y_{-i}]$ compares to $\mathbb{P}[Y_i|X, Y_{-i}]$. In order to approximate $\mathbb{P}[Y_i|X, Y_{-i}]$, I modify my method by attach an additional encoder with X as the input. This encoder follows a standard ViT encoder that takes the global staining as input without any masking and produces a latent representation for every patch. This representation is then added to the latent representation of the segmentation mask if the patch was present or to the mask token if it was not. Subsequently, the updated representation is forwarded through the decoder to produce the final segmentation mask. In this way, I allow the method to be able to access the global context X as well as the surrounding initial segmentation masks Y_{-i} and I follow the same training and inference procedure outlined in Section 5.2.3.

Comparing the results in Table 5.2, in which the method is labelled MAE+X, shows that adding the additional context deteriorates the performance. I speculate that the model focuses too much on reproducing the association between the initial segmentation mask Y_i and the input X rather than capturing the interactions with the surrounding neighbourhood Y_{-i} during the training.

5.2.5 Postprocessing

For the following analysis, I apply the 3D version of the presented reconstruction algorithm to the control group of the HeLa cells. For the drug treated cells, I work with the raw prediction to prevent a possible bias of the results as part of the analysis is to pick up irregularities induced by the drugs.

Moreover, I apply additional postprocessing steps like removing small isolated pixels in the segmentation masks. On top of this, I remove separated components which include no intermembrane space or consist solely out of intermembrane space as this is biologically not possible.

Finally, I remove segmented areas which are very fragmented as this is a sign of segmentation mistake. Post-processing the model prediction represents a trade-off as being to aggressive can lead to an increased number of false negatives. Depending

on the application, different thresholds and strategies can be adopted and their influence on the subsequent analysis compared to allow for robust results..

5.3 Parameter Extraction

5.3.1 Instance Segmentation

Based on the NHS input image, the model predicts one label for every pixel. In order to be able to make statements about individual mitochondria, it is necessary to identify the separate mitochondria in the resulting segmentation mask. Instance segmentation is an active field of research (Zou et al., 2023) with several adoption for biological contexts like nuclei or cell segmentation (Stringer et al., 2021). These methods often approach the problem of identifying two separate instances which have their borders touching by imposing some shape constraint (Stringer et al., 2021).

In the present case however, biological information can be leveraged for the instance segmentation: Because of the high resolution it is possible to identify and segment the inner structure and membrane of an mitochondrion which is separated by the outer membrane and the intermembrane space. Therefore, I make the assumption that two mitochondria which are touching are two separate instances as long as the inner part of the mitochondria are not touching but are separated by some intermembrane space.

Computationally, I therefore identify all connected components of inner structure and marks those as individual instances. These instances extend their label then to the surrounding intermembrane space based on a watershed algorithm (Walt et al., 2014) as visualised in Figure 5.7.

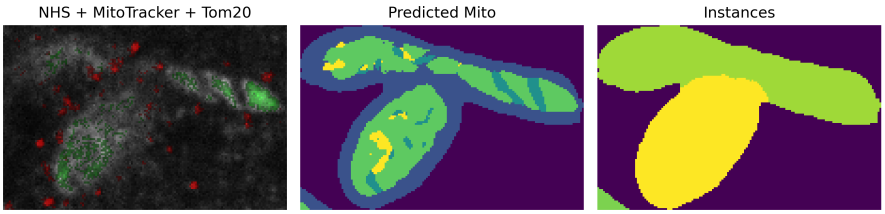


Figure 5.7: Instance separation: Visualisation of how touching but separate mitochondria instances are identified based on their inner structure.

5.3.2 Skeletonisation

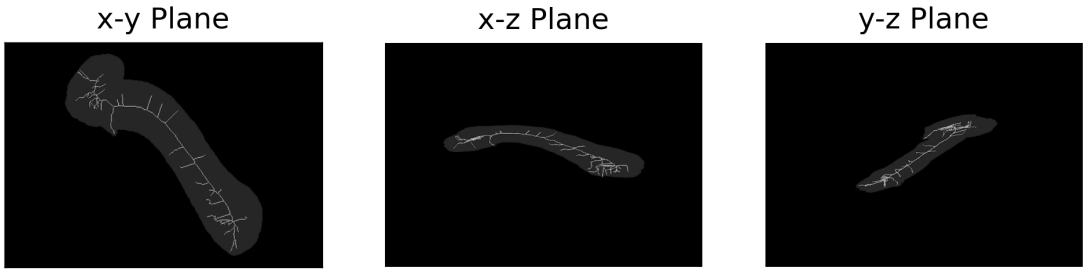
In order to analyse key parameter of the mitochondria like the length or diameter, a skeleton of the segmentation outlines is required. Based on these skeletons, it is subsequently possible to approximate the lengths of the mitochondrion by the lengths of the corresponding skeleton and base the diameter calculation on the distance from the outer boundary to the skeleton.

Popular image processing packages like skimage (Walt et al., 2014) implement the classical algorithm proposed by (Lee et al., 1994) but 3D skeletonisation is still an active area of research (Tagliasacchi et al., 2016). Lee's algorithm works by repeatedly passing over the binary mask and remove pixels which do not break the connectivity until the image stops changing. While this procedure closely resembles the intuitive idea of what the skeletons should look like and match the topology of the mask, I observe major artefacts when apply the algorithm to our mitochondria data.

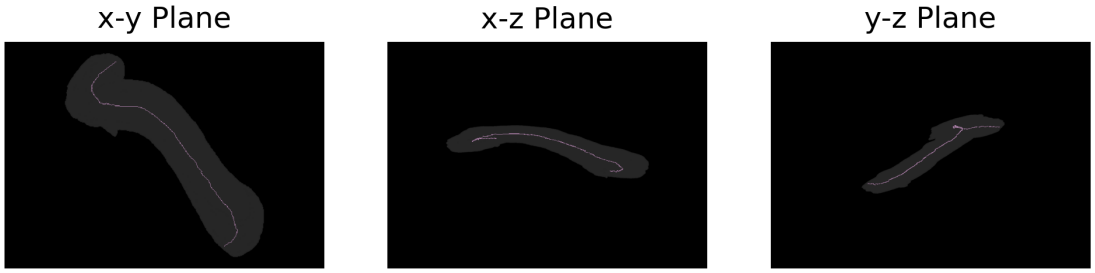
In the following paragraphs, I highlight several of these artefacts and demonstrate how I modify the skeletons.

Most prominently, the skeletons produced by skimage produce a large number of smaller side branches as seen in Figure 5.8. As a first step, I therefore prune these branches. In order to do so, I identify the leaf nodes and check whether its distance walking along the skeleton is shorter than some threshold t and remove it if it is. However, because several side branches consist of several branches, I repeat this procedure several time with increasing thresholds. The chosen thresholds are $t \in [10, 20, 25]$ and subsequently the algorithm is repeated with $t = 25$ until the skeleton does not change further. However, it is important not to accidentally remove the main branch of the skeleton which is identified as the two leaf nodes with the furthest distance from each other.

Moreover, it is observed that the skeletons do not extend until the boundary of the mitochondria outline. This is because the initial skeletonisation is based only on the inner structure of the mitochondria (excluding intermembrane space) and because the skimage algorithm also carves away some prediction in the process.



(a) Skeleton produced by Skimage



(b) Skeleton after branch removal

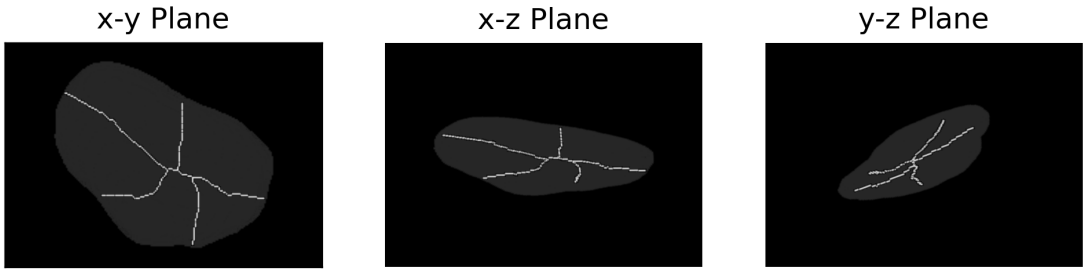
Figure 5.8: Skeletons produced by the standard skimage (Walt et al., 2014) algorithm which is shown in the top row produce a lot of additional branches. The pruned skeleton is shown in the bottom row.

In order to reliably measure the length, the ends of the mitochondria are therefore extended until it they reach the boundary of the outline by extrapolating the direction based on the last 10 skeleton pixel.

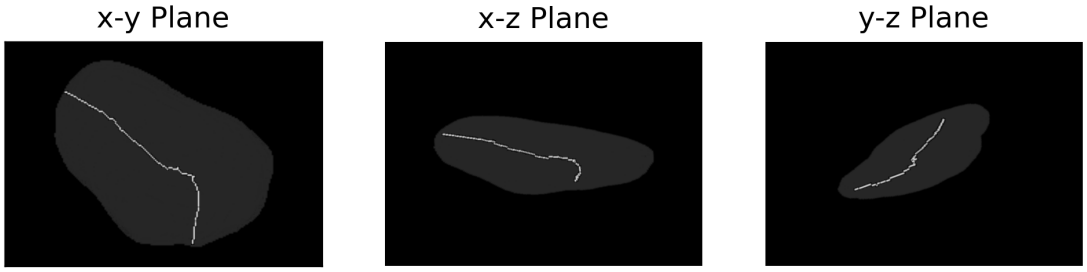
The result is shown in the second row of Figure 5.8 which validates the approach.

However, as seen in the upper row of Figure 5.9 round mitochondria can lead to a situation in which there remain be several large branches that are longer than all thresholds in the previous step but using larger thresholds deteriorate smaller mitochondria. In order to identify and remove these branches, connecting lines between each leaf node are drawn. If this line is completely contained within the segmentation mask, I remove the resulting leave. Again, no leave that corresponds to the main branch of the skeleton is removed.

Additionally, it is observed that some of the annotations contain missing pixels, both in ground truth from the annotator and (more rarely) in our model predictions. These missing pixels change the genus in a topological sense of the segmentation mask and lead to artefacts in the form of shallow balls or cycles in the skeleton



(a) Long convex branches



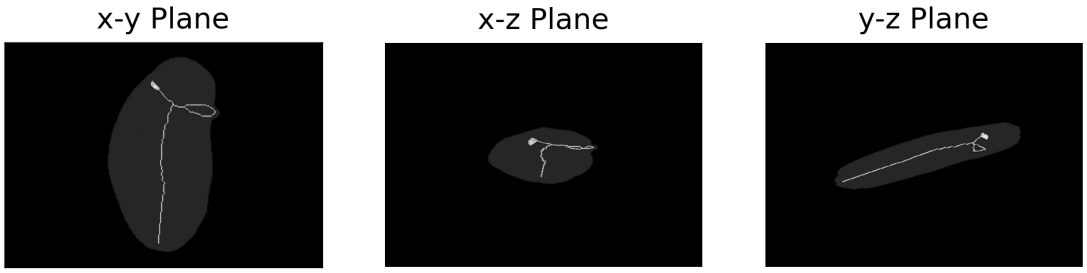
(b) After branch removal

Figure 5.9: Additional skeleton pruning by deleting branches for which the connecting lines between leaves are completely contained within the segmentation mask.

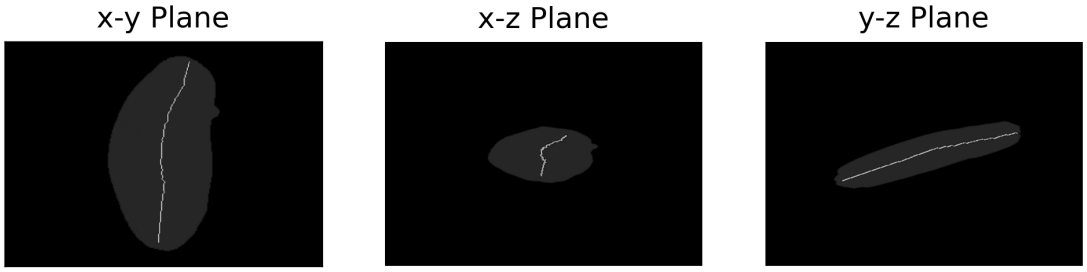
graph as shown in the first row of Figure 5.10. In order to fill small holes, I dilate the binary segmentation mask before the skeletonisation but often this turns out not to be sufficient. Therefore, I search through the skeleton graph for dense areas and identify these dense areas by computing the pairwise distance between skeleton nodes and remove areas in which clusters are found. When it comes to thin cycles, it is important to differentiate whether they actually represent the genus of the mitochondria shape or if they stem from a segmentation mistake. To make this decision, several connecting lines between the member nodes of these cycles are drawn. If all of these lines lie completely within the segmentation mask, the cycle is disregarded and removed it from the skeleton.

5.3.3 Length and Diameter Calculation

Previous work that quantitatively analyses mitochondria (Fischer et al., 2020; Lee et al., 2013; Lefebvre et al., 2021; Picard et al., 2013) has often based the calculation of the length and diameter of a mitochondrion on crude approximations: Given a mitochondrion, the direction of its largest extension was calculated and interpreted



(a) Skeleton with artefacts from missing pixels



(b) Skeleton after processing.

Figure 5.10: Missing pixels in the segmentation mask can lead to either hallow balls with a dense surface or cycles which are removed.

as the length and the maximal extension orthogonal to this was seen as its diameter. While these approximation might be reasonable in the 2D case of a textbook like elliptical mitochondrion, it fails ungracefully in the face of the large shape variations observed in 3D as seen in Figure 4.1. Therefore, the presented calculations are based on the skeletons described in the previous section.

When it comes to calculating the length from the skeletons, it is important to note that it is not sufficient to just calculate the number of skeleton voxels. Rather it is important to account for the fact that when walking along the skeleton, the next pixel can be up to $\sqrt{3}$ units away from the previous one if it lies diagonally to it. This becomes even more important when accounting for the anisotropy of the data as a step in the z-plane represents $0.25\mu m$ compared to $0.1\mu m$ in x and y.

Regarding the diameter, one simple approach to calculate the diameter is to approximate the shape of a mitochondrion by a cylinder which allows to easily compute the diameter as $2\sqrt{\frac{V}{\pi l}}$ where l denotes the lengths and V the volume. However, this might lead to biased results if the assumption is not fulfilled and

having access to the skeletons previously described, I calculate the average diameter of a mitochondrion by measuring for every point on the segmentation boundary the distance to the closest point on the skeleton and form the average.

Nucleus and Cytoplasm detection

Because mitochondria are dynamic in the cell and can grow or shrink it is interesting to investigate how much of the cytoplasm is occupied by mitochondria. However, because the field of view of the microscope is not large enough to cover the whole cell, every image stack shows a different composition between background, cytoplasm and nucleus area.

Therefore, estimating the volume of cytoplasm occupied by mitochondria first requires us to produce reliable estimates of the cytoplasm volume. Initial hopes that this problem could be quickly solved by simple clustering algorithms that sort the pixel intensities or thresholding were disappointed by the large variation seen in the NHS intensity and the inconsistent amount of nucleus area and nucleoli within the nucleus. Instead I had to resort to training another segmentation network that allows to segments the background, nucleus and cytoplasm.

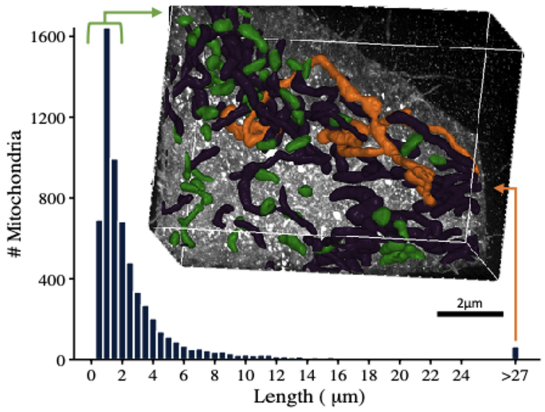
5.4 Discussion

In this chapter, I have proposed a post-processing framework which is capable of correcting biologically invalid segmentation masks. The framework combines deep ensembles in order to identify areas of poor signal and leverages the strength of MAEs to learn repetitive patterns and selectively reconstruct the segmentation masks for these regions. The methods has the advantages that it infers the structure of biologically plausible segmentation masks directly from the data and it can be applied as a post-processing step independently of the underlying primary segmentation algorithm. Moreover, the reconstruction based on the MAE requires a single forward pass and is therefore able to scale to large data volumes.

Limitations

One challenge associated with the presented approach is that the use of strong priors can detach the prediction to some degree from the observed data. This implies that some of the segmentation masks which are corrected by the MAE are actually false positives and should rather be removed completely. However, since it is not possible to reliably differentiate between false and true positives, I argue that it is still beneficial to modify the false positives so that their influence on any subsequent analysis is limited. While I demonstrated the effectiveness of the approach with the example of mitochondria segmentation and expansion microscopy, it would be exciting to see future work apply the proposed method to other segmentation tasks in which the biological structures show similar large degree of degree of self-similarity and biological constraints.

With regard to the skeletonisation, some incorrect branches might not have been removed which could bias the calculation of the length but a visual inspection did not indicate that this is the case. Moreover, I limit the analysis to shape descriptions that have a clear interpretation like length and diameter instead of trying to capture the shape in more abstract terms like the Kendall shape space or elastic shape space that are based on landmarks or parametrised curves (Zhang et al., 2022). The decision was made in consultation with our biologist collaborators, as defining these shape spaces in 3D for objects as variable in size and shape as the observed mitochondria is highly complex and complicates the subsequent communication of results.



6

Application to Biological Studies

Contribution

This chapter finalises the analysis pipeline by quantitatively investigating the segmented mitochondria on a large scale.
Starting with the control group, I analyse the shape and ultrastructure of mitochondria based on over 6,000 mitochondria in HeLa cells before investigating the difference induced by two different drug treatments based on over 35,000 mitochondria. Despite the large scale of the analyse, I demonstrate how every measurement can be traced back to the underlying mitochondrion which can be visualised and inspected in 3D.
Additionally, I extend the analysis to mouse kidney tissues and quantify the difference in mitochondria morphology between healthy tissue and mitochondria from cells with acute kidney injury based on over 24,000 mitochondria which reveals substantial changes in the cristae ultrastructure.
Finally, I demonstrate how the segmentation masks enable the analysis of additional biomarkers and their localisation in the three dimensional context of the mitochondria.

Contents

6.1 Mitochondria Analysis in HeLa Cells 124
6.1.1 Global Descriptors 125
6.1.2 Ultrastructure Analysis 129
6.1.3 Statistical Analysis 133
6.2 Mitochondria in Kidney Tissue 135
6.3 Immunolabeling Analysis 136
6.4 Discussion 141

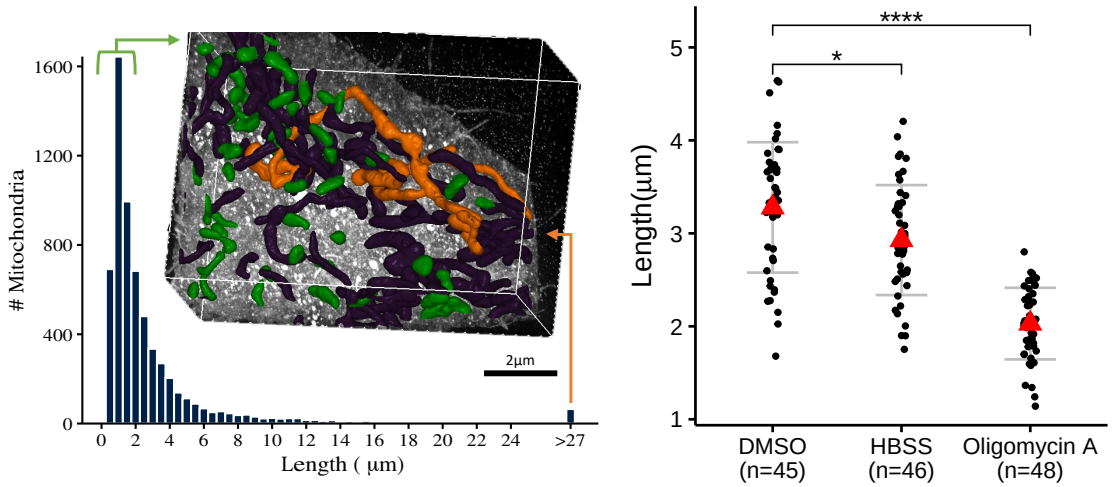
6.1 Mitochondria Analysis in HeLa Cells

In this section, I provide a quantitative analysis of the mitochondria in HeLa cells based on the extracted features to compare the effect of the different drug treatments. Therefore, I first describe the distribution of the observed feature for a control of 30 non-treated HeLa cells containing 6344 mitochondria and subsequently describe how the drug treatment alters that distribution compared to the DMSO treated group.

To ensure reproducibility, the experiment of treating the drugs was repeated three times resulting in three different batches. In the following figures, all batches are included resulting in 46 DMSO files with 8067 mitochondria, 46 HBSS files containing 12,397 mitochondria and 49 Oligomycin A files with 15,329 mitochondria. To keep the discussion concise and because no significant differences were found, Antimycin is excluded from the following discussion.

When investigating the distribution of the control group, the data is viewed at the mitochondrion level, but when comparing the effect of drugs I aggregate the results via the mean on a per cell level and compare the cells with each other. In all drug comparisons, the mean value over all cells per drug is shown in red and one standard deviation as error bars. The plots also show the significance in the shift of the distribution as measured by a Mann-Whitney test (which gives the same result as an unpaired t test in all reported cases) on the cell level but a more detailed statistical analysis is provided in Section 6.1.3. In the plots, the corresponding p-value is indicated by * for $p < 0.05$ and **** indicating $p < 0.0001$. All presented measurements in this chapter are adjusted by the expansion factor and the anisotropy of the data to represent the true underlying dimensions of the mitochondria.

In the analysis, every mitochondrion is assigned a unique ID which allows to trace back every abstract data point to a concrete mitochondrion. Therefore, it is always possible to verify the measurements and go beyond a simple number by visualising their meaning on a specific example. I demonstrate this capability by not only providing histograms of the metrics but augmenting them with examples of mitochondria that visualise what the corresponding instances look like.



(a) Control Group **(b)** Comparison between drugs

Figure 6.1: Length: (a) Visualisation of the distribution of diameter values in the control group. Mitochondria with a length smaller than 1.5µm are highlighted in green and mitochondria with a length greater than 27µm are shown in orange. (b) Both drugs induce a significant reduction in the average length of the mitochondria with a much stronger effect created by Oligomycin.

6.1.1 Global Descriptors

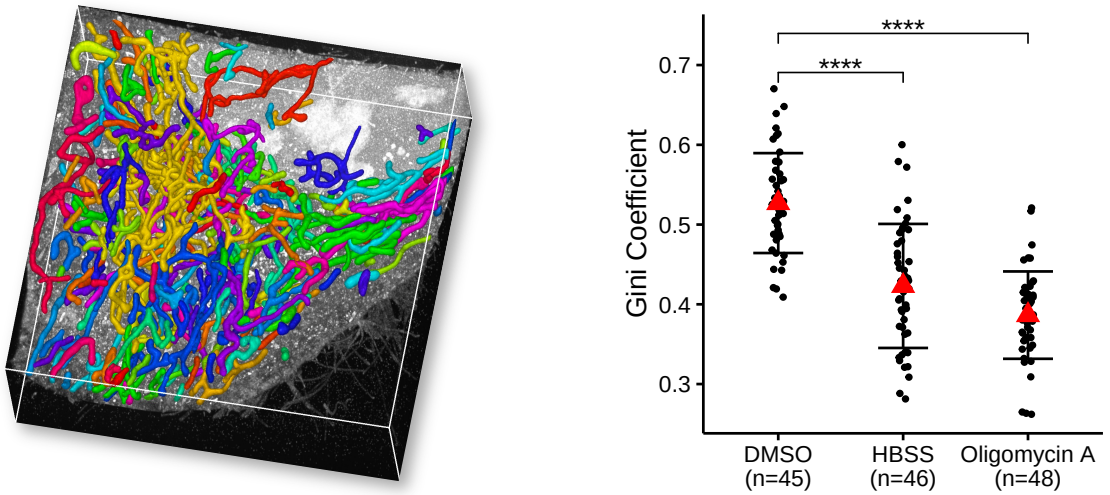
The next section starts discussing several parameter which describe the overall, global structure of the mitochondria before focusing on the ultrastructure in section 6.1.2.

Length

The mitochondria in the control group show a length of 3.27µm on average with a highly skewed distribution shown in Figure 6.1(a). The 10%, 50% and 90% quantiles are 0.7µm, 1.65µm and 10.4µm respectively and Figure 6.1(a) is capped at the 99% quantile with a value of 27µm.

Figure 6.1b shows that both drugs drugs significantly reduce the average length of the mitochondria with a stronger effect for Oligomycin as the effect for HBSS is only borderline significant and Section 6.1.3 reveals that this effect is not robust if batch and cell effects are included.

Regarding the right tail of the distribution, a number of very long outliers is present which represent networks of connected mitochondria with a length as as high as 239 µm for the maximum. Moreover, on average only 6 ± 4.5 (mean ±



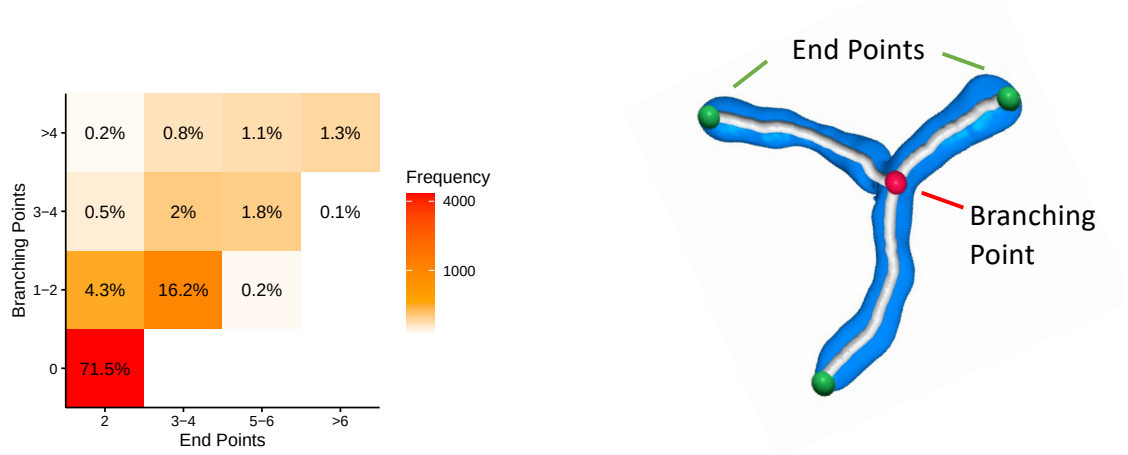
(a) Mitochondria Networks **(b)** Comparison between drugs

Figure 6.2: Mitochondria networks: (a) Example of a cell with one particularly complex mitochondria network in yellow (b) Gini coefficient for the volume to capture large networks. A high Gini coefficient implies that most of the mitochondria volume is concentrated in a few networks.

SD) mitochondria contain 20% of the overall mitochondria volume per cell even though the average number of mitochondria per cell is found to be 211 ± 80 (mean ± SD). Finding these networks by pure visual inspection of the dense images is nearly impossible as these networks only reveal themselves in 3D and following the rich branching structure in a manual way is very challenging.

A large network within one cell means implies that a lot of the mitochondria volume is concentrated in a few mitochondria instances which implies a higher Gini coefficient of the volume per cell. Figure 6.2(b) shows that both drugs are significantly associated with a decrease in the Gini coefficient which implies that there are fewer dominating networks. Potential biases in the analysis of the HBSS samples are discussed in Sections 4.6 and 6.4.

The topology of the mitochondria can be further explored based on the skeletons. Figure 6.3 confirms that the majority of the mitochondria instances have a simple shape without any branching points and two end points. However, only about 38% of all mitochondria volume is contained in mitochondria with a single branch.



(a) Control Group **(b)** Comparison between drugs

Figure 6.3: Topology of mitochondria: (a) shows that most of the mitochondria have a textbook like shape with no branching points and two endpoints. (b) illustrates the identified end points in green and a branching point in read on an example mitochondria.

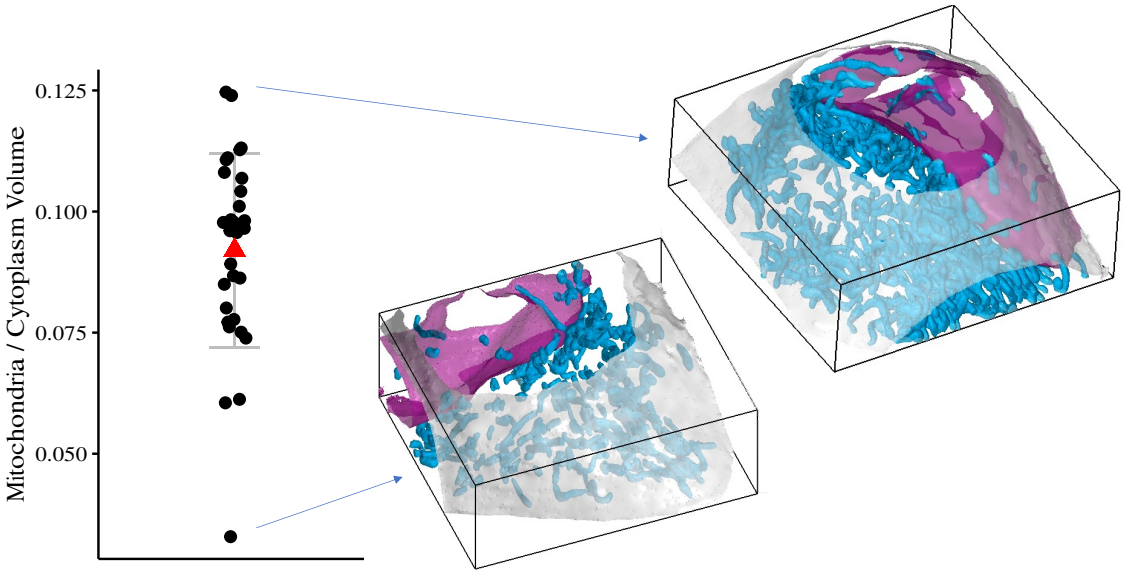


Figure 6.4: Mitochondria volume in relation to the cytoplasm of the FoV. Example 3D renderings of FoVs in which mitochondria occupy 3.3% or 12.4% of the cytoplasm shown. Mitochondria are shown in blue, the cell wall in grey and the nucleus in purple.

Volume

The volume of individual mitochondria instances is largely driven by the length and shows a similar skewed distribution with an average value of $0.13 \pm 0.28 \mu m^3$ (mean ± SD). However, more informative than the mitochondria instance volume might be the fraction of the cell volume which is occupied by mitochondria.

Because the FoVs only cover parts of the cell and include different parts of the nucleus, the computation is reliant on the additional segmentation of the FoVs in background, nucleus and cytoplasm (which includes all other organelles). Figure 6.4 shows that the mean fraction is around 9.2% and 3D renderings on the right visualise the extreme cases with either 3% or 12%. These renderings indicate that the mitochondria density might be higher closer to the nucleus. No significant differences in this metric are found with respect to the drug treatments.

Diameter

In the control group, the diameter is found to lie mostly within a range of 180nm and 370nm which represent the 1% and 99% percentiles and show a mean of 250 ± 40nm (mean ± SD). Figure 6.5a shows the full distribution and a 3D rendered example of thin and thicker mitochondria in green and orange respectively in comparison.

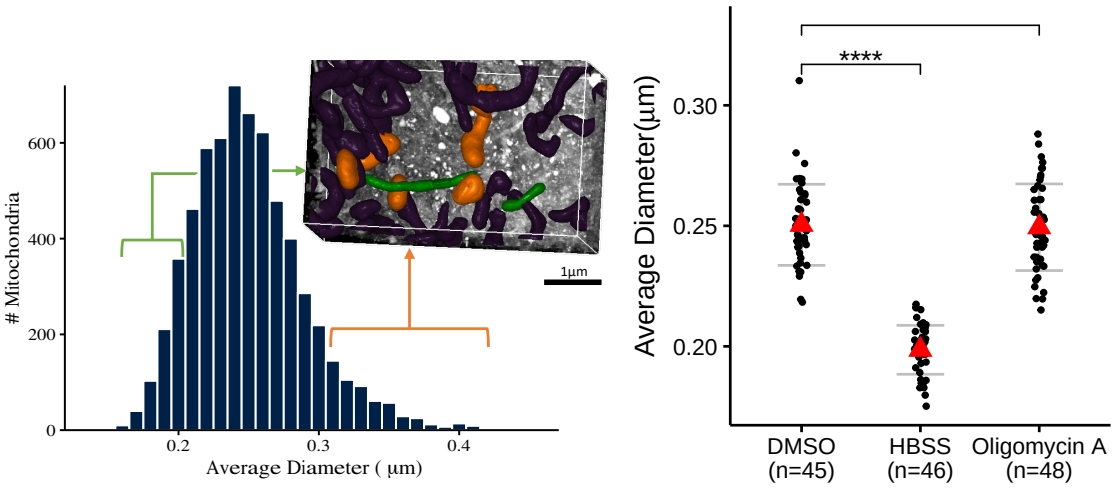
By comparing the effect of the drugs, the visual impression of Figure 4.18 can be quantitatively confirmed that on average the mitochondria shrink down by almost 25% for the HBSS treatment. On the other hand, no significant change is found for Oligomycin.

Sphericity

Additionally, I investigate several other derived metrics like the sphericity (Wadell, 1935) for example which is defined as

$$\Psi = \frac{\pi^{1/3}(6V)^{2/3}}{A},$$

where V is the volume and A is the surface area. The sphericity is a monotonic transformation of the mitochondria complexity index (MCI) introduced by Vincent



(a) Control Group **(b)** Comparison between drugs

Figure 6.5: Diameter: (a) Visualisation of the distribution of diameter values in the control group. Mitochondria with a diameter smaller than 0.2 μm are highlighted in green and mitochondria with a diameter greater than 0.3 μm are shown in orange. (b): HBSS significantly reduces the average diameter of the mitochondria.

et al. (2019) which it relates to as $MCI = \frac{1}{3}\frac{1}{\Psi}$ and therefore only the sphericity is reported. The bivariate plot in Figure 6.6, shows an almost linear relationship ($R^2 = 0.83$ of a linear regression) between the logarithmic length and the sphericity and provides three exemplary 3D renderings to illustrate the corresponding mitochondria.

6.1.2 Ultrastructure Analysis

With regard to the ultrastructure of the mitochondria, I consider both the percentage of each mitochondria occupied by cristae volume and the percentage that could not unambiguously be classified.

Figure 6.7(a) shows the fraction of cristae volume with respect to the overall mitochondria volume which has a mean and standard deviation of 8% ± 4%. If the fraction of only the inner structure and excluding the intermembrane space the value increases to 20% ± 9%. Moreover, the Figure reveals that a high fraction of cristae is often associated with a lamellar shape. Figure 6.7(b) reveals a significant drop in the percentage of cristae across all drugs. For the HBSS treatment, this confirms the strong visual impression. The drop is less severe for Oligomycin and as seen in Figure 6.8(b), this drop in clear cristae is offset by a large increase in

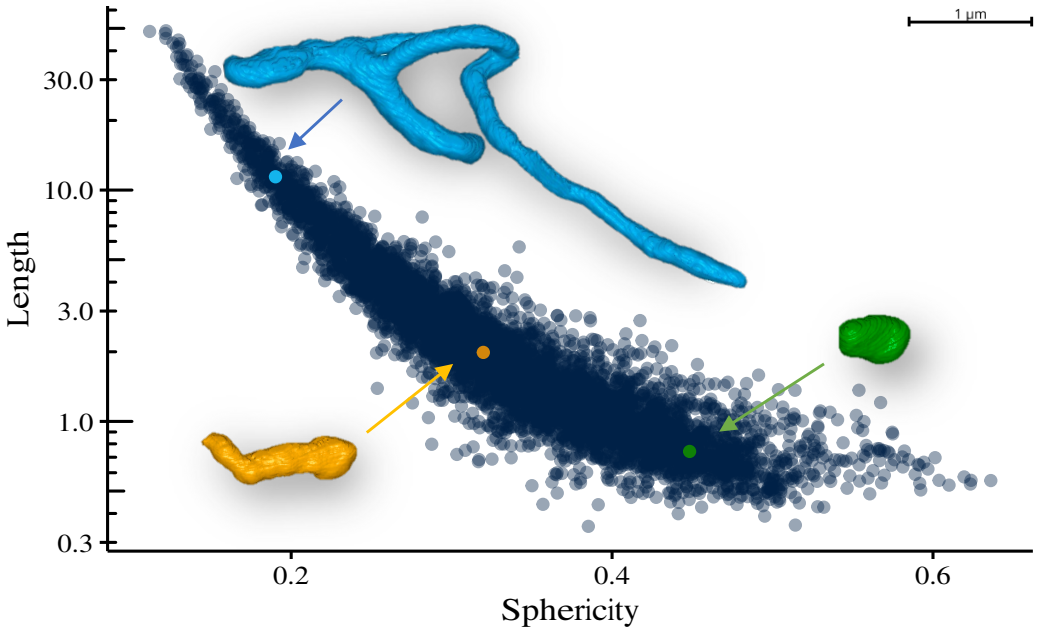
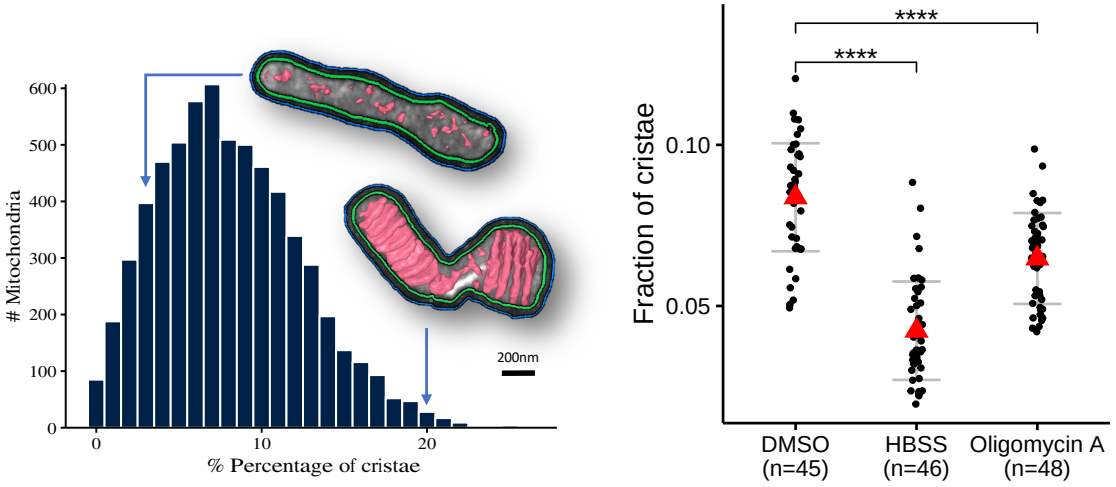
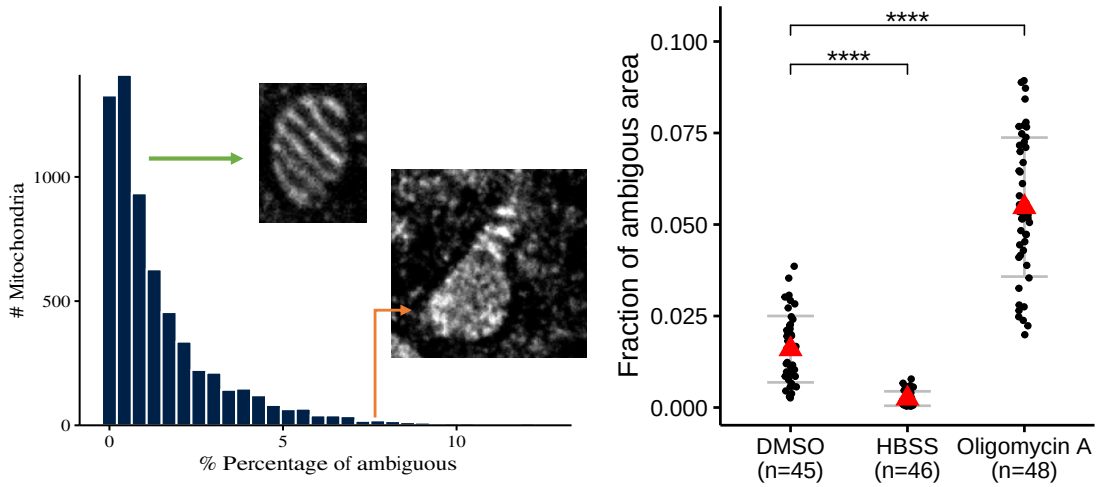


Figure 6.6: Bivariate plot of length and sphericity: Length plotted on a logarithmic scale against the sphericity reveals a close relationship between these metrics.



(a) Control Group **(b)** Comparison between drugs

Figure 6.7: Cristae area: (a) Visualisation of the distribution of cristae percentage in the control group. The top example shows a mitochondrion with 4% of cristae which is sparsely distributed through the mitochondrion. The bottom example shows lamellar cristae structures that account for 20% of the mitochondrion. (b) Both drugs induce a significant drop in the area classified as cristae.

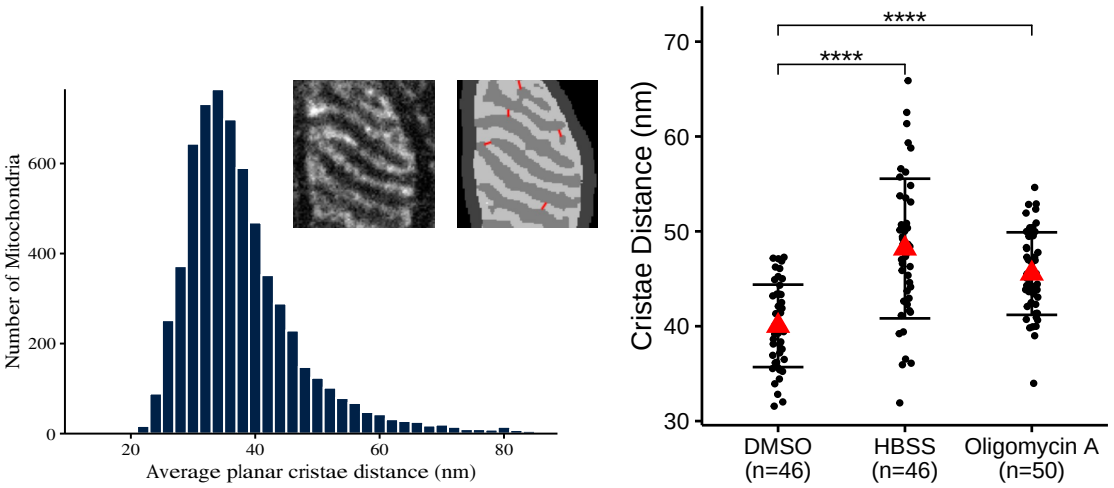


(a) Control Group **(b)** Comparison between drugs

Figure 6.8: Ambiguous area: For Oligomycin a significant increase in the area which cannot be clearly assigned to either matrix or cristae is identified.

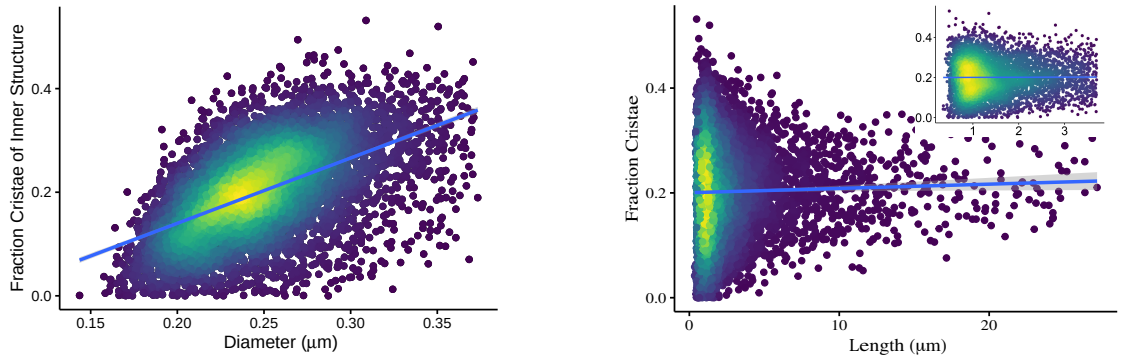
area that cannot be uniquely characterised as either cristae or matrix. For the control group shown in Figure 6.8(a), a mean fraction of of 0.8% with a standard deviation of 1.4% is found. As seen in the provided examples, the ambiguous area shows no clear patterns and I find a significant correlation of -0.12^{***} between the fraction of cristae and ambiguous volume with indicates that these areas might contain smaller cristae structures which cannot be clearly imaged with the currently possible resolution. The significance of this result however might largely be a result of the large sample size as the absolute size of the effect is small.

Moreover, I analyse the distance between cristae components. However, defining what exactly is meant by this is far from trivial, especially in 3D (Wang et al., 2024a). To formalise the statement, I identify individual cristae components in 2D and find for each component its closest neighbour. This allows to define the planar distance as the distance between the two closest points as visualised in top right corner of Figure 6.9(a) in red. The mean and standard deviation are found to be of 39nm and 10 nm respectively with 10% of distances smaller than 29 nm and 10% greater than 50 nm. Figure 6.9(b) shows that both drugs are associated with larger distances which is in line with general decrease in cristae volume shown previously in Figure 6.7.



(a) Control Group **(b)** Comparison between drugs

Figure 6.9: Cristae Distance: (a) The measured planar distance is visualised in red. (b) Both drugs are associated with significant larger distances between cristae components.



(a) Relationship Diameter and Cristae **(b)** Relationship Length and Cristae

Figure 6.10: Bivariate Plots of length and diameter against the cristae fraction: Brighter colours indicate a higher density of observations and the blue line represents a regression line. (a) shows a positive association between the cristae fraction and mitochondria diameter. No similar association is found with respect to the length in (b) which also includes a zoomed in version.

Finally, the combination of super-resolution with a large sample size allows to investigate the relationship between the ultrastructure and macro-parameters like the length or diameter. Figure 6.10 shows a positive relationship between the fraction occupied by cristae per mitochondrion and the diameter. No similar relationship is observed for the length.

6.1.3 Statistical Analysis

To thoroughly quantify the effects of the drug treatments on the mitochondria morphology, I estimate several mixed-effects models. The presented data has been collected in three batches with up to 15 cells per batch. This implies that the observed mitochondria do not represent independent observations but some correlation within each cell and batch should be expected. Mixed-effects model are the appropriate statistical tool to address this challenge (Hox et al., 2017). Denoting the batch by b, the cell by c and each mitochondrion by i, the model to estimate is given by

$$Y_{bci} = \mu + \alpha_b + \gamma_c + \beta_{HBSS} + \beta_{Oligo} + \epsilon_i \tag{6.1}$$

where Y_{bci} represent the metric of interest and $\alpha_b \sim \mathcal{N}(0, \sigma_b^2)$, $\gamma_c \sim \mathcal{N}(0, \sigma_c^2)$ and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ so that α allows for to account for batch effects and γ captures cell effects.

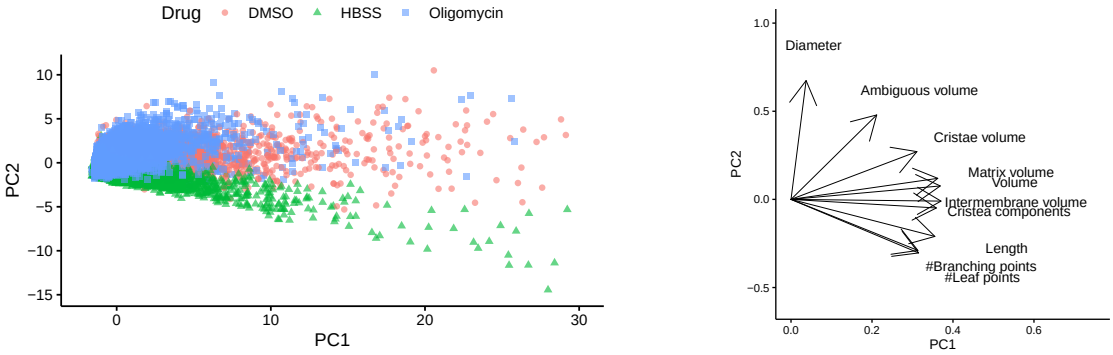
If the target metric shows a highly skewed distribution like for the length and volume, a log transformation is applied to the metric as this allows for multiplicate effects. Moreover, for discrete metrics like the branching points, I estimate a Poisson generalised mixed-linear model with a log-link function.

The results are shown in Table 6.1 and align with the cell based aggregations shown in the previous figures. All of the highly significant results (i.e. $P < 0.001$) are confirmed but no significant difference between the length of the control group and HBSS is detected any more which was borderline significant at the FoV level. Moreover, Table 6.1 provides quantitative estimates for the effects induced by the drug treatments. For example, it is estimated that the mean diameter of 0.25µm for the control group is associated with a decrease by 0.05µm through the HBSS treatment whereas no significant change is found with respect to Oligomycin.

I also visualise all mitochondria in a low dimensional space by conducting a principal component analysis (PCA) based on the features volume, cristae volume, intermembrane volume, matrix volume, ambiguous volume, number of cristae components, length, diameter, number of branching points, and number of leaf points. PCA (Hastie et al., 2009) finds the direction in which the data shows

Table 6.1: Mixed-Effects Analysis: Results based on the regression in equation 6.1 confirms the highly significant results presented in earlier Figures. For each coefficient the estimate, the estimated standard error and the resulting p value is reported.

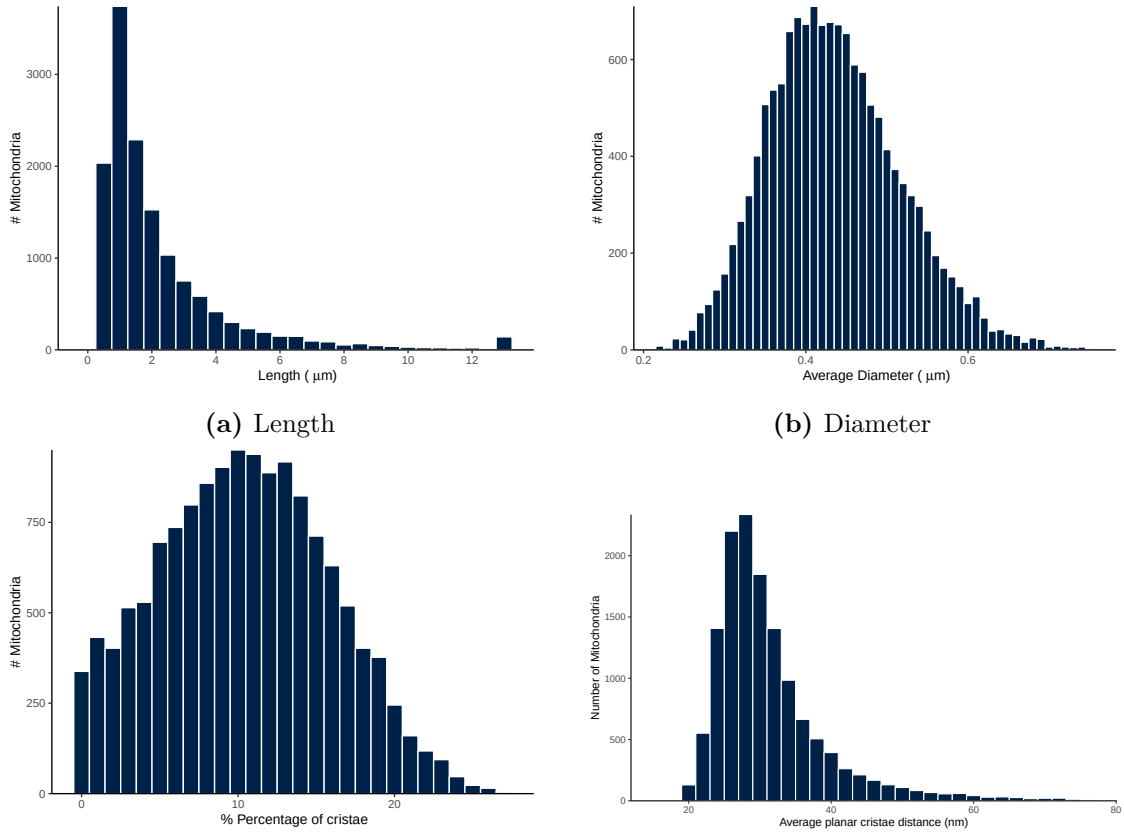
	μ	$\hat{\sigma}_\mu$	p_μ	$\hat{\beta}_{HBSS}$	$\hat{\sigma}_{HBSS}$	p_{HBSS}	$\hat{\beta}_{Olig}$	$\hat{\sigma}_{Oligo}$	p_{Oligo}
Cristae Percentage	0.084	0.005	0.002	-0.042	0.003	< 0.001	-0.018	0.003	< 0.001
Ambiguous Percentage	0.016	0.005	0.088	-0.013	0.002	< 0.001	0.038	0.002	< 0.001
Diameter (µm)	0.250	0.004	< 0.001	-0.052	0.003	< 0.001	-0.002	0.003	0.608
Length per Branch (µm)	1.624	0.090	0.002	-0.091	0.038	0.017	-0.244	0.037	< 0.001
Sphericity	0.328	0.007	< 0.001	-0.024	0.004	< 0.001	0.030	0.004	< 0.001
MCI	5.164	0.199	< 0.001	0.440	0.089	< 0.001	-0.852	0.088	< 0.001
Log(Volume)	-2.456	0.100	0.001	-0.455	0.033	< 0.001	-0.331	0.032	< 0.001
Log(Length)	0.781	0.073	0.005	0.012	0.031	0.687	-0.321	0.031	< 0.001
#Branching points	-0.258	0.054	< 0.001	-0.018	0.056	0.747	-0.681	0.056	< 0.001
#Leaf points	0.912	0.012	< 0.001	0.022	0.011	0.038	-0.103	0.010	< 0.001



(a) PCA Plot **(b)** Biplot

Figure 6.11: PCA plot: Low dimensional representation reveals large overlap between the mitochondria with differences in the tails. The biplot shows that differences are driven by the ultrastructure and mitochondria topology.

the highest variances as the first principal component (PC) and the direction orthogonal to that with the next highest variances as the second PC. The data is then projected into this two dimensional space and the result can be seen in Figure 6.11. It reveals that there is both a substantial overlay between the mitochondria but the drugs also differentiate in the outer regions of the plot. As the biplot in Figure 6.11(b) shows, the change in the mitochondria due to HBSS is mostly driven by the change in diameter.



(a) Length **(b)** Diameter

(c) Cristae Percentage **(d)** Cristae Distance

Figure 6.12: Quantification WT Mitochondria: Histograms showing the distribution of the (a) length, (b) diameter, (c) cristae percentage and (d) average planar cristae distance

6.2 Mitochondria in Kidney Tissue

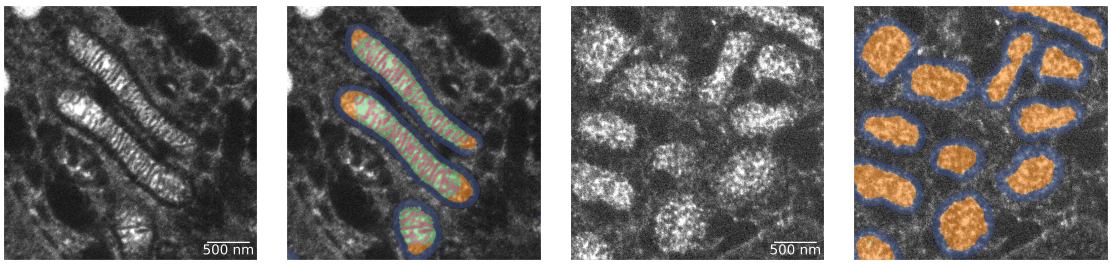
Based on segmentations provided in Section 4.4 for the mouse kidney tissue, I derive quantitative metrics describing the mitochondria and their ultrastructure. Exemplarily, Figure 6.12 shows the derived length and proportion of cristae but any metric investigated in the previous section can be visualised. It is found that the mitochondria have mean length of 2.6µm with a standard deviation of 2.9µm which is slightly shorter than the value we found for HeLa cells (3.3µm). Moreover, the mitochondria are also thicker with a mean of 0.4µm and standard deviation of 0.08µm. The fraction of the volume identified as cristae is similar with 10.4% ± 5.4% (mean ± SD). However, Figure 4.24 already indicated at more fine-grained cristae structure which is represented in a smaller average planar cristae to cristae distance of 32nm with a standard deviation of 11nm.

Regarding the difference between healthy samples and AKI tissue, Figure 6.13(a)-(d) shows the predictions for WT and AKI samples and Figure 6.13(e) shows the quantitative comparison of the lamellar cristae within each cell. To account for the uncertainty in the identification of the lamellar cristae, the mitochondria are divided into groups based on the fraction of lamellar cristae and aggregate the mitochondria for each FoV. The right plot within Figure 6.13(e) shows the mean average fraction of lamellar ultrastructure per FoV with means and standard deviations of 0.26 ± 0.16 and 0.007 ± 0.006 for WT and AKI respectively. The difference is striking and confirms the visual expression that except a few cases, most of the mitochondria from AKI samples do not show any lamellar structure any more. Additionally, the measurement reveals the large heterogeneity between mitochondria within the same condition.

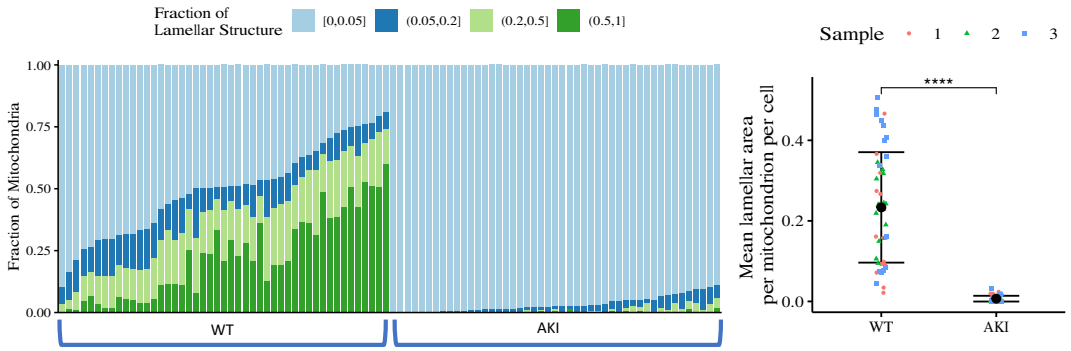
I also compare the mitochondria in terms of other macro parameters like the volume, length or diameter but find no significant differences as shown in Figure 6.14. However, it is found that the mitochondria in AKI samples are significantly rounder (with a p value < 0.001) with matches the visual impression. The significance is calculated based on a mixed-effects model analogous to Section 6.1.3. This result underlines the importance of investigating the mitochondria at the ultrastructure as only this reveals the drastic changes within the cristae structure.

6.3 Immunolabeling Analysis

One of the benefits of ExM over electron microscopy is the possibility to easily stain the sample with additional anti-bodies in addition to the pan-staining. While the previous Chapter 4 explored how these additional stainings can be used to improve the segmentation model, biologist also have an intrinsic interest in understanding the protein composition and localisation. In this section, I investigate Tom20, Mic60 and COXIV in combination with the segmentation masks generated by our model from Chapter 4 as shown in Figure 6.15. The 3D rendering shown in the last row already provides interesting insights to biologists as it allows to understand the binding sites within the three dimensional mitochondria context interactively.



(a) WT Pan Staining **(b)** WT Segmentation **(c)** AKI Pan Staining **(d)** AKI Segmentation



(e) Quantitative Comparison of WT and AKI

Figure 6.13: WT and AKI results: (a)-(d) compares the segmentation result between two WT and AKI crops showing the lamellar structures in the WT.(e) shows a substantial decline in the lamellar volume between 32 WT (14,065 mitochondria) and 33 AKI cells (10,227 mitochondria).

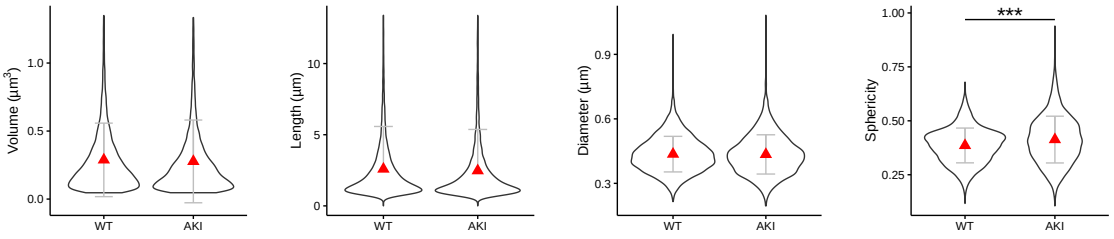


Figure 6.14: WT and AKI morphology comparison: No significant difference is found with respect to the volume, length or diameter. However, the mitochondria in AKI samples are significantly more round.

Tom20 targets the translocase of the outer mitochondrial membrane (TOM) complex located in the outer membrane which facilitates the protein exchange between the mitochondrion and the celloplasma (Yamamoto et al., 2011). The formation of the cristae is governed by the mitochondria contact site and cristae organisation system (MICOS) and Mic60 is reported to accumulate at cristae junctions (van der Laan et al., 2016). Finally, cytochrome c oxidase (COX) is a critical enzyme in the mitochondrial electron transport chain crucial for the cellular

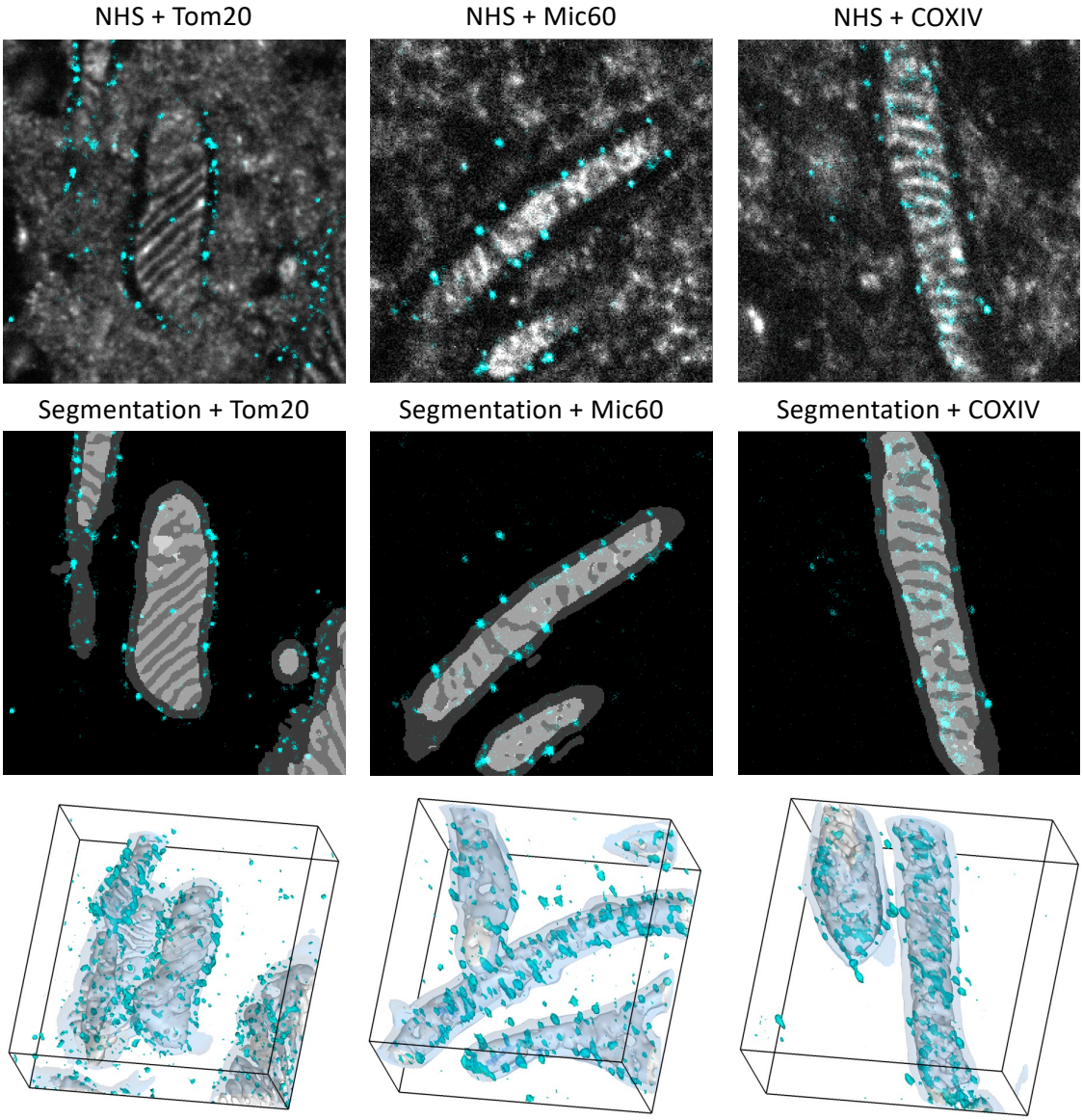


Figure 6.15: Antibody Labelling: Example of the antibodies Tom20, Mic60 and COXIV together with the global NHS staining in the first row. The second row displays a single 2D slice of the antibody with the produced segmentation mask while the third row shows a 3D rendering including the outer and inner membrane of the mitochondria.

respiration and production of ATP (Lodish et al., 2021). COXIV is therefore dominantly enriched at the inner mitochondrial membrane.

Localising the Binding

The first step for a quantitative analysis requires to identify the location of the antibody bindings. As shown in Figure 6.16 for the example of COXIV, the intensity

of the antibodies does not show any clear cut-offs but relatively smoothly transitions between background (roughly intensity values smaller than 0.65 of the maximum value) and the true binding locations and comes in pixel clusters of different sizes.

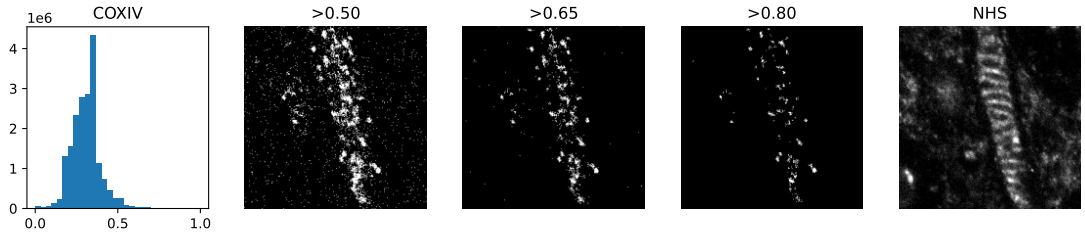


Figure 6.16: Antibody Intensities: The histogram of the intensity values shows the variation in the intensity values of the COXIV antibody. Three examples of a zoomed in crop demonstrate how different thresholds identify different antibody binding locations.

When comparing the mean intensity of the different immunolabeling within the identified mitochondria (to account for different number of mitochondria per cell) it is found that the expressiveness differs substantially. This is problematic as it leads to biased results when comparing the minimal distance between a cristae junction and the closest antibody binding.

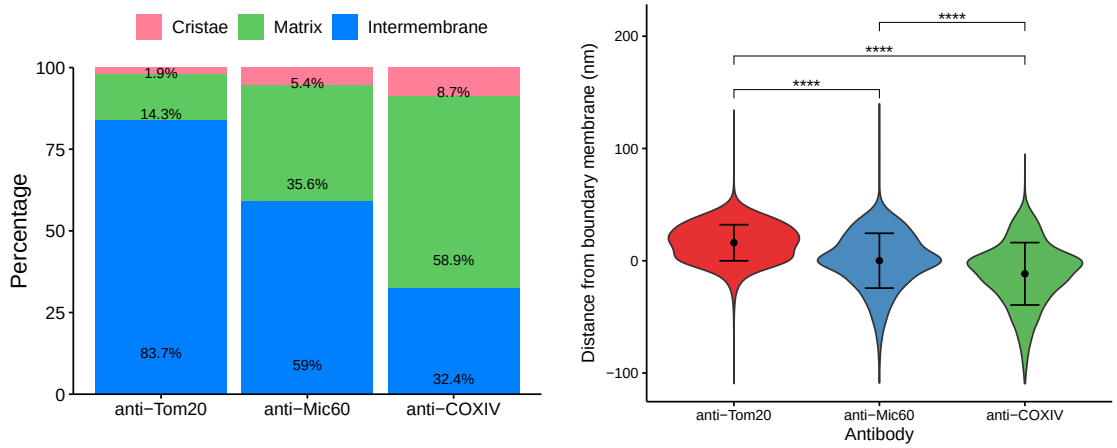
To normalisation over the number of bindings, I therefore choose the threshold for each cell individually such that the final set of samples has the most similar expressiveness of antibodies as measured by mean binding number per mitochondria voxel.

After thresholding to a binary mask, a total variation denoising step (Chambolle, 2004) is performed to remove isolated pixels. Subsequently, I identify the centre pixel of the remaining pixel clusters and record it as the localisation of the antibody. If a cluster is excessively large, it is split iteratively using k-Means until no cluster exceeds the initial 99 percentile size.

Analysis

I analyse three cells with Tom20 containing 833 mitochondria, three cells with Mic60 containing 320 mitochondria and two cells with COXIV containing 193 mitochondria. The first analysis step calculates the proportion of each segmentation label where the antibodies locate. Because on average only 4% of the image stacks of interest are occupied by mitochondria and some noisy binding is unavoidable, the percentage

of antibody bindings within the mitochondria are 26%, 37% and 68% for Tom20, Mic60 and COXIV respectively. The subsequent analysis therefore focuses on the mitochondria localising within the mitochondria which is shown in Figure 6.17(a). With 84% the fast majority of Tom20 can be found within the intermembrane space. As Mic60 binds very close to the boundary membrane, it is found that 60% of the antibodies are localised in the intermembrane space and 36% in the matrix. For COXIV, with 59% most the antibodies are found within the matrix.



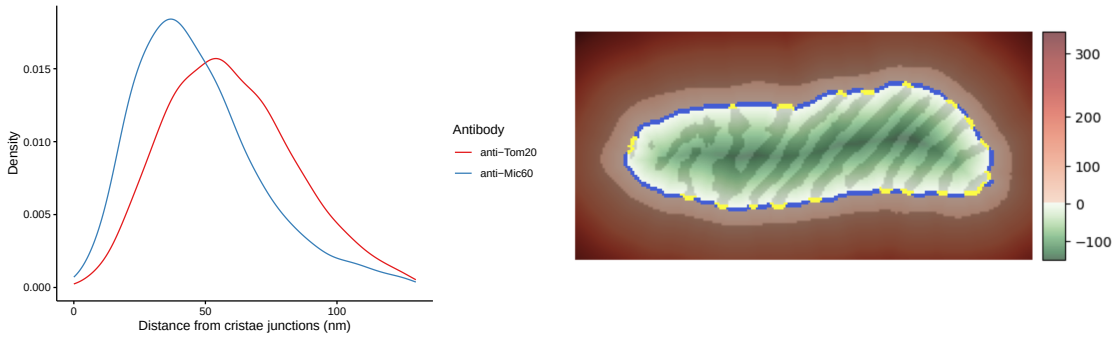
(a) Binding Proportion **(b)** Distance Boundary

Figure 6.17: Immunolabeling Analysis: (a) shows the distribution of anti-body bindings over the segmented areas for antibodies within the identified mitochondria. (b) shows the 3D distance to the boundary membrane. ****P<0.0001 based on a Wilcoxon test.

The visual inspection of the samples in Figure 6.15 also suggests that the Mic60 is binding closer to the boundary membrane, the part of the inner membrane bordering the intermembrane space, than the Tom20. In order to differentiate between the inner and outer part of the mitochondrion I use a signed distance with negative values indicating a position within the mitochondrion as shown in Figure 6.18(b). Figure 6.17(b) shows a violin plot with the mean and standard error of the distances of the antibody binding locations to the inner part of the mitochondrion and confirms that the Mic60 is statistically closer to the boundary membrane than the Tom20.

Moreover, I investigate the distance between the cristae junctions (as shown in yellow in Figure 6.18(b)) and the Tom20 and Mic60 antibodies. Specifically, for each cristae junctions the distance to the next closest antibody is recorded. The

resulting distribution is shown in Figure 6.18 and shows a clear and significant shift towards zero for the Mic60 which indicates that Mic60 binds closer to the cristae junctions which is also hinted at in Figure 6.15.



(a) Distance Cristae **(b)** Visualisation

Figure 6.18: Antibody Distance from Cristae Junctions and Visualisation: (a) Distribution of the distance of closest anti-body labelling for each cristae junction in nm. (b) Visualising the distance in nm from the boundary membrane. The boundary membrane is depicted by the blue and yellow line where the yellow parts separately indicate the cristae junctions. The distance from the boundary membrane is shown outwards in red and inwards with a negative sign in green overlaying the mitochondria segmentation (in grey).

6.4 Discussion

In this chapter, I completed the analysis pipeline and demonstrated how the framework can be used to analyse different aspects of the mitochondria morphology in HeLa cells as well as in mouse kidney tissue samples. In total, I analysed over 65,000 mitochondria from over 200 field of views. Moreover, I highlight how every measurement can be connected back to the underlying mitochondrion, which can be inspected in 3D and included several examples into the presented histograms.

The analysis allowed a detailed description of the mitochondria in HeLa cells in terms of macro parameters like the volume, length, diameter and network topology. Moreover, based on the high resolution provided by panExM I also provided a description of the ultrastructure in terms of the volume of the cristae and the distance between individual cristae invaginations. Moreover, I extended the analysis to drug treated HeLa cells to quantify the morphological differences of

the mitochondria. The analysis revealed for example a significant decrease of the diameter of mitochondria from nutrient starved cells and challenged the assumption that nutrient starvation is associated with larger networks.

With respect to the kidney tissue, I was able to quantify the change in the cristae ultrastructure between healthy samples and tissue affected by AKI which lead to an almost complete disappearance of the lamellar structure. Moreover, I was able to show that this change is not visible in terms of macro parameters like the length or diameter. Only the sphericity revealed significantly rounder mitochondria in the AKI samples.

Finally, I demonstrated how the segmentation masks can aid to understand additional biomarkers by analysis Tom20, Mic60 and COXIV as examples. I provide 3D renderings which allow to inspect the distribution of the immunolabeling in the context of the mitochondria without distracting background. Moreover, I analyse the localisation of the labelling clusters quantitatively and quantitatively validate differences in the accumulation sides.

Limitations

The limitations discussed in section 4.6 extend to the analysis presented in this chapter and future work would benefit from investigating ways of including the mitochondria DNA in the segmentation masks. Additionally, the assessment of the network topology within the HBSS treated cells could be effected by small segmentation mistakes which miss very thin connecting parts. This could be a potential reason why the analysis did not reveal a significant increase in the mitochondria length or connectedness for cells treated with HBSS. However, an extensive visual examination of the segmentation mask does not indicate that these missing parts are the missing pieces of one large network. Nevertheless, assessing the network structure in 3D visually is a very challenging task.

With regard to the kidney tissue, one limitation of the current analysis is that the appearance of the mitochondria in some cases changes substantially between different FoV. Even though I analyse over 30 FoVs, this number might still be too

small to completely remove a potential bias due to the choice of the FoV made by the biologist. One potential solution is to replace the spinning disk with a high-content screening microscope to capture the entire cells.

With regard to the comparison of the localisation of additional biomarkers, one key issue is the consistent identification of labelling cluster across different antibodies. The challenge is due to the noisy character as shown in Figure 6.16 but I am confident that the proposed approach, which combines denoising with normalisation based on mitochondria segmentation, offers a solution that is both reproducible and objective.

Additionally, the nature of the presented investigation into HeLa cells and kidney tissue was rather exploratory and meant as a proof of work for our method but not guided by a specific biological objective from the beginning. Therefore a potential criticism of the kidney tissue analysis might be that the change is so obvious visually that no quantification is necessary. However, the proportion of lamellar cristae within the wild type is less clear visually the visual impression of the global parameters also ambiguous. Therefore, the value in the presented work not only in quantifying the differences between healthy and AKI tissue but also in describing the ultrastructure in the WT and allowing to combine parameters that describe the global morphology with the ultrastructure. Moreover, I hope that the work presented can act as a blueprint for future studies which examine mitochondria to answer targeted biological questions.

7
Conclusion and Future Work

In the 20th century, the field of microscopy has been revolutionised by the invention of electron microscopy and other superresolution techniques which allowed to break through the diffraction limit of light. These advances enabled biologist to visualise structures as small as single proteins which form the fundamental building blocks of biology.

The 21st century might see another revolution in microscopy, however, it might not come in terms of new imaging capabilities but rather in the way we analyse the resulting images. With the current microscopy techniques already blurring the boundaries between biology and chemistry in their resolution capabilities, further improvements might rather be seen in other areas like the imaging throughput, live-imaging capabilities, more precise targeted staining and the general availability of these technologies to researchers.

All these future developments have the potential to lead to an unseen increase in the produced data and require the efficient use of computers to analyse them. Instead of researchers qualitatively describing what they observe in a few samples, computers process huge datasets and allow to quantitatively analyse the samples. While there already exist commercial products in this space for well established microscopy modalities, a new modality like panExM faces the challenge that the model development cannot rely on a history of freely accessible annotated datasets

like they exist in the EM community and it might not be financially attractive enough for companies while the initial user base is small.

This thesis presents a pipeline that allows to analyse mitochondria and their ultrastructure in pan-stained ExM images in an automated way with minimal human annotation required. Pan-stained ExM allows to 3D image a much larger number of cells compared to EM and my pipeline enables researchers to analyse this large amount of data. To overcome the challenge of missing pre-existing annotated data, I demonstrate how the annotation effort can be substantially reduced by exploiting the opportunity to add additional dyes (like MitoTracker in Section 4.2) or immunolabeling (Section 4.4) in addition to the pan-staining with ExM. I show how this additional information can be used despite the high noise level to train an initial model capable of segmenting the outline of the mitochondria. Subsequently, this model only requires a limited amount of annotation data to also segment the inner structure of the mitochondria. Based on this approach, I analyse over 6,000 mitochondria in 30 HeLa cells and investigate how the mitochondria react to different drug treatments based on over 35,000 mitochondria in about 130 cells.

Moreover, I also quantify the dramatic change in the ultrastructure that is caused by acute kidney injury in mouse tissue cells in a dataset consisting of about 60 cells and over 25,000 mitochondria. To the best of our knowledge, at the time of writing, there is no other study that investigates the mitochondria ultrastructure on a similar scale.

With the ability to precisely analyse the mitochondria cristae only emerging recently, one limitation of the current study is that biologists haven't formed very precise quantitative hypotheses with regard to the cristae structure yet and therefore I mostly focused on quantifying its volume. However, we expect that the wider availability of superresolution microscopy enabled by panExM might lead to a growing interest and understanding of the ultrastructure that ultimately leads to more precise quantitative questions. I therefore publish the code for the whole segmentation pipeline and analysis under an open licence at `https://github.com/AlexSauer/MAPS` for other researchers to use.

Segmentation Uncertainty

While I showed that the produced segmentation mask are almost completely accepted by a human inspector, I also established that there is substantial uncertainty in the true underlying segmentation in Chapter 3. I therefore developed an algorithm which is able to utilise the information contained in the disagreement between different annotators to increase the performance in a semi-supervised setting. Moreover, I theoretically analysed the popular Dice loss in a setting with noisy segmentation labels and show its favourable properties. Over the course of this project, we have seen a substantial improvement in the image quality with panExM and further improvements might allow to decrease some of the segmentation ambiguity. The ultimate goal in this direction would be to find a dye or immunolabeling that visualises the targeted structure precisely enough to act as the sole training target for the DL model. This would not only remove the need for additional human annotations but also eliminate the ambiguity coming with them.

Reconstruction of Plausible Segmentations

One of the challenges in microscopy data compared to natural images is the substantially higher annotation uncertainty and increased local variability in the imaging quality. This poses a challenge for segmentation models but biological samples on the other hand often have an inherent structure derived from the underlying biological building blocks that imposes useful constraints on any valid segmentation. For the example of mitochondria, this implies that any biologically plausible segmentation should respect the double membrane structure and account for the fact that the cristae are by definition surrounded by matrix volume. Translating these constraints for a 3D organelle into mathematical formulas is very challenging and the approach does not scale to new organelles. That is why I propose a purely data-driven approach in Chapter 5. The method consists of two steps in which the first step identifies areas of high uncertainty and therefore potentially misleading segmentation masks by an ensemble approach. In the second step, I build on a masked autoencoder to apply the learned prior of biologically

plausible segmentations to these areas and finally, I demonstrate how this approach performs superior to competing methods with regard to several metrics which approximate the biological validity. The method offers a new perspective on the use-cases for masked autoencoders and tackles the common of uneven imaging quality in microscopy. We are excited to see this approach being applied to other segmentation problems in which the organelle of interest has similar biological constraints.

Lightsheet and High-Content Screening Microscopes

The main objective of this work is to enable the analysis of large amounts of imaging data. Our collaborators at Yale University are building a new light sheet microscope specifically designed for the pan-stained expansion microscopy. This has the potential to increase the throughput of super-resolution data even further by imaging even more cells and thicker samples. Moreover, the light sheet microscopy has the potential to image a larger field of view and increase the resolution in the z direction. This will allow for a more exhaustive analysis as with the current spinning disk microscope our collaborators had restrict themselves to parts of a HeLa cell. This effect was even more prominent with the kidney samples as the morphology of the mitochondria varied substantially between the FoV that our collaborators chose to image.

Initial results are promising according to our collaborators but it remains to be seen whether the resulting imaging are similar enough to directly apply our models. However, this is not a major concern as one of the main focus of this work was to make the adoption to new situations as easy as possible by exploiting an additional staining like the MitoTracker. This allows to easily retrain the model to a new data like this.

Moreover, in a new collaboration with the Charité in Berlin, first attempts have shown promising results of using high-content screening (HCS) microscopes like an Pera Phenix for panExM imaging. As the presented analysis pipeline essentially lifts the bottleneck of analysing the imaged data, pairing it with a HCS microscope is an exciting opportunity to truly scale up the presented method.

Expansion to Other Organelles

This thesis focused on the analysis of mitochondria but panExM allows to visualise other organelles as well. Thus, there are ambitions to segment other targets within the cell like the endoplasmic reticulum (ER) or Golgi complex. The ER is a network of membranes responsible for protein and lipid synthesis. The rough ER synthesizes and folds proteins, while the smooth ER produces lipids and detoxifies certain chemicals. It also transports these molecules to the Golgi complex. The Golgi complex processes, modifies, and packages the received proteins and lipids and prepares them for transport to their final destinations inside or outside the cell. It sorts and tags these molecules, ensuring they reach the correct location. The Golgi also plays a role in synthesizing certain molecules, like carbohydrates used in cell membranes.

As discussed by M'Saad et al. (2020) in the original panExM paper, the ER does not show a characteristic NHS-ester pan-staining but can still be visualised with an anti-GFP antibody. Moreover, it is possible to visualise the Golgi complex based on a maleimide staining and can be combined with the NHS ester which also shows some of the Golgi cisternae. However, the Golgi complex is much harder to identify in the NHS-ester pan staining and often only partially visible. Therefore, a complete analysis of these organelles would require further research into a different approach that focuses more on the combination of the NHS-ester staining and the additional stainings.

Additionally, a start-up[1] was founded by the author of the original panExM paper which is interested in integrating my analysis pipeline into their commercial applications and investigate brain tissue which I will be consulting for.

Expansion to Human Tissue

Finally, after this work demonstrated the automated analysis pipeline for cell cultures and mouse tissue in this thesis, the goal is to apply our method to images of human patient samples.

[1] https://www.panluminate.com

The Rittscher Lab established a new collaboration with Philip Mergenthaler at the Charité in Berlin and Mergenthaler's groups conducted first experiments with an HCS microscope that allows to increase the amount of imaged data substantially.

One project, which the group is working on, aims to uncover the link between genetic mutations in the mitochondria DNA, the resulting morphological changes in the mitochondrial ultrastructure and the dysfunction of the cell. Specifically, they are investigating the MELAS (Mitochondrial Encephalomyopathy and Lactic Acidosis and Stroke-like-episodes) syndrome in which a mutation in the mitochondrial genome leads to a defect in the respiratory chain and therefore insufficient cellular energy supply. This causes stroke-like episodes and dementia due to neuronal cell death (El-Hattab et al., 2015). Moreover, electron microscopy analysis of tissue from MELAS patients revealed mitochondria with abnormal ultrastructure (Felczak et al., 2017). However, because even neighbouring cells and mitochondria can show substantial differences in the mutation burden, it is essential to conduct a large scale analysis by imaging large numbers of cells which is very challenging with EM. Scaling up an analysis like this is exactly the setting which panExM and my analysis pipeline were developed for, and it is exciting to see the insights they will generate.

References

Abbe, Ernst (1882). "The Relation of Aperture and Power in the Microscope". In: *Journal of the Royal Microscopical Society* 2.4, pp. 460–473.

Abdar, Moloud et al. (2021). "A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges". In: *Information Fusion* 76, pp. 243–297.

Allen, Terence (2015). *Microscopy: A Very Short Introduction*. Oxford University Press.

Alzubaidi, Laith et al. (2021a). "Deepening into the Suitability of Using Pre-Trained Models of ImageNet against a Lightweight Convolutional Neural Network in Medical Imaging: An Experimental Study". In: *PeerJ Computer Science* 7, e715.

Alzubaidi, Laith et al. (2021b). "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions". In: *Journal of Big Data* 8.1, p. 53.

Aswath, Anusha et al. (2023). "Segmentation in Large-Scale Cellular Electron Microscopy with Deep Learning: A Literature Survey". In: *Medical Image Analysis* 89, p. 102920.

Banerji, Sugata and Sushmita Mitra (2022). "Deep Learning in Histopathology: A Review". In: *WIREs Data Mining and Knowledge Discovery* 12.1, e1439.

Beal, M.F. (2007). "Mitochondria and Neurodegeneration". In: *Novartis Foundation Symposium* 287, pp. 183–192.

Bearman, Amy et al. (2016). "What's the Point: Semantic Segmentation with Point Supervision". In: *Computer Vision – ECCV 2016*. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 549–565.

Benda, Carl (1898). "Über Die Spermatogenese Der Vertebraten Und Höherer Evertebraten, II. Theil: Die Histiogenese Der Spermien". In: *Arch. Anat. Physiol* 73, pp. 393–398.

Berg, Stuart et al. (2019). "Ilastik: Interactive Machine Learning for (Bio)Image Analysis". In: *Nature Methods* 16.12, pp. 1226–1232.

Betzig, Eric et al. (2006). "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution". In: *Science* 313.5793, pp. 1642–1645.

Bianchini, Paolo et al. (2015). "STED Nanoscopy: A Glimpse into the Future". In: *Cell and Tissue Research* 360.1, pp. 143–150.

Bílý, Tomáš et al. (2021). "Ultrastructural Changes of the Mitochondrion During the Life Cycle of Trypanosoma Brucei". In: *Journal of Eukaryotic Microbiology* 68.3, e12846.

Boers, Tim G. W. et al. (2022). "Comparing Training Strategies Using Multi-Assessor Segmentation Labels for Barrett's Neoplasia Detection". In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 131–138.

Bohlender, Simon, Ilkay Oksuz, and Anirban Mukhopadhyay (2023). "A Survey on Shape-Constraint Deep Learning for Medical Image Segmentation". In: *IEEE Reviews in Biomedical Engineering* 16, pp. 225–240.

Brandt, Tobias et al. (2017). "Changes of Mitochondrial Ultrastructure and Function during Ageing in Mice and Drosophila". In: *eLife* 6, e24662.

Breiman, Leo (2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32.

Broughton, Richard E. and Paulette C. Reneau (2006). "Spatial Covariation of Mutation and Nonsynonymous Substitution Rates in Vertebrate Mitochondrial Genomes". In: *Molecular Biology and Evolution* 23.8, pp. 1516–1524.

Brown, Tom B. et al. (2020). *Language Models Are Few-Shot Learners*. URL: http://arxiv.org/abs/2005.14165 (visited on 12/03/2023). Pre-published.

Cadonic, Chris, Mohammad Golam Sabbir, and Benedict C. Albensi (2015). "Mechanisms of Mitochondrial Dysfunction in Alzheimer's Disease". In: *Molecular Neurobiology 2015 53:9* 53.9, pp. 6078–6090.

Caldeira, João and Brian Nord (2020). "Deeply Uncertain: Comparing Methods of Uncertainty Quantification in Deep Learning Algorithms". In: *Machine Learning: Science and Technology* 2.1, p. 015002.

Can, Yigit B. et al. (2018). "Learning to Segment Medical Images with Scribble-Supervision Alone". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Ed. by Danail Stoyanov et al. Cham: Springer International Publishing, pp. 236–244.

Canny, John (1986). "A Computational Approach to Edge Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6, pp. 679–698.

Cao, Hu et al. (2023). "Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation". In: *Computer Vision – ECCV 2022 Workshops*. Ed. by Leonid Karlinsky, Tomer Michaeli, and Ko Nishino. Cham: Springer Nature Switzerland, pp. 205–218.

Cao, Xuyang et al. (2021). "Uncertainty Aware Temporal-Ensembling Model for Semi-Supervised ABUS Mass Segmentation". In: *IEEE Transactions on Medical Imaging* 40.1, pp. 431–443.

Caron, Mathilde et al. (2021). "Emerging Properties in Self-Supervised Vision Transformers". In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9650–9660.

Casser, Vincent et al. (2020). "Fast Mitochondria Detection for Connectomics". In: *Proceedings of Machine Learning Research* 121, pp. 111–120.

Chambolle, Antonin (2004). "An Algorithm for Total Variation Minimization and Applications". In: *Journal of Mathematical Imaging and Vision* 20.1, pp. 89–97.

Chang, Gary Han et al. (2024). "Isotropic Multi-Scale Neuronal Reconstruction from High-Ratio Expansion Microscopy with Contrastive Unsupervised Deep Generative Models". In: *Computer Methods and Programs in Biomedicine* 244, p. 107991.

Chang, Jae Byum et al. (2017). "Iterative Expansion Microscopy". In: *Nature Methods* 14.6, pp. 593–599.

Chen, Fei, Paul W. Tillberg, and Edward S. Boyden (2015). "Expansion Microscopy". In: *Science* 347.6221, pp. 543–548.

Chen, Tianping and Hong Chen (1995). "Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamical Systems". In: *IEEE Transactions on Neural Networks* 6.4, pp. 911–917.

Chen, Tianqi and Carlos Guestrin (2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. New York, NY, USA: Association for Computing Machinery, pp. 785–794.

Chen, Ting et al. (2020). "A Simple Framework for Contrastive Learning of Visual Representations". In: *Proceedings of the 37th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, pp. 1597–1607.

Cheng, D et al. (2019). “Darnet: Deep Active Ray Network for Building Segmentation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7431–7439.

Cheng, Hsueh Chien and Amitabh Varshney (2018). “Volume Segmentation Using Convolutional Neural Networks with Limited Training Data”. In: *Proceedings - International Conference on Image Processing, ICIP*, pp. 590–594.

Çiçek, Özgün et al. (2016). “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Ed. by Sebastien Ourselin et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 424–432.

Codella, Noel et al. (2019). “Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)”.

Cogliati, Sara, Jose A. Enriquez, and Luca Scorrano (2016). “Mitochondrial Cristae: Where Beauty Meets Functionality”. In: *Trends in Biochemical Sciences* 41.3, pp. 261–273.

Cogliati, Sara et al. (2013). “Mitochondrial Cristae Shape Determines Respiratory Chain Supercomplexes Assembly and Respiratory Efficiency”. In: *Cell* 155.1, pp. 160–171.

Cohen, Jacob (1960). “A Coefficient of Agreement for Nominal Scales”. In: *Educational and Psychological Measurement* 20.1, pp. 37–46.

Cohen, Y. et al. (1992). “Characterization of Inhomogeneous Polyacrylamide Hydrogels”. In: *Journal of Polymer Science Part B: Polymer Physics* 30.9, pp. 1055–1067.

Cover, TL, LY Reddy, and MJ Blaser (1993). “Effects of ATPase Inhibitors on the Response of HeLa Cells to Helicobacter Pylori Vacuolating Toxin”. In: *Infection and immunity* 61.4, pp. 1427–1431.

Czolbe, Steffen et al. (2021). “Is Segmentation Uncertainty Useful?” In: *Information Processing in Medical Imaging*. Ed. by Aasa Feragen et al. Springer International Publishing, pp. 715–726.

Dard, L. et al. (2020). “Mitochondrial Functions and Rare Diseases”. In: *Molecular Aspects of Medicine* 71, p. 100842.

Dehghani, Mostafa et al. (2023). *Scaling Vision Transformers to 22 Billion Parameters*. URL: `http://arxiv.org/abs/2302.05442` (visited on 12/03/2023). Pre-published.

Deng, Jia et al. (2009). “ImageNet: A Large-Scale Hierarchical Image Database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255.

Desideri, Enrico, Giuseppe Filomeni, and Maria Rosa Ciriolo (2012). “Glutathione Participates in the Modulation of Starvation-Induced Autophagy in Carcinoma Cells”. In: *Autophagy* 8.12, pp. 1769–1781.

Dietterich, Thomas G. (2000). “Ensemble Methods in Machine Learning”. In: *Multiple Classifier Systems*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 1–15.

Dimmer, K.S. et al. (2008). “LETM1, Deleted in Wolf-Hirschhorn Syndrome Is Required for Normal Mitochondrial Morphology and Cellular Viability”. In: *Human Molecular Genetics* 17.2, pp. 201–214.

Dong, Shi, Ping Wang, and Khushnood Abbas (2021). “A Survey on Deep Learning and Its Applications”. In: *Computer Science Review* 40, p. 100379.

Dosovitskiy, Alexey et al. (2021). “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *International Conference on Learning Representations*.

Duann, Pu and Pei-Hui Lin (2017). "Mitochondria Damage and Kidney Disease". In: *Mitochondrial Dynamics in Cardiovascular Medicine*. Ed. by Gaetano Santulli. Cham: Springer International Publishing, pp. 529–551.

Duranova, Hana et al. (2020). "Mitochondria: A Worthwhile Object for Ultrastructural Qualitative Characterization and Quantification of Cells at Physiological and Pathophysiological States Using Conventional Transmission Electron Microscopy". In: *Acta Histochemica* 122.8, p. 151646.

Faulkner, Emma L., Steven G Thomas, and Robert K. Neely (2020). "An Introduction to the Methodology of Expansion Microscopy". In: *The International Journal of Biochemistry & Cell Biology* 124, p. 105764.

Felczak, Paulina et al. (2017). "Pathology of Mitochondria in MELAS Syndrome: An Ultrastructural Study". In: *Polish Journal of Pathology* 68.2, pp. 173–181.

Fischer, Christian A. et al. (2020). "MitoSegNet: Easy-to-use Deep Learning Segmentation for Analyzing Mitochondrial Morphology". In: *iScience* 23.10, p. 101601.

Fleiss, Joseph L., Bruce Levin, and Myunghee Cho Paik (2003). *Statistical Methods for Rates and Proportions*. 1st ed. Wiley Series in Probability and Statistics. Wiley.

Franco-Barranco, Daniel, Arrate Muñoz-Barrutia, and Ignacio Arganda-Carreras (2021). "Stable Deep Neural Network Architectures for Mitochondria Segmentation on Electron Microscopy Volumes". In: *Neuroinformatics* 1, pp. 1–14.

Franco-Barranco, Daniel et al. (2022). "Deep Learning Based Domain Adaptation for Mitochondria Segmentation on EM Volumes". In: *Computer Methods and Programs in Biomedicine* 222, p. 106949.

Franco-Iborra, Sandra, Miquel Vila, and Celine Perier (2016). "The Parkinson Disease Mitochondrial Hypothesis: Where Are We At?" In: *The Neuroscientist* 22.3, pp. 266–277.

Freund, Yoav and Robert E. Schapire (1997). "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1, pp. 119–139.

Frey, Terrence G and Carmen A Mannella (2000). "The Internal Structure of Mitochondria". In: *Trends in Biochemical Sciences* 25.7, pp. 319–324.

Fry, Michelle Y et al. (2024). "In Situ Architecture of Opa1-dependent Mitochondrial Cristae Remodeling". In: *The EMBO Journal* 43.3, pp. 391–413.

Frye, Richard E. et al. (2021). "Mitochondrial Morphology Is Associated with Respiratory Chain Uncoupling in Autism Spectrum Disorder". In: *Translational Psychiatry* 11.1, pp. 1–17.

Gal, Yarin and Zoubin Ghahramani (2016). "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *Proceedings of The 33rd International Conference on Machine Learning (PMLR) 48*, pp. 1050–1059.

Galbraith, Catherine G. and James A. Galbraith (2011). "Super-Resolution Microscopy at a Glance". In: *Journal of Cell Science* 124.10, pp. 1607–1611.

Gambarotto, Davide et al. (2019). "Imaging Cellular Ultrastructures Using Expansion Microscopy (U-ExM)". In: *Nature Methods* 16.1, pp. 71–74.

Ganaie, M. A. et al. (2022). "Ensemble Deep Learning: A Review". In: *Engineering Applications of Artificial Intelligence* 115, p. 105151.

Gao, Feng et al. (2022). "Segmentation Only Uses Sparse Annotations: Unified Weakly and Semi-Supervised Learning in Medical Images". In: *Medical Image Analysis* 80, p. 102515.

Gao, Ruixuan, Shoh M. Asano, and Edward S. Boyden (2017). "Q & A: Expansion Microscopy". In: *BMC Biology* 15.1, pp. 1–9.

Gatti, Domenico L. et al. (2019). *Deep Learning Strategies for Differential Expansion Microscopy*. URL: `https://www.biorxiv.org/content/10.1101/743682v1` (visited on 08/22/2024). Pre-published.

Gaynes, Robert P. (2023). *Germ Theory: Medical Pioneers in Infectious Diseases*. 2nd ed. Washington Hoboken, NJ: American society for microbiology Wiley.

Gilchrist, James M. et al. (1996). "Adult-Onset MELAS". In: *Stroke* 27.8, pp. 1420–1423.

Grandvalet, Yves and Yoshua Bengio (2004). "Semi-Supervised Learning by Entropy Minimization". In: *Advances in Neural Information Processing Systems 17*.

Graves, Alex (2011). "Practical Variational Inference for Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 24.

Grill, Jean-Bastien et al. (2020). "Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., pp. 21271–21284.

Guo, Chuan et al. (2017). "On Calibration of Modern Neural Networks". In: *Proceedings of the 34th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, pp. 1321–1330.

Guo, Lihong et al. (2021). "Learned Snakes for 3D Image Segmentation". In: *Signal Processing* 183, p. 108013.

Guo, Xiaojia et al. (2023). "APOBEC-1 Deletion Enhances Cisplatin-Induced Acute Kidney Injury". In: *Scientific Reports* 13.1, p. 22255.

Gur, Shir, Tal Shaharabany, and Lior Wolf (2019). "End to End Trainable Active Contours via Differentiable Rendering". In: *Proceedings of the International Conference on Learning Representations*.

Gustafsson, M. G. L. (2000). "Surpassing the Lateral Resolution Limit by a Factor of Two Using Structured Illumination Microscopy". In: *Journal of Microscopy* 198.2, pp. 82–87.

Han, Hongqing et al. (2022). "Self-Supervised Voxel-Level Representation Rediscovers Subcellular Structures in Volume Electron Microscopy". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1874–1883.

Hastie, Trevor, Robert Tibshirani, and J H Friedman (2009). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. New York.

Hatamizadeh, Ali et al. (2022a). "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by Alessandro Crimi and Spyridon Bakas. Cham: Springer International Publishing, pp. 272–284.

Hatamizadeh, Ali et al. (2022b). "UNETR: Transformers for 3D Medical Image Segmentation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 574–584.

El-Hattab, Ayman W. et al. (2015). "MELAS Syndrome: Clinical Manifestations, Pathogenesis, and Treatment Options". In: *Molecular Genetics and Metabolism*. SI:Therapy 116.1, pp. 4–12.

Hausen, Peter and Christine Dreyer (1981). "The Use of Polyacrylamide as an Embedding Medium for Immunohistochemical Studies of Embryonic Tissues". In: *Stain Technology* 56.5, pp. 287–293.

He, Kaiming, Ross Girshick, and Piotr Dollar (2019). "Rethinking ImageNet Pre-Training". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE, pp. 4917–4926.

He, Kaiming et al. (2015). *Deep Residual Learning for Image Recognition*. URL: `http://arxiv.org/abs/1512.03385` (visited on 01/03/2023). Pre-published.

He, Kaiming et al. (2020). "Momentum Contrast for Unsupervised Visual Representation Learning". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, pp. 9726–9735.

He, Kaiming et al. (2022). "Masked Autoencoders Are Scalable Vision Learners". In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15979–15988.

Hell, Stefan W and Jan Wichmann (1994). "Breaking the Diffraction Resolution Limit by Stimulated Emission: Stimulated-Emission-Depletion Fluorescence Microscopy". In: *Optics letters* 19.11.

Henriques, Ricardo et al. (2011). "PALM and STORM: Unlocking Live-Cell Super-Resolution". In: *Biopolymers* 95.5, pp. 322–331.

Hinton, Geoffrey E. et al. (2012). *Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors*. arXiv.org. URL: `https://arxiv.org/abs/1207.0580v1` (visited on 10/07/2024).

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Computation* 9.8, pp. 1735–1780.

Holt, I. J., A. E. Harding, and J. A. Morgan-Hughes (1988). "Deletions of Muscle Mitochondrial DNA in Patients with Mitochondrial Myopathies". In: *Nature* 331.6158, pp. 717–719.

Hooker, Sara (2021). "The Hardware Lottery". In: *Commun. ACM* 64.12, pp. 58–65.

Hox, Joop, Mirjam Moerbeek, and Rens van de Schoot (2017). *Multilevel Analysis: Techniques and Applications, Third Edition*. 3rd ed. New York: Routledge. 364 pp.

Huang, Cheng, Kun Deng, and Minghua Wu (2023). "Mitochondrial Cristae in Health and Disease". In: *International Journal of Biological Macromolecules* 235, p. 123755.

Hüllermeier, Eyke and Willem Waegeman (2021). "Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods". In: *Machine Learning* 110.3, pp. 457–506.

Ioffe, Sergey and Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *Proceedings of the 32nd International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, pp. 448–456.

Isensee, Fabian et al. (2019). "No New-Net". In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer International Publishing, pp. 234–244.

Isensee, Fabian et al. (2021). "nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation". In: *Nature Methods* 18.2, pp. 203–211.

Isensee, Fabian et al. (2024). *nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation*. URL: `http://arxiv.org/abs/2404.09556` (visited on 04/17/2024). Pre-published.

Ishimoto, Yu and Reiko Inagi (2016). "Mitochondria: A Therapeutic Target in Acute Kidney Injury". In: *Nephrology Dialysis Transplantation* 31.7, pp. 1062–1069.

Jenkins, Brenita C. et al. (2024). "Mitochondria in disease: changes in shapes and dynamics". In: *Trends in biochemical sciences* 49.4, pp. 346–360.

Johns, Donald R. (1995). "Mitochondrial DNA and Disease". In: *New England Journal of Medicine* 333.10, pp. 638–644.

Jorstad, Anne and Pascal Fua (2014). "Refining Mitochondria Segmentation in Electron Microscopy Imagery with Active Surfaces". In: *European Converence on Computer Vision*. Vol. 8928. Springer, Cham, pp. 367–379.

Kass, Michael and Andrew Witkin (1988). "Snakes: Active Contour Models". In: *International Journal of Computer Vision*, pp. 321–331.

Ke, Alexander et al. (2021). "CheXtransfer: Performance and Parameter Efficiency of ImageNet Models for Chest X-Ray Interpretation". In: *Proceedings of the Conference on Health, Inference, and Learning*. ACM CHIL '21: ACM Conference on Health, Inference, and Learning. Virtual Event USA: ACM, pp. 116–124.

Ke, Guolin et al. (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc.

Kervadec, Hoel and Marleen de Bruijne (2023). *On the Dice Loss Gradient and the Ways to Mimic It*. URL: `http://arxiv.org/abs/2304.04319` (visited on 01/09/2024). Pre-published.

Khadangi, Afshin, Thomas Boudier, and Vijay Rajagopal (2021). "EM-stellar: Benchmarking Deep Learning for Electron Microscopy Image Segmentation". In: *Bioinformatics* 37.1, pp. 97–106.

Kohl, Simon A A et al. (2018). "A Probabilistic U-Net for Segmentation of Ambiguous Images". In: *Advances in Neural Information Processing Systems 31*.

Krähenbühl, Philipp and Vladlen Koltun (2011). "Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials". In: *Advances in Neural Information Processing Systems*. Vol. 24.

Kraus, Oren et al. (2024). "Masked Autoencoders for Microscopy Are Scalable Learners of Cellular Biology". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11757–11768.

Kreshuk, Anna and Chong Zhang (2019). "Machine Learning: Advanced Image Segmentation Using Ilastik". In: *Computer Optimized Microscopy: Methods and Protocols*. Ed. by Elena Rebollo and Manel Bosch. New York, NY: Springer, pp. 449–463.

Krithika, M. and K. Suganthi (2022). "Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET". In: *Diagnostics* 12.12, p. 3064.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc.

Ku, Taeyun et al. (2016). "Multiplexed and Scalable Super-Resolution Imaging of Three-Dimensional Protein Localization in Size-Adjustable Tissues". In: *Nature Biotechnology* 34.9, pp. 973–981.

Kumar, Neeraj et al. (2017). "A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology". In: *IEEE Transactions on Medical Imaging* 36.7, pp. 1550–1560.

Kumar, Neeraj et al. (2020). "A Multi-Organ Nucleus Segmentation Challenge". In: *IEEE Transactions on Medical Imaging* 39.5, pp. 1380–1391.

Laine, Samuli and Timo Aila (2016). "Temporal Ensembling for Semi-Supervised Learning". In: *5th International Conference on Learning Representations*.

Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell (2017). "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles". In: *Advances in Neural Information Processing Systems*, pp. 6402–6413.

Lane, Nick (2006). "Powerhouse of Disease". In: *Nature* 440.7084, pp. 600–602.

— (2018). *Power, Sex, Suicide: Mitochondria and the Meaning of Life*. Oxford University Press.

Lane, Nick and William Martin (2010). "The Energetics of Genome Complexity". In: *Nature* 467.7318, pp. 929–934.

Larrazabal, Agostina J., Cesar Martinez, and Enzo Ferrante (2019). "Anatomical Priors for Image Segmentation via Post-processing with Denoising Autoencoders". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Ed. by Dinggang Shen et al. Lecture Notes in Computer Science. Springer International Publishing, pp. 585–593.

Last, John M. and Miquel Porta, eds. (2018). *A Dictionary of Public Health*. Second edition. Oxford Reference Online Premium. Oxford, New York: Oxford University Press. 1 p.

Lea, Peter J. and Martin J. Hollenberg (1989). "Mitochondrial Structure Revealed by High-Resolution Scanning Electron Microscopy". In: *American Journal of Anatomy* 184.3, pp. 245–257.

Lebrat, Léo et al. (2021). "CorticalFlow: A Diffeomorphic Mesh Deformation Module for Cortical Surface Reconstruction". In: *Advances in Neural Information Processing Systems 34*.

LeCun, Yann and Yoshua Bengio (1995). "Convolutional Networks for Images, Speech, and Time Series". In: *The handbook of brain theory and neural networks* 3361.10, p. 1995.

Lee, Dong-Hyun (2013). "Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks". In: *Workshop on Challenges in Representation Learning (ICML)*. Vol. 3. 2, p. 896.

Lee, Gi Ja et al. (2013). "Quantitative and Qualitative Analysis of Heart Mitochondria for Evaluating the Degree of Myocardial Injury Utilizing Atomic Force Microscopy". In: *Micron* 44.1, pp. 167–173.

Lee, T. C., R. L. Kashyap, and C. N. Chu (1994). "Building Skeleton Models via 3-D Medial Surface Axis Thinning Algorithms". In: *CVGIP: Graphical Models and Image Processing* 56.6, pp. 462–478.

Lefebvre, Austin E.Y.T. et al. (2021). "Automated Segmentation and Tracking of Mitochondria in Live-Cell Time-Lapse Images". In: *Nature Methods 2021 18:9* 18.9, pp. 1091–1102.

Li, Haichun, Ao Li, and Minghui Wang (2019). "A Novel End-to-End Brain Tumor Segmentation Method Using Improved Fully Convolutional Networks". In: *Computers in Biology and Medicine* 108, pp. 150–160.

Li, Weifu et al. (2016). "Segmentation of Mitochondria Based on SEM Images". In: *2016 IEEE International Conference on Mechatronics and Automation, IEEE ICMA 2016*. Institute of Electrical and Electronics Engineers Inc., pp. 1953–1958.

Li, Xiaomeng et al. (2018). "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes". In: *IEEE Transactions on Medical Imaging* 37.12, pp. 2663–2674.

Li, Yuexiang et al. (2020). "Self-Loop Uncertainty: A Novel Pseudo-Label for Semi-Supervised Medical Image Segmentation". In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Vol. 12261 LNCS. Springer Science and Business Media Deutschland GmbH, pp. 614–623.

Li, Zhili et al. (2021). "Contrastive Learning for Mitochondria Segmentation". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 3496–3500.

Lionetti, Lillà et al. (2014). "High-Lard and High-Fish-Oil Diets Differ in Their Effects on Function and Dynamic Behaviour of Rat Hepatic Mitochondria". In: *PLOS ONE* 9.3, e92753.

Liu, Guilin et al. (2018). "Image Inpainting for Irregular Holes Using Partial Convolutions". In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 85–100.

Liu, Zhichao et al. (2021). "A Survey on Applications of Deep Learning in Microscopy Image Analysis". In: *Computers in Biology and Medicine* 134, p. 104523.

Lodish, H et al. (2021). *Molecular Cell Biology*. New York: Macmillan.

Lopez, J. and S. W. G. Tait (2015). "Mitochondrial Apoptosis: Killing Cancer Using the Enemy Within". In: *British Journal of Cancer* 112.6, pp. 957–962.

Lucchi, Aurelien et al. (2015). "Learning Structured Models for Segmentation of 2-D and 3-D Imagery". In: *IEEE Transactions on Medical Imaging* 34.5, pp. 1096–1110.

Lucchi, Aurlien et al. (2012). "Supervoxel-Based Segmentation of Mitochondria in Em Image Stacks with Learned Shape Features". In: *IEEE Transactions on Medical Imaging* 31.2, pp. 474–486.

M'Saad, Ons and Joerg Bewersdorf (2020). "Light Microscopy of Proteins in Their Ultrastructural Context". In: *Nature Communications* 11.1, pp. 1–15.

Ma, Jun et al. (2021). "Loss Odyssey in Medical Image Segmentation". In: *Medical Image Analysis* 71, p. 102035.

Ma, Xiuquan et al. (2011). "Mitochondrial Electron Transport Chain Complex III Is Required for Antimycin A to Inhibit Autophagy". In: *Chemistry & biology* 18.11, pp. 1474–1481.

Maier-Hein, Lena et al. (2024). "Metrics Reloaded: Recommendations for Image Analysis Validation". In: *Nature Methods* 21.2, pp. 195–212.

Mannella, Carmen A. (2006). "Structure and Dynamics of the Mitochondrial Inner Membrane Cristae". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. Mitochondrial Dynamics in Cell Life and Death 1763.5, pp. 542–548.

Mannella, Carmen A. et al. (1994). "The Internal Compartmentation of Rat-Liver Mitochondria: Tomographic Study Using the High-Voltage Transmission Electron Microscope". In: *Microscopy Research and Technique* 27.4, pp. 278–283.

Mansoor, RAZA, Nathali Dilshani Jayasinghe, and Muhana Magboul Ali Muslam (2021). "A Comprehensive Review on Email Spam Classification Using Machine Learning Algorithms". In: *2021 International Conference on Information Networking (ICOIN)*. IEEE, pp. 327–332.

Marcos, Diego et al. (2018). "Learning Deep Structured Active Contours End-to-End". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8877–8885.

Masters, Barry R (2009). "History of the Electron Microscope in Cell Biology". In: Wiley. *Encyclopedia of Life Sciences*. 1st ed. Wiley.

McBride, Heidi M., Margaret Neuspiel, and Sylwia Wasiak (2006). "Mitochondria: More Than Just a Powerhouse". In: *Current Biology* 16.14, R551–R560.

McHugh, Mary L. (2012). "Interrater Reliability: The Kappa Statistic". In: *Biochemia Medica* 22.3, pp. 276–282.

Mehta, Ravindra L. et al. (2015). "International Society of Nephrology's 0by25 Initiative for Acute Kidney Injury (Zero Preventable Deaths by 2025): A Human Rights Case for Nephrology". In: *The Lancet* 385.9987, pp. 2616–2643.

Menze, Bjoern et al. (2022). *QUBIQ: Quantification of Uncertainties in Biomedical Images*. 2021. URL: `https://qubiq21.grand-challenge.org/` (visited on 10/25/2022).

Milletari, Fausto, Nassir Navab, and Seyed Ahmad Ahmadi (2016). "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation". In: *Proceedings International Conference on 3D Vision*, pp. 565–571.

Mittal, Sparsh and Shraiysh Vaishay (2019). "A Survey of Techniques for Optimizing Deep Learning on GPUs". In: *Journal of Systems Architecture* 99, p. 101635.

Mohamad, Nora et al. (2005). "Mitochondrial Apoptotic Pathways". In: *Biocell* 29.2, pp. 149–161.

Morgan, C. et al. (1954). "Structure and Development of Viruses Observed in the Electron Microscope. II. Vaccinia and Fowl Pox Viruses". In: *The Journal of Experimental Medicine* 100.3, pp. 301–310.

Mumcuoglu, E. U. et al. (2012). "Computerized Detection and Segmentation of Mitochondria on Electron Microscope Images". In: *Journal of Microscopy* 246.3, pp. 248–265.

Murphy, Kevin P. (2022). *Probabilistic Machine Learning: An Introduction*. MIT Press.

Narasimha, Rajesh et al. (2009). "Automatic Joint Classification and Segmentation of Whole Cell 3D Images". In: *Pattern Recognition*. Vol. 42. Pergamon, pp. 1067–1079.

Neubert, Peer and Peter Protzel (2014). "Compact Watershed and Preemptive SLIC: On Improving Trade-offs of Superpixel Segmentation Algorithms". In: *2014 22nd International Conference on Pattern Recognition*. 2014 22nd International Conference on Pattern Recognition, pp. 996–1001.

Nguyen, Nguyen P., Tommi A. White, and Filiz Bunyak (2021). "Mitochondria Instance Segmentation in Electron Microscopy Image Volumes Using 3D Deep Learning Networks". In: *2021 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, pp. 1–6.

Nieves, D. J. and M. A.B. Baker (2021). "Pushing the Super-Resolution Limit: Recent Improvements in Microscopy below the Diffraction Limit". In: *Biochemical Society Transactions* 49.1, pp. 431–439.

Nightingale, Luke et al. (2021). "Automatic Instance Segmentation of Mitochondria in Electron Microscopy Data". In: *bioRxiv*, p. 2021.05.24.444785.

Otsu, Nobuyuki (1979). "A Threshold Selection Method from Gray-Level Histograms". In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1, pp. 62–66.

Oztel, Ismail et al. (2017). "Mitochondria Segmentation in Electron Microscopy Volumes Using Deep Convolutional Neural Network". In: *Proceedings IEEE International Conference on Bioinformatics and Biomedicine*. Institute of Electrical and Electronics Engineers Inc., pp. 1195–1200.

Pagliarini, David J. and Jared Rutter (2013). "Hallmarks of a New Era in Mitochondrial Biochemistry". In: *Genes & Development* 27.24, pp. 2615–2627.

Palade, George E. (1952). "The Fine Structure of Mitochondria". In: *The Anatomical Record* 114.3, pp. 427–451.

Palay, Sanford L. and George E. Palade (1955). "The Fine Structure of Neurons". In: *The Journal of Biophysical and Biochemical Cytology* 1.1, pp. 69–88.

Parlakgül, Güneş et al. (2022). "Regulation of Liver Subcellular Architecture Controls Metabolic Homeostasis". In: *Nature* 603.7902, pp. 736–742.

Peiris, Himashi et al. (2021). "Duo-SegNet: Adversarial Dual-Views for Semi-supervised Medical Image Segmentation". In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 428–438.

Peng, Sida et al. (2020). "Deep Snake for Real-Time Instance Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8533–8542.

Perkins, Guy et al. (2012). "Ultrastructural Modifications in the Mitochondria of Hypoxia-Adapted Drosophila Melanogaster". In: *PLOS ONE* 7.9, e45344.

Perlman, Steve J. et al. (2015). "Maternal Transmission, Sex Ratio Distortion, and Mitochondria". In: *Proceedings of the National Academy of Sciences* 112.33, pp. 10162–10168.

Picard, Martin, Douglas C. Wallace, and Yan Burelle (2016). "The Rise of Mitochondria in Medicine". In: *Mitochondrion* 30, pp. 105–116.

Picard, Martin, Kathryn White, and Douglass M. Turnbull (2013). "Mitochondrial Morphology, Topology, and Membrane Interactions in Skeletal Muscle: A Quantitative Three-Dimensional Electron Microscopy Study". In: *Journal of Applied Physiology* 114.2, pp. 161–171.

Porter, Keith R., Albert Claude, and Ernest F. Fullam (1945). "A Study of Tissue Culture Cells By Electron Microsccopy". In: *The Journal of Experimental Medicine* 81.3, p. 233.

Qiao, Siyuan et al. (2020). *Micro-Batch Training with Batch-Channel Normalization and Weight Standardization*. URL: `http://arxiv.org/abs/1903.10520` (visited on 01/30/2023). Pre-published.

Radford, Alec et al. (2018). "Improving Language Understanding by Generative Pre-Training". In: *https://openai.com/research/language-unsupervised*.

Ralto, Kenneth M. and Samir M. Parikh (2016). "Mitochondria in Acute Kidney Injury". In: *Seminars in Nephrology* 36.1, pp. 8–16.

Reinke, Annika et al. (2021). *Common Limitations of Image Processing Metrics: A Picture Story*. arXiv.org. URL: `https://arxiv.org/abs/2104.05642v8` (visited on 09/05/2024).

Reinke, Annika et al. (2024). "Understanding Metric-Related Pitfalls in Image Analysis Validation". In: *Nature Methods* 21.2, pp. 182–194.

Ren, Pengzhen et al. (2021). "A Survey of Deep Active Learning". In: *ACM Comput. Surv.* 54.9, 180:1–180:40.

Ribeiro, Vinicius, Sandra Avila, and Eduardo Valle (2019). "Handling Inter-Annotator Agreement for Automated Skin Lesion Segmentation".

Rizve, Mamshad Nayeem et al. (2021). "In Defense of Pseudo-Labeling: An Uncertainty-Aware Pseudo-label Selection Framework for Semi-Supervised Learning". In: *International Conference on Learning Representations (ICLR)*.

Robust Non-rigid Registration Challenge for Expansion Microscopy - Grand Challenge (2024). grand-challenge.org. URL: `https://rnr-exm.grand-challenge.org/` (visited on 08/22/2024).

Roger, Andrew J., Sergio A. Muñoz-Gómez, and Ryoma Kamikawa (2017). "The Origin and Diversification of Mitochondria". In: *Current Biology* 27.21, R1177–R1192.

Rohani, Ali et al. (2020). "Mito Hacker: A Set of Tools to Enable High-Throughput Analysis of Mitochondrial Network Morphology". In: *Scientific Reports* 10.1 (1), p. 18941.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 234–241.

Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." In: *Psychological Review* 65.6, pp. 386–408.

Rust, Michael J., Mark Bates, and Xiaowei Zhuang (2006). "Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM)". In: *Nature Methods* 3.10, pp. 793–796.

Sagi, Omer and Lior Rokach (2018). "Ensemble Learning: A Survey". In: *WIREs Data Mining and Knowledge Discovery* 8.4, e1249.

Sajjadi, Mehdi, Mehran Javanmardi, and Tolga Tasdizen (2016). "Regularization With Stochastic Transformations and Perturbations for Deep Semi-Supervised Learning". In: *Advances in Neural Information Processing Systems*. Vol. 29.

Sandler, Mark et al. (2018). "MobileNetV2: Inverted Residuals and Linear Bottlenecks". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520.

Sauer, Alexander et al. (2023). "Leveraging Inter-Annotator Disagreement for Semi-Supervised Segmentation". In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), pp. 1–5.

Sauer, Alexander et al. (2024). "Refining Biologically Inconsistent Segmentation Masks with Masked Autoencoders". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6904–6912.

Scheffler, Immo E. (2008). *Mitochondria*. 2nd ed. Hoboken, N.J: Wiley-Liss. xviii+462.

Schermelleh, Lothar et al. (2019). "Super-Resolution Microscopy Demystified". In: *Nature Cell Biology* 21.1, pp. 72–84.

Schlemper, Jo et al. (2019). "Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images". In: *Medical image analysis* 53, pp. 197–207.

Schuh, Claus D. et al. (2018). "Combined Structural and Functional Imaging of the Kidney Reveals Major Axial Differences in Proximal Tubule Endocytosis". In: *Journal of the American Society of Nephrology : JASN* 29.11, pp. 2696–2712.

Shelhamer, Evan, Jonathan Long, and Trevor Darrell (2017). "Fully Convolutional Networks for Semantic Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4, pp. 640–651.

Shi, Yinghuan et al. (2022). "Inconsistency-Aware Uncertainty Estimation for Semi-Supervised Medical Image Segmentation". In: *IEEE Transactions on Medical Imaging* 41.3, pp. 608–620.

Siddique, Nahian et al. (2021a). "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications". In: *IEEE Access* 9, pp. 82031–82057.

— (2021b). "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications". In: *IEEE Access* 9, pp. 82031–82057.

Siekevitz, Philip (1957). "Powerhouse of the Cell". In: *Scientific American* 197.1, pp. 131–144.

Sjostrand, F. S. (1953). "Electron Microscopy of Mitochondria and Cytoplasmic Double Membranes". In: *Nature* 171.4340, pp. 30–32.

Sjöstrand, Fritiof S. (1956). "The Ultrastructure of Cells as Revealed by the Electron Microscope". In: *International Review of Cytology*. Ed. by G. H. Bourne and J. F. Danielli. Vol. 5. Academic Press, pp. 455–533.

Smith, Samuel L. et al. (2023). *ConvNets Match Vision Transformers at Scale*. URL: `http://arxiv.org/abs/2310.16764` (visited on 12/03/2024). Pre-published.

Sobel, Irwin and G Feldman (1968). "An Isotropic 3x3 Image Gradient Operator for Image Processing". In: *Mach. Vis. Three–Dimens. Scenes* (June), pp. 376–379.

Sommer, Christoph et al. (2011). "Ilastik: Interactive Learning and Segmentation Toolkit". In: *Proceedings - International Symposium on Biomedical Imaging*, pp. 230–233.

Song, Hwanjun et al. (2023). "Learning From Noisy Labels With Deep Neural Networks: A Survey". In: *IEEE Transactions on Neural Networks and Learning Systems* 34.11, pp. 8135–8153.

Stenman, Sebastian et al. (2021). "Antibody Supervised Training of a Deep Learning Based Algorithm for Leukocyte Segmentation in Papillary Thyroid Carcinoma". In: *IEEE Journal of Biomedical and Health Informatics* 25.2, pp. 422–428.

Stephan, Till et al. (2020). "MICOS Assembly Controls Mitochondrial Inner Membrane Remodeling and Crista Junction Redistribution to Mediate Cristae Formation". In: *The EMBO Journal* 39.14, e104105.

Stoldt, Stefan et al. (2018). "Spatial Orchestration of Mitochondrial Translation and OXPHOS Complex Assembly". In: *Nature Cell Biology* 20.5, pp. 528–534.

Stringer, Carsen et al. (2021). "Cellpose: A Generalist Algorithm for Cellular Segmentation". In: *Nature Methods* 18.1, pp. 100–106.

Sudre, Carole H. et al. (2019). "Let's Agree to Disagree: Learning Highly Debatable Multirater Labelling". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 665–673.

Suga, Shogo et al. (2021). "An Interactive Deep Learning-Based Approach Reveals Mitochondrial Cristae Topologies". In: *bioRxiv*, p. 2021.06.11.448083.

Suga, Shogo et al. (2023). "An Interactive Deep Learning-Based Approach Reveals Mitochondrial Cristae Topologies". In: *PLOS Biology* 21.8, e3002246.

Sutton, Richard (2019). *The Bitter Lesson*. URL: `http://www.incompleteideas.net/IncIdeas/BitterLesson.html` (visited on 05/26/2023).

Szendroedi, Julia, Esther Phielix, and Michael Roden (2012). "The Role of Mitochondria in Insulin Resistance and Type 2 Diabetes Mellitus". In: *Nature Reviews Endocrinology* 8.2, pp. 92–103.

Tagliasacchi, Andrea et al. (2016). "3D Skeletons: A State-of-the-Art Report". In: *Computer Graphics Forum* 35.2, pp. 573–597.

Tait, Stephen W. G. and Douglas R. Green (2012). "Mitochondria and Cell Signalling". In: *Journal of Cell Science* 125.4, pp. 807–815.

Tan, Mingxing and Quoc Le (2019). "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, pp. 6105–6114.

Tanaka, Toyoichi et al. (1980). "Phase Transitions in Ionic Gels". In: *Physical Review Letters* 45.20, pp. 1636–1639.

Tarvainen, Antti and Harri Valpola (2017). "Mean Teachers Are Better Role Models: Weight-averaged Consistency Targets Improve Semi-Supervised Deep Learning Results". In: *Advances in Neural Information Processing Systems*. Vol. 30.

Thompson, Neil et al. (2023). "The Computational Limits of Deep Learning". In: *Ninth Computing within Limits 2023*. Ninth Computing within Limits 2023. Virtual: LIMITS.

Tilborghs, Sofie et al. (2022). "The Dice Loss in the Context of Missing or Empty Labels: Introducing Φ and ϵ". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Ed. by Linwei Wang et al. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, pp. 527–537.

Toldo, Marco et al. (2020). "Unsupervised Domain Adaptation in Semantic Segmentation: A Review". In: *Technologies* 8.2 (2), p. 35.

Tolstikhin, Ilya O. et al. (2021). "Mlp-Mixer: An All-Mlp Architecture for Vision". In: *Advances in neural information processing systems* 34, pp. 24261–24272.

Trimmer, Patricia A. et al. (2000). "Abnormal Mitochondrial Morphology in Sporadic Parkinson's and Alzheimer's Disease Cybrid Cell Lines". In: *Experimental Neurology* 162.1, pp. 37–50.

Troeder, Simon (2014). *HeLa Cells with Fluorescent Stained Mitochondria.* URL: `https://commons.wikimedia.org/wiki/File:HeLa_mtGFP.tif` (visited on 08/13/2024).

Van der Laan, Martin, Susanne E Horvath, and Nikolaus Pfanner (2016). "Mitochondrial Contact Site and Cristae Organizing System". In: *Current Opinion in Cell Biology*. Cell Organelles 41, pp. 33–42.

Vaswani, Ashish et al. (2017). "Attention Is All You Need". In: *arXiv* (Nips).

Villarreal, Mariana Ruiz (2006). *File:Animal Mitochondrion Diagram En.Svg - Wikipedia.* URL: `https://commons.wikimedia.org/wiki/File:Animal_mitochondrion_diagram_en.svg` (visited on 08/13/2024).

Vincent, Amy E. et al. (2016). "The Spectrum of Mitochondrial Ultrastructural Defects in Mitochondrial Myopathy". In: *Scientific Reports* 6.1, p. 30610.

Vincent, Amy E. et al. (2019). "Quantitative 3D Mapping of the Human Skeletal Muscle Mitochondrial Network". In: *Cell Reports* 26.4, 996–1009.e4.

Vitaladevuni, Shiv et al. (2008). "Mitochondria Detection in Electron Microscopy Images". In: *Workshop on Microscopic Image Analysis with Applications in Biology*.

Wadell, Hakon (1935). "Volume, Shape, and Roundness of Quartz Particles". In: *The Journal of Geology* 43.3, pp. 250–280.

Yu-Wai-Man, Patrick, Philip G. Griffiths, and Patrick F. Chinnery (2011). "Mitochondrial Optic Neuropathies – Disease Mechanisms and Therapeutic Strategies". In: *Progress in Retinal and Eye Research* 30.2, pp. 81–114.

Walt, Stéfan van der et al. (2014). "Scikit-Image: Image Processing in Python". In: *PeerJ* 2, e453.

Wang, Chenhao et al. (2024a). "A Semi-Automatic Method for Extracting Mitochondrial Cristae Characteristics from 3D Focused Ion Beam Scanning Electron Microscopy Data". In: *Communications Biology* 7.1, pp. 1–11.

Wang, Nanyang et al. (2021a). "Pixel2Mesh: 3D Mesh Model Generation via Image Guided Deformation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.10, pp. 3600–3613.

Wang, Xinsheng et al. (2024b). *MitoStructSeg: A Comprehensive Platform for Mitochondrial Structure Segmentation and Analysis*. URL: https://www.biorxiv.org/content/10.1101/2024.06.28.601295v1 (visited on 08/22/2024). Pre-published.

Wang, Yisen et al. (2019). "Symmetric Cross Entropy for Robust Learning With Noisy Labels". In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 322–330.

Wang, Yuchao et al. (2022). "Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4248–4257.

Wang, Yunpeng et al. (2021b). "DeepSDM: Boundary-aware Pneumothorax Segmentation in Chest X-ray Images". In: *Neurocomputing* 454, pp. 201–211.

Warfield, Simon K., Kelly H. Zou, and William M. Wells (2004). "Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation". In: *IEEE Transactions on Medical Imaging* 23.7, p. 903.

Wei, Chen et al. (2023). "Diffusion Models as Masked Autoencoders". In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, pp. 16238–16248.

Wei, Donglai et al. (2020a). "MitoEM Dataset: Large-Scale 3D Mitochondria Instance Segmentation from EM Images". In: *Medical Image Computing and Computer Assisted Intervention*. Springer, Cham, pp. 66–76.

— (2020b). "MitoEM Dataset: Large-Scale 3D Mitochondria Instance Segmentation from EM Images". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Vol. 12265 LNCS, pp. 66–76.

Weigert, Martin et al. (2020). "Star-Convex Polyhedra for 3D Object Detection and Segmentation in Microscopy". In: *IEEE Winter Conference on Applications of Computer Vision*, pp. 3666–3673.

Weinstein, John N. et al. (2013). "The Cancer Genome Atlas Pan-Cancer Analysis Project". In: *Nature Genetics* 45.10 (10), pp. 1113–1120.

Wen, Gang et al. (2023). "Current Progress in Expansion Microscopy: Chemical Strategies and Applications". In: *Chemical Reviews* 123.6, pp. 3299–3323.

Wickramasinghe, Udaranga et al. (2020). "Voxel2Mesh: 3D Mesh Model Generation from Volumetric Data". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 299–308.

Włodarczyk, Tomasz et al. (2019). "Estimation of Preterm Birth Markers with U-Net Segmentation Network". In: *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis*. Cham: Springer International Publishing, pp. 95–103.

Wu, Ching-An et al. (2013). "Nutrient Deprivation Induces the Warburg Effect through ROS/AMPK-dependent Activation of Pyruvate Dehydrogenase Kinase". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1833.5, pp. 1147–1156.

Wu, Yuxin and Kaiming He (2018). "Group Normalization". In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19.

Xia, Yingda et al. (2020). "Uncertainty-Aware Multi-View Co-Training for Semi-Supervised Medical Image Segmentation and Domain Adaptation". In: *Medical Image Analysis* 65.

Xiang, Jinyi, Peng Qiu, and Yang Yang (2022). "FUSSNet: Fusing Two Sources of Uncertainty for Semi-supervised Medical Image Segmentation". In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 481–491.

Xiao, Yunpeng, Youpeng Zhao, and Ge Yang (2022). "Semi-Supervised Segmentation of Mitochondria from Electron Microscopy Images Using Spatial Continuity". In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5.

Xie, Ronald et al. (2023). "MAESTER: Masked Autoencoder Guided Segmentation at Pixel Resolution for Accurate, Self-Supervised Subcellular Structure Recognition". In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, pp. 3292–3301.

Xue, Yuan et al. (2020). "Shape-Aware Organ Segmentation by Predicting Signed Distance Maps". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.07, pp. 12565–12572.

Yamamoto, Hayashi et al. (2011). "Dual Role of the Receptor Tom20 in Specificity and Efficiency of Protein Import into Mitochondria". In: *Proceedings of the National Academy of Sciences* 108.1, pp. 91–96.

Yoon, Yisang et al. (2011). "Mitochondrial Dynamics in Diabetes". In: *Antioxidants & Redox Signaling* 14.3, p. 439.

Zahedi, Atena et al. (2018). "Deep Analysis of Mitochondria and Cell Health Using Machine Learning". In: *Scientific Reports* 8.1, pp. 1–15.

Zeng, Dewen et al. (2021). "Segmentation with Multiple Acceptable Annotations: A Case Study of Myocardial Segmentation in Contrast Echocardiography". In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 478–491.

Zhang, Aston et al. (2023). *Dive into Deep Learning*. Cambridge University Press.

Zhang, Le et al. (2020a). "Learning to Segment When Experts Disagree". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 179–190.

Zhang, Ruiyi et al. (2022). "Nonparametric K-Sample Test on Shape Spaces with Applications to Mitochondrial Shape Analysis". In: *Journal of the Royal Statistical Society. Series C: Applied Statistics* 71.1, pp. 51–69.

Zhang, Zhilu and Mert Sabuncu (2018). "Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels". In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc.

Zhang, Ziang et al. (2020b). "DENSE-INception U-net for Medical Image Segmentation". In: *Computer Methods and Programs in Biomedicine* 192, p. 105395.

Zheng, Xu et al. (2022). "Uncertainty-Aware Deep Co-training for Semi-supervised Medical Image Segmentation". In: *Computers in Biology and Medicine* 149, p. 106051.

Zick, Michael, Regina Rabl, and Andreas S. Reichert (2009). "Cristae Formation—Linking Ultrastructure and Function of Mitochondria". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. Assembly of the Mitochondrial Respiratory Chain 1793.1, pp. 5–19.

Zou, Zhengxia et al. (2023). "Object Detection in 20 Years: A Survey". In: *Proceedings of the IEEE* 111.3, pp. 257–276.