



Automated pre-processing of fluorescence microscopy data for the analysis of mitophagy

by

Richard Batt

Thesis presented in fulfilment of the requirements for the degree of
Master of Engineering (Electronic) in the Faculty of Engineering at
Stellenbosch University.

Supervisor: Dr. R.P. Theart

November 2023

Acknowledgements

I would like to thank my dog, Muffin. I also would like to thank the inventor of the incubator; without him/her, I would not be here. Finally, I would like to thank Dr Herman Kamper for this amazing report template.

Plagiaatverklaring / Plagiarism Declaration

1. Plagiaat is die oorneem en gebruik van die idees, materiaal en ander intellektuele eiendom van ander persone asof dit jou eie werk is.

Plagiarism is the use of ideas, material and other intellectual property of another's work and to present it as my own.

2. Ek erken dat die pleeg van plagiaat 'n strafbare oortreding is aangesien dit 'n vorm van diefstal is.

I agree that plagiarism is a punishable offence because it constitutes theft.

3. Ek verstaan ook dat direkte vertalings plagiaat is.

I also understand that direct translations are plagiarism.

4. Dienooreenkomsdig is alle aanhalings en bydraes vanuit enige bron (ingesluit die internet) volledig verwys (erken). Ek erken dat die woordelikse aanhaal van teks sonder aanhalingsstekens (selfs al word die bron volledig erken) plagiaat is.

Accordingly all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism

5. Ek verklaar dat die werk in hierdie skryfstuk vervat, behalwe waar anders aangedui, my eie oorspronklike werk is en dat ek dit nie vantevore in die geheel of gedeeltelik ingehandig het vir bepunting in hierdie module/werkstuk of 'n ander module/werkstuk nie.

I declare that the work contained in this assignment, except where otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

| | |
|---|--|
| 99652154 |  |
| Studentenommer / Student number | Handtekening / Signature |
| L. Skywalker Voorletters en van / Initials and surname | 1 October 2099 Datum / Date |

Abstract

The English abstract.

Uittreksel

Die Afrikaanse uittreksel.

Contents

| | |
|---|------|
| Declaration | ii |
| Abstract | iii |
| Opsomming | iv |
| Nomenclature | viii |
| 1. Introduction | 1 |
| 1.1. Background | 1 |
| 1.1.1. Motivation for mitophagy research | 1 |
| 1.1.2. The investigation of mitophagy | 1 |
| 1.1.3. Preparation of images prior to analysis | 2 |
| 1.1.4. High-throughput pre-processing | 4 |
| 1.2. Problem Statement | 4 |
| 1.3. Project objectives | 5 |
| 1.4. Scope | 6 |
| 1.5. Contributions | 6 |
| 1.5.1. Contributions towards the Memantine research paper | 6 |
| 1.5.2. Contributions towards relevant fields | 7 |
| 1.6. Structure of thesis | 8 |
| 2. Theoretical Background & Fundamentals | 9 |
| 2.1. Mitophagy and involved organelles | 9 |
| 2.1.1. Mitochondria | 9 |
| 2.1.2. Lysosomes | 11 |
| 2.1.3. Autophagosomes | 11 |
| 2.1.4. Mitophagy | 12 |
| 2.2. Fluorescence Microscopy | 13 |
| 2.2.1. Light microscopy and motivations for fluorescence microscopy | 13 |
| 2.2.2. Factors regarding fluorophore selection for fluorescence microscopy . | 15 |
| 2.2.3. Constraints on the performance of fluorescence microscopy | 17 |
| 2.2.4. Confocal Microscopy | 19 |
| 2.2.5. Wide-field and confocal microscope apparatus | 22 |
| 2.2.6. Summary | 23 |

| | |
|---|------------|
| 2.3. Colocalization | 23 |
| 2.4. Image preprocessing | 24 |
| 2.4.1. Sources of interference in image quality | 25 |
| 2.4.2. Standard pipeline for image pre-processing | 26 |
| 2.4.3. Overview of the methods to restore image quality | 28 |
| 2.5. Further investigation into prominent deconvolution and image quality restoration methods | 37 |
| 2.5.1. Otsu thresholding | 37 |
| 2.5.2. Hysteresis thresholding | 37 |
| 2.5.3. Triangle thresholding | 40 |
| 2.5.4. Richardson-Lucy deconvolution | 41 |
| 2.6. Other techniques implemented within this research | 43 |
| 2.6.1. Kneedle algorithm | 43 |
| 2.7. Prominent literature used in this research | 46 |
| 2.8. Chapter conclusions | 48 |
| 3. Methodology | 50 |
| 3.1. Dataset acquisition process and preparation | 50 |
| 3.1.1. Sample preparation prior to imaging | 50 |
| 3.1.2. Image acquisition via microscopy | 51 |
| 3.1.3. Preprocessing prior to thresholding | 51 |
| 3.2. Automatic thresholding method | 53 |
| 3.2.1. Choice of threshold method | 53 |
| 3.2.2. Acquisition of the low threshold | 54 |
| 3.2.3. High threshold acquistion | 62 |
| 3.2.4. Iterative hysteresis histogram (IHH) | 63 |
| 3.2.5. Selecting a high threshold | 69 |
| 3.2.6. Applying biases to the centroid result | 71 |
| 4. System performance and observations | 82 |
| 4.1. Shortlisting viable thresholding methods for the evaluation | 82 |
| 4.1.1. Global threshold shortlisting | 85 |
| 4.1.2. Local threshold shortlisting | 86 |
| 4.2. Visual analysis | 90 |
| 4.2.1. Lysosome structures | 91 |
| 4.2.2. Autophagosome structures | 99 |
| 5. Conclusion | 107 |
| 5.1. | 107 |

| | |
|---|------------|
| Bibliography | 108 |
| A. Threshold outcomes for method selection | 114 |
| A.1. Global threshold outcomes of shortlist samples | 114 |
| A.2. Local threshold outcomes of shortlist samples | 118 |
| B. Visual analysis thresholding outcome images | 128 |
| B.1. Global threshold outcomes | 131 |
| B.2. Local threshold outcomes | 140 |
| B.3. AHT outcomes | 144 |

Nomenclature

Variables and functions

| | |
|---------------------------------------|--|
| $p(x)$ | Probability density function with respect to variable x . |
| $P(A)$ | Probability of event A occurring. |
| ε | The Bayes error. |
| ε_u | The Bhattacharyya bound. |
| B | The Bhattacharyya distance. |
| s | An HMM state. A subscript is used to refer to a particular state, e.g. s_i refers to the i^{th} state of an HMM. |
| \mathbf{S} | A set of HMM states. |
| \mathbf{F} | A set of frames. |
| \mathbf{o}_f | Observation (feature) vector associated with frame f . |
| $\gamma_s(\mathbf{o}_f)$ | A posteriori probability of the observation vector \mathbf{o}_f being generated by HMM state s . |
| μ | Statistical mean vector. |
| Σ | Statistical covariance matrix. |
| $L(\mathbf{S})$ | Log likelihood of the set of HMM states \mathbf{S} generating the training set observation vectors assigned to the states in that set. |
| $\mathcal{N}(\mathbf{x} \mu, \Sigma)$ | Multivariate Gaussian PDF with mean μ and covariance matrix Σ . |
| a_{ij} | The probability of a transition from HMM state s_i to state s_j . |
| N | Total number of frames or number of tokens, depending on the context. |
| D | Number of deletion errors. |
| I | Number of insertion errors. |
| S | Number of substitution errors. |

Acronyms and abbreviations

| | |
|------|---|
| AE | Afrikaans English |
| AID | accent identification |
| IAT | Improved automatic thresholding |
| SNR | Signal-to-noise ratio |
| PSNR | Peak signal-to-noise ratio |
| NA | Numerical aperture |
| PSF | Point spread function |
| RACC | Regression adjusted colocalisation colour mapping |
| DMEM | Dulbecco's Modified Eagle Medium |
| IHH | Iterative hysteresis histogram |
| MIP | Maximum intensity projection |
| RGB | Red, green and blue |

Chapter 1

Introduction

1.1. Background

1.1.1. Motivation for mitophagy research

The human body is a multicellular organism comprised of many different organs and tissues, each of which is composed of an immense number of different types of specialised cells working together to enable the biological functions that allow us to live and act. The vast majority of these cells contain a number of specialised components fulfilling a number of functions and important among them is the mitochondria. The mitochondrion is an organelle that has developed to efficiently process food molecules, absorbed by the body, and convert them into a form of chemical energy that is easily used by the other organelles within the cell. This is crucial as cells with specialised functions in turn have specialised organelles to achieve them and typically the energy requirement is relatively high. Due to chemical wear and tear from performing their functions, or due to external forces, the efficiency and health of mitochondria can diminish and at their worst can become a hindrance to the healthy supply of energy for the cell.

When this occurs a process called mitophagy activates where the irreparably damaged mitochondria are flagged, isolated, and broken down for the biomolecules composing it to be reused by the cell [1]. There are two organelles involved in this process and they are the autophagosome, involved with isolating and transporting the flagged mitochondria, and the lysosome, triggering the decomposition process which breaks down the mitochondria. There is a risk though of this mitophagy not activating correctly, or being dysfunctional, and is associated with a number of human illnesses particularly those that are neurodegenerative such as Alzheimer's disease and Parkinson's disease [2]. It is for this reason that research into the mechanisms behind mitophagy, the causes of the dysfunction, and the correction of it is of value.

1.1.2. The investigation of mitophagy

The investigation behind the causes of dysfunction in mitophagy, and methods by which to correct it, can be approached in a number of ways. For studying the regulation and

correction of mitophagy dysfunction it is more accessible to observe the behaviours of the mitophagy-relevant organelles using cell cultures. These observations can be performed either visually, chemically, or both to determine which organelle behaviours are associated with functional or dysfunctional mitophagy. To achieve this, a common approach is to apply treatments that target specific factors (chemical, thermal, physical, etc.) that stress the mitochondria. Once the mitochondria are sufficiently stressed, mitophagy will be triggered allowing us to observe the actions of the mitochondria and involved organelles (autophagosomes and lysosomes). Treatments can target different factors to measure their influence in triggering mitophagy or can induce dysfunction in mitophagy. Visual observation of these events provides greater spatial information with regard to the organelles (position, size, count) especially with many interactions being spatial (engulfment, joining, separation, etc.) and is achieved through microscopy imaging.

1.1.3. Preparation of images prior to analysis

The purpose of imaging is to capture a representation of the specimen that can be measured by visual and image analysis with the former being performed by humans with qualitative (e.g. estimating the shape of the subjects in the image) or quantitative (e.g. manually counting each individual subject) measures while the latter is performed by machines through algorithms that are able to provide more detailed quantitative measurements (e.g. the calculated distribution of subject sizes within a population which provides measures such as the mean or variance of the sizes). An obstacle to the accuracy of both the visual and image analysis is image degradation which is the loss of image quality due to the presence of a number of factors in the image. There are a number of different factors causing this image degradation but the three most prominent are noise, distortion and artefacts. Noise typically causes a misrepresentation of the intensities within an image and generally acts as an additive component [3]. Distortion is a deformation in the spatial representation of the specimen being imaged with the shape and position of structures primarily being affected. Artefacts are a broad group that refers to any anomalous elements in the image that do not fall under the prior factors and could affect the intensity or spatial representations of the specimen. A common attribute among these is that they are undesired in the image and either obscure the structures of interest, misrepresent their features (e.g. size, shape, texture, etc.); or describe non-existent, or unintended, structures. If this image degradation remains then analysis of the image is less reliable with the results being skewed depending on the degradation factors present. Unfortunately, there is always some degradation present in all images but the impact depends on both the quantity of these factors and the sensitivity of the features that are being analyzed (e.g. structures or structural features that are small or fine in size are usually more sensitive).

Restoring image quality: The restoration of image quality and clarity has motivated the development of a number of techniques over many years. These techniques typically aim to reduce the impact of the aforementioned factors such that the image becomes a closer representation of the specimen improving the accuracy and reliability of further analyses. These techniques have typically been developed to target specific factor(s) (e.g. removing specific types of noise or distortion specifically) with each having optimal conditions (e.g. consistent illumination, high contrast between foreground and background, etc.) to perform effectively and when these conditions are not met the outcomes can be unsatisfactory for any analysis. It is for this reason that the applied techniques need to be correctly chosen based on the factors present within the image. These techniques can be roughly grouped together based on their intended outcome whether it be deconvolution techniques reducing the distortion caused by microscope diffraction; denoising techniques which reduce the presence of noise; image enhancement techniques which emphasise or correct image properties such as improving contrast or smoothing unequal illumination across the image; and binarizing techniques which segment structures between the foreground and background with the foreground typically being what is of interest to analysis. These are merely categories and based on the criteria and desired granularity sub-groups can be made such as “wavelet filters” and “patch-based filters” which may still share overlapping outcomes.

Binarizing techniques: These techniques focus on the separation of pixels between a foreground and background resulting in a binary image. The application of this can be quite broad but for the purposes of this research, the binary separation of pixels allows clear differentiation between what is and is not of interest to succeeding analyses. A commonly employed set of techniques in image analysis, to achieve binarization, are thresholding techniques where numeric threshold(s) are calculated from the image and used to determine whether a pixel belongs to the foreground or background based on whether the pixel intensity is greater than the threshold or not, respectively [4]. Effective binarization enables the selectivity of structures that can be of particularly great benefit to any image analyses employing quantitative measurements that are highly sensitive to the image quality (e.g. Gaussian noise can greatly compromise structure count measurements). Through this, the measurements taken can be isolated to only the foreground structures where the degradation factors may be weaker.

Thresholding groups: As mentioned previously, thresholding is a common method of achieving binarization but typically falls into two groups for image processing. These are global thresholding and local thresholding methods with each having their own respective pros and cons.

| Global methods | Local methods |
|---|---|
| <ul style="list-style-type: none"> • Relies on the image histogram to determine the threshold value. • Applies this threshold to all image pixels naively. • Sensitive to both distortions of the image histogram and localised variations in image intensity. • Require few, if any, parameters for optimal performance. | <ul style="list-style-type: none"> • Determines the threshold of each pixel by evaluating the intensity characteristics of the localised region centred on it. • Robust to localised characteristics in images (e.g. variations in contrast around structures). • Performance is dependent on one or more parameters that may need to be tuned for each image for optimal results. |

Despite this, both methods can still be sensitive to other degradation factors present in the image regardless of whether a global or local method is used.

1.1.4. High-throughput pre-processing

High-throughput pre-processing is where the pre-processing of images is optimised for both speed and the resultant image quality. This is usually applied to large batches of images and can form part of a larger high-throughput analysis sequence with the goal of expediting image analysis while maintaining accuracy. To optimize the improvements to the image quality achieved by the pre-processing it is not uncommon to sequentially apply pre-processing techniques to images in a pipeline. This acts as a synergistic process where the outcomes of preceding techniques in the pipeline can optimize the conditions for succeeding techniques allowing them to produce higher quality outcomes than if they were applied in isolation. A key component of high-throughput analysis is speed which is achieved by minimizing interruptions in the pre-processing by using automated techniques that do not require parameters to be tuned to achieve adequate results.

1.2. Problem Statement

After imaging a large number of various cells, all under identical imaging conditions (e.g. resolution, brightness, gain, etc.), it is often the case that high-throughput pre-processing be implemented as part of preparing them for analysis. This approach ideally enables efficient and effective pre-processing, thereby saving time for studies and projects dealing with extensive batches of images. However, the involvement of binarization in this process can pose a challenge to the goals of high-throughput pre-processing.

Binarization is challenging in high-throughput pre-processing due to the constraints that automatic global and local thresholding methods can impose. Global thresholding has a high sensitivity to intensity-related factors (e.g. noise) and image properties, such as contrast or brightness, that affect the histogram particularly when these vary over localised regions of the image. This can result in inconsistent results that can vary greatly based on each image. Local thresholding is more robust to localised variations across the image and is unaffected by the histogram but the quality of the results is dependent on the parameter tuning. Parameter tuning is a human-involved process of adjusting parameter values to attain the most accurate thresholding results but can be time-consuming to perform and the degree of tuning necessary can be worsened by the presence of degradation factors. This results in a situation where global thresholds provide simplicity of application but uncertainty in behaviour while local thresholding can provide robust results given that the parameters are sufficiently tuned for each image, disrupting the pacing of the high-throughput pre-processing.

This is not a constant where pre-selected parameter values may be sufficient for a local threshold technique or the intensity-related image properties are consistent enough that there is minimal variation for global thresholding techniques. Resolving this problem through a solution that provides both the robustness of local thresholds with the simplicity of global thresholds would provide more opportunities for studies requiring the pre-processing of large image batches to expedite this research and save time. Providing an effective binarization solution could reduce the outcome of poor binarization which ranges from the inclusion of noise as structures of interest for analysis or the exclusion of structures important to the study. While machine learning solutions to similar problems exist [5], their performance can be dependent on both the type and quantity of data they are trained on but this training data can be expensive or time-consuming to acquire.

1.3. Project objectives

To resolve this problem of binarization in high-throughput pre-processing a system will be designed based on these objectives:

1. Optimize the binarization outcomes without parameter tuning being required.
2. Be an automated method that can be integrated into a high-throughput pipeline.
3. Most provide results similar to or exceeding established methods such as global or thresholding methods in a high-throughput context (no parameter tuning).

These objectives will guide the design of a deterministic binarization system which will be referred to as Improved Automatic Threshold (IAT).

1.4. Scope

The constraints on the expectations for the development and evaluation of this binarization solution are:

1. Commonly extracted structure features of the image that can be quantified by existing libraries (e.g. scikit-image [6], SciPy [7], etc.) will be used to determine the binarization. Structure count, structure volumes, intensity distribution across structures, and structure-specific intensities. Combinations and transformations (e.g. averages, derivatives, etc.) of these features will also be considered.
2. The system will be agnostic to image resolution and freely be able to process 2D and 3D images.
3. The developed method will be evaluated against existing automatic global and local thresholding methods both quantitatively, using metrics such as signal-to-noise ratio (SNR), and qualitatively. The qualitative evaluations will entail visual analysis of the thresholding outcomes and will also be compared to manually thresholded images but not as extensively as the quantitative evaluation due to time constraints.
4. This research is pursued application to imaging relevant to the study of mitophagy thus only the imaged mitochondria, autophagosomes, and lysosomes will be used for development and evaluation.
5. Only images for a single type of microscopy will be used for both the development and evaluation of this method due to both time and resource constraints.

1.5. Contributions

Contributions made through this research were to the paper *The Highs and Lows of Memantine—An Autophagy and Mitophagy Inducing Agent That Protects Mitochondria* [8] by assisting in sample pre-processing and image quality evaluation prior to analysis. Further contributions that are discussed are the potential contributions that this research could provide towards the image analysis, and related, fields.

1.5.1. Contributions towards the Memantine research paper

The IAT developed in this project originated from the application in high-throughput pre-processing of mitophagy-related organelle images to expedite data set preparation for mitophagy analysis. Coincidentally, research regarding the application of Memantine treatments and the analysis of its effects on mitophagy behaviour [8] was pursued. Despite the IAT project shifting focus from mitophagy analysis to image pre-processing, with a

focus on mitophagy-related organelle images, support was provided to the Memantine research with regard to the image pre-processing and the evaluation of the image quality prior to mitophagy analysis.

First contribution: The development of a pre-processing pipeline macro for Fiji [9] to automate the pre-processing of 3D confocal fluorescence microscopy images of cellular organelles which are mitochondria, lysosomes, and autophagosomes. The pipeline used was based off of the pipeline described in [10] (excluding the binarization and all succeeding algorithms) with adjustments made detailed in Section 3.1.3.

Second contribution: With pre-processing having been applied, the quality of these images were evaluated under three criteria:

1. Sufficient contrast of the structures of interest against the background, and any residual noise, for accurate differentiation.
2. The lack of fluorescence falsely describing in-focus structures originating from either out-of-focus fluorophores, fluorescence cross-talk, and autofluorescence.
3. Confident clustering of organelles in the images between separate cells if present.

The first criterion being met implies that structures of interest can be differentiated and selected for algorithmic analyses where those results could be disturbed by erroneously including noise. The second criterion being met implies that aberrations in the image that are more challenging to automatically remove, without degrading the image signal, are not strongly present and will not distort measured results. The final criterion being met implies that per cell measurements are possible to capture for an image which is necessary for certain analyses.

1.5.2. Contributions towards relevant fields

The fields of relevance to this research are image analysis, with the focus on image binarization, and cellular physiology, as the data used in throughout the IAT development involved cellular organelle images. This will provide a binarization solution accessible to those with low expertise in image pre-processing while providing viable results particularly for high-throughput applications. For image analysis, the contribution will be the establishment of a new automated thresholding method with a novel approach to approximating binarized images.

1.6. Structure of thesis

This thesis is composed of 6, including Chapter 1, with the remaining 5 Chapters summarised below:

Chapter 2 provides context around many facets of this research project to build a fundamental understanding of certain decisions in the system design and the phenomena that are observed. These fundamentals are composed of context for the organelles being imaged in regard to their characteristics; the microscopy techniques used and how they function; a high-level overview of the pre-processing method groups and the roles each play in improving image quality; and specific methods that are of note to the system design.

Chapter 3 will present the rationale behind the preparation of the specimen which is imaged, the pre-processing techniques applied to improve the image quality, and the design of the system IAT to binarize the objects of interest with observations and tests to justify said design choices.

Chapter 4 will present and discuss the performance results of the IAT against established thresholding methods.

Chapter 5 will provide a conclusion regarding the performance of the IAT method, the shortcomings, and improvements to be made through future work.

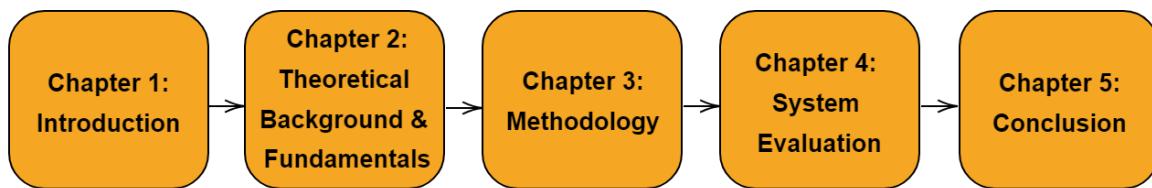


Figure 1.1: Diagram of thesis structure

Chapter 2

Theoretical Background & Fundamentals

In Chapter 1 the research regarding mitophagy is mentioned and the research to develop automated thresholding methods to streamline the data preparation process is detailed in terms of objectives and the proposed use case. Before detailing related research or the design of the proposed thresholding system an initial establishment of theoretical information regarding the mitophagy research use case will be presented. This theoretical information will regard the organelles being imaged for mitophagy analysis, the mitophagy process itself, the microscopy techniques employed for imaging, an overview of image preprocessing techniques that may be employed (noise removal, deconvolution, thresholding) and other particular techniques of interest to the thresholding system design.

2.1. Mitophagy and involved organelles

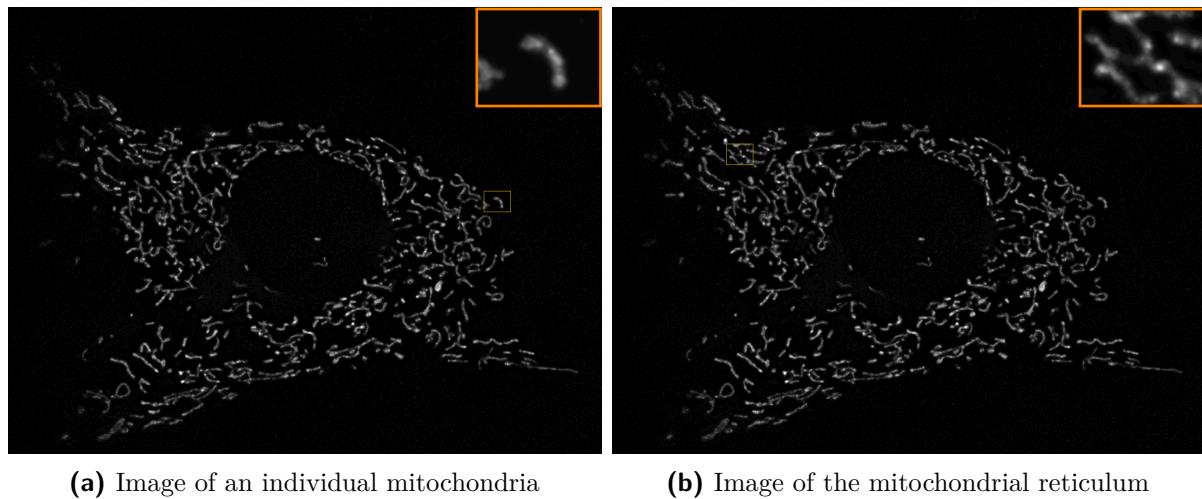
In Section 1.1 a number of potential applications of mitophagy research were presented. Of importance were the overview of the organelles involved in mitophagy, the purpose of these organelles and the importance of mitophagy with regard to these organelles. As these organelles and the mitophagy process will be studied via visual and image analysis the function but also the structure will be discussed further in this section.

2.1.1. Mitochondria

Mitochondria are organelles found in plants, animals and many other multicellular organisms that are primarily responsible for supplying the energy required by the cell. The primary functions and mitochondria structure will be discussed in this section.

Functions

The mitochondria perform a number of functions within the cell but the most notable function is the provision of fuel for the cell for the other contained organelles to play their role in maintaining the cell and executing their actions relative to the rest of the cells.



(a) Image of an individual mitochondria

(b) Image of the mitochondrial reticulum

Figure 2.1: Comparison between mitochondria structure types with (a) singular punctate mitochondria and (b) branching mitochondria forming a reticulum

This fuel is synthesized by the mitochondria in a molecule called adenosine triphosphate (ATP) which is an efficient fuel format accepted by all other energy-consuming organelles within the cell. This is crucial to the survival of cells as the many chemical and mechanical actions performed by cells can be energy expensive and the extraction of energy from food energy molecules is of a low yield with much of the desired energy remaining chemically trapped [1, p. 33, 35, 40]. For this reason, mitochondria are dedicated to producing ATP which effectively yields energy when required and the number of mitochondria in cells ranges from hundreds to thousands depending on the cell functions. This is not the sole function of mitochondria as they are instrumental in the management of controlled cell death (apoptosis) which is employed when the cell itself is too damaged for repair deeming it to be of risk to the surrounding cells [1, p. 40-41].

Structure

The structure of mitochondria is that of an oval-like structure with many cristae (internal folds) increasing the internal surface area and improving the efficiency of ATP synthesis process. A visualisation of this internal structure can be seen in Figure 2.1a and these mitochondria coexist separately from each other in the cell in a formation called ‘punctate’ or they can appear in a formation called a ‘reticulum’. A mitochondrial reticulum (shown in Figure 2.1b) is formed primarily in cells with a higher energy demand than usual such as skeletal muscle and this reticulum is akin to a connected branching structure. This branching allows mitochondria to support each other by sharing functional components as ATP production can cause wear to build up as a byproduct.

2.1.2. Lysosomes

Lysosomes are organelles involved with the recycling and degradation of compounds and organelles within the cell.

Function

The purpose of lysosomes is to decompose compounds within the cell into the base components which can be used to synthesize new organelles or provide needed nutrients. The lysosome is a carrier of hydrolases (an enzyme that reacts with water to break down molecules) that break down specific compounds within the cell ranging from biomolecules absorbed from outside the cell such as proteins, carbohydrates and lipids; to invasive bacteria; and even damaged organelles. The breaking down of compounds absorbed from outside of the cell (through a process called endocytosis) is used to ‘digest’ these compounds to provide a source of new nutrients. The decomposition of invasive bacteria and damaged organelles acts similarly to a defence response to maintain the health of the cell while also allowing the damaged organelles to be recycled to produce new organelles maintaining the population of healthy organelles [1, 11]. The process of decomposing these compounds is called *autophagy* and relies on autophagosomes to deliver the compounds, to be decomposed, to the lysosomes. The lysosome fuses with the autophagosome to become an *autolysosome* where the contained compounds are introduced to hydrolase enzymes initiating decomposition [11].

Structure

Lysosomes are small structures of spherical bodies composed of a membrane-like structure containing the hydrolase enzymes [1]. This membrane structure is impermeable to most interfacing to prevent the unintentional release of the contained enzymes [11].

2.1.3. Autophagosomes

Autophagosomes are the name for the membrane structure that carries compounds to be decomposed by autophagy.

Function

Autophagosomes are formed from phagophores (a cup-shaped membrane) that are located in the cytosol (cellular fluid) and stretch to engulf the compounds to be decomposed. Once the compounds have been engulfed the phagophore is referred to as an autophagosome which proceeds to carry the contents to a lysosome. An autophagosome is one of the organelles capable of interfacing with lysosomes to form an autolysosome. After fusing

together the lysosome introduces the enzymes to the cargo of the autolysosome and autophagy begins [11].

Structure

An autophagosome is initially a single membrane called a phagophore which stretches and deforms into a cup-like shape when approaching biomolecules that are supposed to undergo autophagy. The phagophore continues to stretch around the biomolecules until they are fully engulfed at which point the edges of the cup-like shape join and fuse to form a double membrane. The phagophore is now referred to as an autophagosome with a complete double membrane which has an ovular/spherical shape.

2.1.4. Mitophagy

With the relevant organelles, and their functions, briefly covered a simple description of the mitophagy process can be presented. As mentioned in Section 1.1, the dysfunction of mitophagy is associated with a number of diseases with neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's disease, etc.) of note. Mitophagy as a process refers to the isolation and subsequent decomposition of dysfunctional mitochondria. This is performed by the autophagosomes engulfing the mitochondria and transporting them to lysosomes as cargo. The autophagosomes fuse with the lysosomes introducing enzymes to the mitochondria and beginning the decomposition process. After the decomposition is completed the resultant molecular components of the mitochondria can be used by the cell in a number of ways such as synthesising new mitochondria.

The reasons why mitochondria are degraded by autophagy is typically due to irreparable mitochondria dysfunction usually caused by either physical, chemical or oxidative damage. This damage to the mitochondria can affect a number of components (mitochondrial DNA, lipids, proteins, etc.) that are necessary for many mitochondrial functions and while this damage is below a certain threshold the impact of this damage on mitochondrial functions can be mitigated through fusion. This fusion entails the joining of mitochondria to share functional components to compensate for those that are damaged allowing mitochondrial functions, like ATP synthesis, to continue [12]. Multiple fusions can occur between mitochondria allowing branching chains of mitochondria to form, mentioned in Section 2.1.1 as a mitochondrial reticulum, which is more frequently seen in cells undergoing high ATP demand such as activated skeletal muscle cells [1]. This makes the network of fused mitochondria robust to interference in ATP production when there are high energy demands.

Previously it was mentioned that fusion could compensate for damage to the mitochondria given that the damage is below a threshold but what happens when this threshold is exceeded? When this occurs the mitochondria can be seen as irreparable and if present

in a reticulum then the density of dysfunctional components could place undue strain on the fused mitochondrial collective. The first step in this process is mitochondrial fission where dysfunctional components are aggregated to a singular region within the mitochondria which are then split into two daughter mitochondria. The daughter mitochondria containing dysfunctional components can then be removed via autophagy while the healthy daughter can resume activities. In the case of a fused mitochondria network, this aggregation still occurs and only the localised region containing the dysfunctional components will be split from the network [12].

With the dysfunctional mitochondria separated from the healthy a number of mitochondria-specific biochemical pathways (mechanisms), which will not be described further here, activate flagging the dysfunctional mitochondria. With the dysfunctional mitochondria flagged and located the phagophores can engulf the mitochondria, becoming an autophagosome, and transport it to a lysosome to begin decomposition [12, 13] (described in Sections 2.1.2 and 2.1.3).

2.2. Fluorescence Microscopy

As described in Section 1.1, a particular method of studying mitophagy behaviours is to study the organelles involved in this process (detailed prior in Section 2.1). There are a number of methods by which to analyse organelle behaviours whether visual or biochemical where each captures different ranges of traits describing these behaviours. For the study of mitophagy, it is not uncommon to employ a combination of visual and biochemical analysis but it is the visual analysis that is of interest to this particular research. This section will explore the high-level motivation behind fluorescence microscopy as opposed to traditional optical microscopy, the factors to consider for suitable fluorophore selection, the limitations of fluorescence microscopy, an overview of confocal fluorescence microscopy, and a brief breakdown of the relevant microscopy apparatus.

2.2.1. Light microscopy and motivations for fluorescence microscopy

In the microscopy of both cellular and subcellular structures, there are numerous biochemical structures, biomolecules and intracellular processes that can be observed. For this research, organelles are the subcellular structures that are of interest and going forwards will be referred to simply as ‘structures’. A traditional, and still used, method to visually observe these structures is through the use of optical microscopy.

Optical microscopy exposes the biological specimen, placed on the microscope stage, to visible spectrum light illuminating the specimen allowing cells, organelles, and other biochemical structures to be seen. When imaging these specimens there are three obstacles

to making accurate and clear observations of the specimen, particularly with subcellular objects. These are the resolution limit where optical microscopes struggle to resolve features (edges, gaps between structures, etc.) that are smaller than $0.2\mu\text{m}$ [14, p.582]; thick specimens reduce the level of clear detail at higher resolutions; and that these specimens are frequently composed of majority water thus structures and features may be indistinct if not nearly invisible [14, p.585]. The challenge of thick specimens can be resolved by slicing the specimen and improving the clarity due to the high water contents can be resolved through the use of stains which bind to cell structures, bacteria and other microscopic organisms based on the stain's preferred conditions. This allows stains to be selected that will localise to specific structures or to specific cellular processes (e.g. metabolic) improving the contrast between these structures from the irrelevant surroundings. Despite this, there can still be visual interference from surrounding structures which can decrease the clarity at which fine structures or features can be discerned.

For the above paragraph, and in future, I use the `paragraph` command to head off new points in the subsubsection that is contained within the paragraph. I realise now, after going through a few sections, that there are extra spaces after the bolded paragraph heading. Do you think this looks fine, should I linebreak after the paragraph heading, or do you think I should bold text the first word (as I have done now) but without the space after which will use `textbf` and `linebreak` to replicate the effect.

Fluorescence microscopy is similar to optical microscopy except for the types of stains utilised and some adjustments to the microscope architectures. Fluorescence microscopy utilises fluorescent stains, called fluorophores, that are identical to the stains described prior except that they emit light of a specific wavelength (based on the fluorophore) when excited. This excitation of the fluorophores is likewise induced by exposing these fluorophores to the light of a fluorophore-specific wavelength allowing the selection of which structures are visible by constraining both the wavelength of light illuminating the specimen and the wavelength that is captured by the microscope. This wavelength selection for both the excitation and emission light is achieved through filters that reject wavelengths outside of the currently selected range allowing structures to be indicated by specific wavelengths [15]. For this reason, fluorescent stains can be referred to as indicators and fluorescently stained structures can be referred to as ‘indicated’ where post-imaging different colours can be assigned to these respective wavelengths improving differentiation between different types of indicated structures. A comparison between optical and fluorescence microscopy can be seen in Figure 2.2.

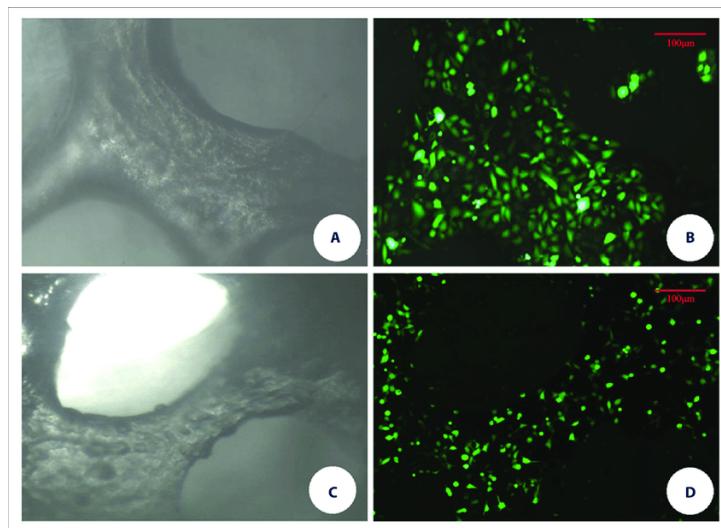


Figure 2.2: A comparison between light (left sub-figures) and fluorescence (right sub-figures) microscopy. The same specimen is represented using each microscopy type with the top pair (A,B) and the bottom pair (B,D) each representing different cell cultures [16].

2.2.2. Factors regarding fluorophore selection for fluorescence microscopy

As described in Section 2.2.1, the primary reason to use fluorescence microscopy is the ability to select structures or biomolecules that are relevant to the analysis through the use of specific fluorophores. When imaged the indicated objects appear with greater clarity and differentiation than in optical microscopy through the rejection of irrelevant fluorescence. These benefits are not acquired without a cost which is the added decision-making of selecting suitable fluorophores for the analysis being performed. There are a number of factors involved in this can be viewed as the types of biomolecules being stained (for our mitophagy imaging purposes it is the organelles) and the fluorophore properties.

Regarding the impact on fluorophore selection the largest is based on what biomolecules or structures are of interest to the analysis as the scope of fluorophores will be narrowed to those that suitably stain them after which the fluorophore properties are to be considered in the context of the imaging requirements. Imaging requirements can describe the status of the specimen whether alive or fixed; the duration of the imaging whether time-lapse is being performed; the desired signal strength of the captured fluorescence; and the number of fluorophore types employed such that the different types of structures of interest can be differentiated between.

The fluorophore properties that will be explored are the excitation spectrum, the emission spectrum, the photostability, the maturation time, and the emission brightness [17–19].

Excitation spectrum is the range of wavelengths that are able to excite the indicators to induce emission. The excitation spectrum of a fluorophore must be evaluated relative to the microscope filters for the illuminant and the excitation spectra of indicators for other structure types. When the excitation spectra of two or more different fluorophores overlap, a phenomenon called ‘cross-talk’ is induced [20]. This overlap can be complete or partial where the quantity of overlap can increase the severity of the cross-talk where cross-talk causes the simultaneous excitation of multiple fluorophores where only one fluorophore type was intended to be excited.

Emission spectrum is the range of fluorescence wavelengths that the fluorophore emits post-excitation. This spectrum must be evaluated relative to both the filter preceding the photon detector and the emission spectra of different fluorophores employed for other structure types. When these emission spectra experience some quantity of overlap a phenomenon called ‘bleed-through’ occurs [20]. Bleed-through where fluorescence captured by the microscope that is expected to be from a specific fluorophore is intruded by the fluorescence of unintended fluorophores interfering with the accurate differentiation of structures since structures can be incorrectly labelled as a structure due to this inappropriate emission spectra overlap. In fluorescence microscopy, it is standard to separate captured fluorescence wavelengths by channels thus grouping different structures under different channels when there is minimal bleed-through.

Photostability is the property of the fluorophore’s resistance to degradation due to photobleaching (will be discussed in Section 2.2.3) which is a phenomenon that limits the total quantity of excitation that indicators can undergo. A higher photostability must be considered when indicators must undergo extended excitation such as time-lapse imaging (to capture temporal behaviours) or when a longer exposure time is required for the captured fluorescence to compose a sufficient signal [19].

Maturation time is the time required before the fluorophore begins fluorescence. This is exclusive to fluorescent protein-type fluorophores and the maturation time begins once the fluorescent proteins have been introduced (transfected) to the specimen [21, p.286-289].

Emission brightness is the property describing the potential intensity of the fluorescence emission relative to the intensity at which it is excited [17, 19]. In this use, intensity refers to the number of photons either absorbed or emitted within some time window. This is under the assumption of optimal wavelengths as the effectiveness of the photon absorption is determined by the wavelength of the excitation photons based on the excitation spectrum and the number of photons of a specific wavelength will be relative to the emission spectrum of the fluorophore. Due to this it is more likely to induce greater excitation in a fluorophore

for certain wavelengths and the brightness of the emitted signal will be greater in certain wavelengths than others. A higher emission brightness property means that lower total excitation experienced by the fluorophore is needed to emit a strong fluorescence signal than a fluorophore with a lower emission brightness. This is beneficial for time-lapses or when the photostability is low as less intense excitation energy is needed to capture a sufficient signal reducing the photobleaching experienced.

I tried to describe this in a simple way without plagiarising the text and while avoiding the heavily mathematical definition. This was intended to remain high-level but I fear I may have still overcomplicated it.

2.2.3. Constraints on the performance of fluorescence microscopy

In comparison to light microscopy, fluorescence microscopy has many advantages in the visual labelling of structure types of interest and the removal of irrelevant structures which clutter the image but they are not without performance constraints. These constraints on the imaging performance are primarily due to the suitability of the fluorophores for the given analysis, the condition of said fluorophores and microscope resolving limits. Prominent constraints that will be discussed are bleed-through, photobleaching and diffraction. While both cross-talk and bleed-through were mentioned individually, in Section 2.2.2, they will be discussed together as the occurrence of both phenomena is required for the effect to be noticeable to the imaging quality.

Bleed-through and by extension **cross-talk**, are referred to either interchangeably [15] or individually [20] depending on the source of the description. What can be seen is that both phenomena must occur simultaneously for there to be some impact since the overlap in emission spectra means little if only the intended fluorophore is excited given that the excitation spectra are fully separated. The benefit of segregating the employed fluorophores by the respective emission wavelengths is so that the different types of structures or biomolecules (e.g. types of proteins, certain organelles, etc.) can be differentiated easily. For this to function, there is the expectation that each structure type that is to be differentiated between has a ‘unique’ emission wavelength, among the emission wavelengths used in that imaging, to identify it [20]. Bleed-through and cross-talk break down this emission wavelength uniqueness thus fluorescence expected to identify only structures of type A (e.g. mitochondria) will also capture structures of type B (e.g. lysosomes). This is avoided by minimising the overlap between the excitation spectra of different fluorophores and likewise for the emission spectra respectively since emission requires excitation and filters can be employed to reject small overlaps (typically occurring along the outer edges of the spectra) by filtering out the edges of the emission and excitation spectrum. This is also not without repercussion as some of the excitation

and emission energy will be at these filtered regions reducing the total captured signal strength.

The above paragraph is where I noticed the space after the paragraph heading. There used to be a comma after “bleed-through” but it was not flush with the word so I removed it.

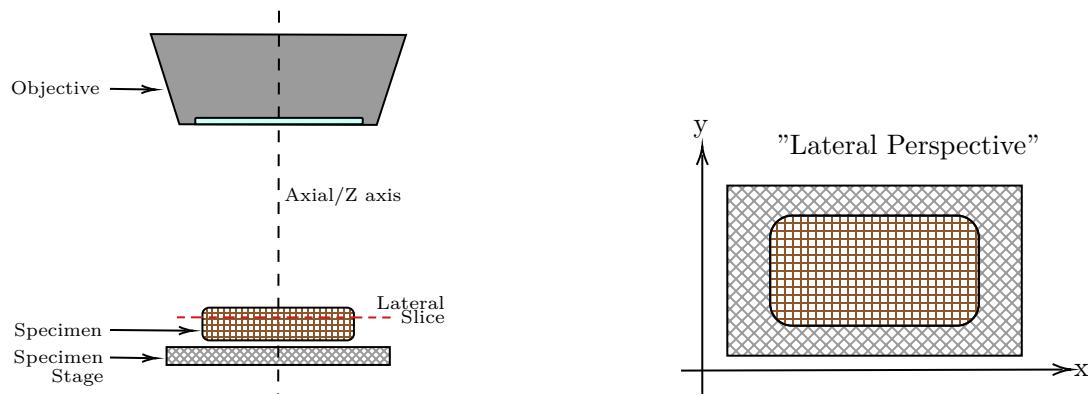
Photobleaching is the degradation of the fluorophore’s ability to fluorescence due to exposure to light [22, 23]. Despite this, it is the exposure to excitation-inducing light that contributes the majority to this degradation [22] with the exposure measured by the number of absorbed photons in this context. Since the excitation light intensity exposes the fluorophore to more photons within a given time frame while exposure duration extends the length of said time frame. Photostability is the property measuring the resistance of the fluorophore to photobleaching while the emission brightness defines the number of photons emitted by the fluorophore relative to the number of absorbed photons thus high emission brightness means lower intensity excitation needs to be applied to capture sufficient signal thus reducing photobleaching indirectly [15].

Diffraction is a phenomenon experienced by photons due to their intrinsic particle-wave nature. This diffraction occurs when the photons travel through a small opening, such as a microscope aperture, and the photons are dispersed in a radial pattern originating from said opening. This radial pattern can also be visualised as a wavefront with amplitude oscillations originating from the opening with the pattern expanding outwards [24]. Under the assumption that diffraction does not occur then each point of fluorescence emitted by the specimen, within the microscope resolution, is captured by the microscope with a corroborating fluorescence point across the image pixels. With diffraction, each of those points will have their intensity distributed radially around them distorting the spatial representation of the specimen in the image. This radial distribution oscillates in intensity as it radiates outward due to the amplitude oscillation of the wavefront which appears as concentric rings and can be referred to as the airy disk [25]. There is also some decay in the intensity of the rings successively but this reduction is non-linear and a few bright rings or none at all will be visible depending on the diffraction pattern. The diffraction pattern itself regarding both the size and the distribution in intensity across the successive rings is dependent on a number of properties but namely the size of the numerical aperture (NA) and the emission wavelength value [24].

As stated prior, this diffraction distorts the spatial representation of the captured fluorescence by scattering the fluorescence signal across diffraction regions. This distortion can be reduced by reconstructing the image without diffraction through deconvolution (described in detail in Section 2.4.3) but the diffraction present in the image has to be understood for it to be reversed. This can be estimated through investigation of the

image diffraction but this involves human bias leading to inconsistency in the results or a mathematical representation called a point spread function (PSF) can be generated using parameters specific to the imaging. The problem is that irrespective of the diffraction being correctly modelled is that the overlapping of diffraction regions beyond a certain amount makes reconstruction impossible for those regions. When these diffraction regions overlap the airy disks compound on each other resulting in a region that fits no diffraction pattern and cannot be reversed. This adds to the resolution limitation of microscopes as the diffraction pattern overlap from the airy disk centre to the first ring from the centre is roughly when effective reconstruction becomes impossible for those regions [24, 25].

2.2.4. Confocal Microscopy



(a) Axial axis relative to Objective, Specimen and Specimen Stage. The Lateral Slice is annotated by a dashed red line

(b) The lateral view of the specimen as seen in the axial direction from the Objective.

Figure 2.3: Two diagrams illustrating the two perspectives with the appropriate axis. a) depicts the specimen from the side to display the hypothetical position of the lateral slice (focal plane) relative to the microscope objective with the z -axis, or axial direction, annotated perpendicular to the specimen stage. b) depicts the specimen as seen from the objective and will be referred to as the "Lateral Perspective" which views the x, y -axis of the specimen.

Microscopy thus far The microscopy that has been discussed thus far (Section 2.2.1) has involved the illumination of the entire specimen and the subsequent capture of all fluorescence emissions, or reflections in the case of optical, limiting the clarity of the captured image by the specimen thickness and is called wide-field microscopy. Before exploring why specimen thickness is a detriment to the image quality the focal plane will be defined in the context of this section to simplify further explanations. The focal plane is a lateral plane perpendicular to the objective lens and light originating from along this plane is in focus thus providing a sharper image with greater clarity. An example of the focal plane can be seen in Figure 2.3a and is referred to as "Lateral Slice" and light originating from above or below this plane is referred to as out-of-focus light [15]. While

wide-field fluorescence microscopy (often Epifluorescence microscopy) can discriminate between fluorescence based on wavelength through the use of filters, separating fluorescence representations of the specimen into groups, it cannot spatially discriminate between light originating from inside or outside of the focal plane. This results in the out-of-focus fluorescence being layered atop the in-focus fluorescence causing blurring and the loss of resolution of details (the pixel resolution is unchanged) [15, 26]. Thicker specimens have larger volumes outside of the focal plane that will contribute more out-of-focus fluorescence thus limiting the image quality by the specimen thickness. An indirect repercussion of this lack of spatial discrimination of fluorescence by focus is that 3D images cannot be captured at high quality since an accurate volumetric representation of the specimen only functions under the assumption that the position, size and shape of specimen objects are accurately reflected in the image which out-of-focus fluorescence prevents.

Confocal microscopy A solution to this is the use of confocal microscopy which is a microscopy architecture which employs a pinhole aperture before the photon detector to reject out-of-focus fluorescence, a comparison between wide-field and confocal imaging of a specimen is shown in Figure 2.4. This rejection using the pinhole aperture works due to fluorescence emitted from different depths through the specimen approaching the pinhole at different angles after refracting through the objective and microscope lenses. This allows the imaging of cross-sections through thicker samples since the problematic out-of-focus fluorescence is rejected, or at least reduced. The implementation of a pinhole is not the sole addition to confocal microscopy, compared to wide-field, as the excitation and capturing of fluorescence originating from a singular point along said the focal plane is the second characteristic. The reason behind this asymmetry in focus across the focal plane is due to the pinhole rejecting fluorescence originating from a specific position and while depth is the major contributor to this so too is lateral position relative to the point of focus of the objective [26, 27]. When the pinhole rejections fluorescence it essentially rejects photons from the detector which are used to construct the image and insufficient photons lead to a weak image signal despite there being less noise from out-of-focus fluorescence leading regions closer to the focal point along the plane to diminish in intensity inversely to their distance from said point.

Using ‘Confocal Laser Scanning Microscopy’ (CLSM) as an example a scanning mirror is used at the objective lens to scan the focal point along the focal plane, maximising the acceptance through the pinhole of fluorescence from that point, while also focusing a high-intensity excitation laser to that same point [26]. The high-intensity laser compensates for the loss of signal intensity due to pinhole rejection while the scanning of the focal plane point-by-point constructs a final image with the focus across the image being relatively uniform. This is only one such architecture with ‘Spinning Disk Confocal Microscopy’ (SDCM) being another such confocal architecture employing multiple pinholes

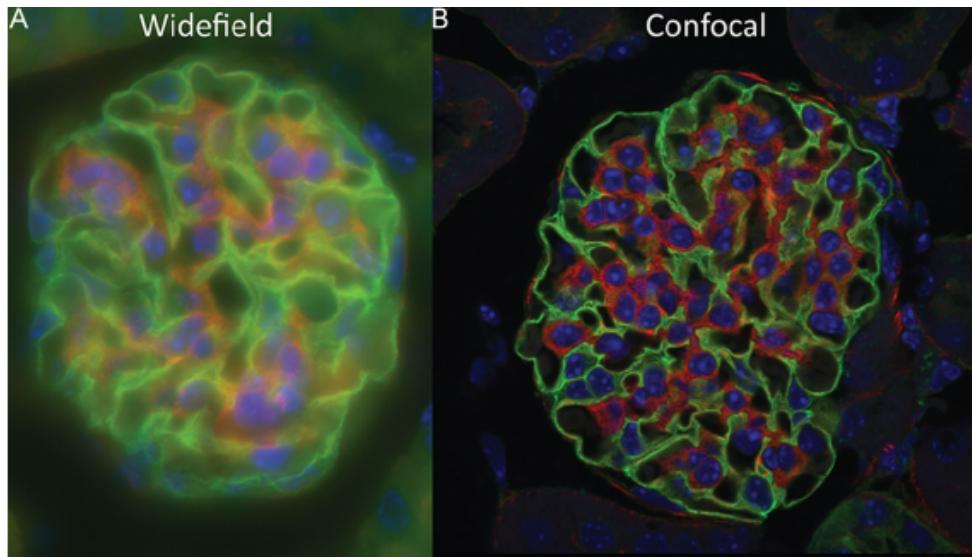


Figure 2.4: A 2D image of a specimen retrieved using wide-field (left) or confocal (right) microscopy [28]

and scanning points, instead of the one in CLSM, but this will not be explored further in this text [15, 26, 27]. With the ability to reject out-of-focus fluorescence comes the opportunity to construct 3D images of the specimen which is performed by adjusting the depth of the focal plane and imaging individual lateral slices through the image while maintaining that only objects located at that depth in the specimen are captured in the corresponding slice of the 3D image. One point of consideration is that higher excitation laser intensities and the duration the laser dwells on a point can render a stronger image signal but the trade-off is that photobleaching is worsened and while the excitation is focused on a single point this laser penetrates through the depth of the specimen exciting out-of-focus points introducing photobleaching without capturing that fluorescence. This penetration through the depth of the specimen by the laser is the limiting factor on the volumetric resolution of the image along the depth as each lateral slice will require an additional excitation to pass along the focal plane increasing photobleaching throughout the image. The size of the pinhole determines the strictness of the out-of-focus fluorescence rejection but it also determines the loss of signal intensity from the emitted fluorescence. Increasing the excitation laser intensity or extending the duration the excitation laser dwells on each point can increase signal intensity but worsens photobleaching thus a compromise must be made when imaging in 3D.

I rewrote this description of Confocal microscopy in the hope that it flowed better and used simpler language than before. I have three sources explaining Confocal microscopy but how do I avoid citing every line? Can I cite at the end of a paragraph or can I cite the key point while assuming the reader will realise the point built upon that subsequently relied on the same source?

2.2.5. Wide-field and confocal microscope apparatus

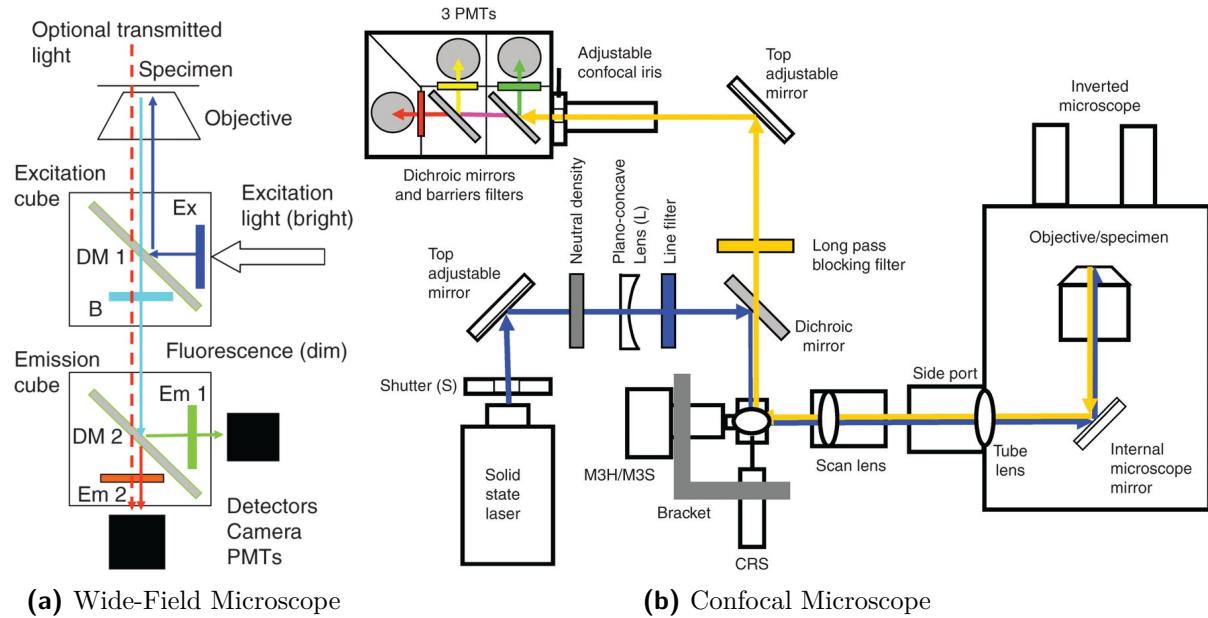


Figure 2.5: Diagrams representing the Wide-Field Microscope and the Confocal Microscope [15]

Both wide-field and confocal microscopy are applications of fluorescence microscopy albeit with different microscope designs. The design aspects of both microscopes, seen in Fig. 2.5, are the use of filters for selecting both the excitation and emission wavelengths, the microscope objective lens and the excitation illuminant. The key components that differ are that wide-field frequently implements cube filters which both direct the excitation light to the specimen through the objective and the emission photons to the photon detectors. Filters within the cube filters select the relevant wavelengths when redirecting the light to the specimen and the detector. Conversely, confocal microscopes implement multiple adjustable, and fixed, mirrors in combination with filters. These mirrors allow the excitation laser to be directed to specific regions of the specimen and the resultant emission fluorescence reflects off of these mirrors on approach to the photon detector. The adjustment of the mirrors determines the angle at which the emission fluorescence approaches the photon detector based on the point of emission relative to the mirror angles. Confocal microscopy also uses a pinhole aperture before the photon detector which rejects photons depending on the angle at which they approach the pinhole. It is the combination of the adjustable mirrors manipulating the emission fluorescence angles with the pinhole that allows the reduction of emission fluorescence outside of the position of focus from reaching the photon detector. This position of focus can also be adjusted in depth using microscope apparatus allowing separated cross-sections through the specimen to be captured along the specimen depth. In the confocal microscope, the excitation illuminant is a precise laser as opposed to the wide-field illuminant which radiates across

the whole specimen [15].

This is one of the few subsections where I performed no revisions besides passing through it to check coherency. If you feel this section is unnecessary please let me know.

2.2.6. Summary

Many aspects of fluorescence microscopy have been discussed ranging from some of the factors under which fluorophores (indicators) are selected to the decisions between the microscopy method applied. The key points covered are that different fluorophores bind to specific environments or subjects based on their chemical or biological composition; the selection of these fluorophores must consider their excitation and emission spectra to avoid overlap; confocal microscopy enables more detailed imaging with less out-of-focus fluorescence with the added option of three-dimensional imaging. I initially wrote this summary a year back and haven't done so for other sections. Do you think it would be beneficial to provide a short summary of a section at the end of a section or should I scrap the summaries as a whole? If I am going to keep summaries I feel I need to do it for every section out of consistency.

2.3. Colocalization

With the specimen imaged using fluorescence microscopy, as described in Section 2.2, mitophagy occurrence is to be measured. Through fluorescence microscopy, each of the organelle types has been separated by the selected fluorophore excitation and emission spectra such that a different colour channel can be assigned to each. Since mitophagy involves the engulfment of mitochondria by autophagosomes which fuse with lysosomes (becoming autolysosomes) it can be assumed that the overlap of mitochondria and autophagosomes (or lysosomes) would be a sufficient indicator of mitophagy. While this holds in a physiological sense, this is not guaranteed for the overlap of structures between two channels, one of which is the mitochondria channel, as that would imply that what was imaged is a perfect representation of the specimen organelles to the highest level of precision. While newer microscopy architectures and methods have increased in imaging resolution to older iterations the required resolution and precision increase as there are attempts to image smaller structures and objects with increased fidelity and detail. To compensate for this shortcoming in the image representation of the specimen a reliable framework is required to evaluate the overlap present by several factors and calculate the certainty that the perceived cross-channel overlap can be classified as colocalization [29].

In "A practical guide to evaluating colocalization in biological microscopy" [29] there is specific mention of two components which derive colocalization and they are co-occurrence and correlation. Co-occurrence is identical to the aforementioned overlap where the

fluorescence from each channel occupies the same position in space. Correlation measures the intensities of the fluorescence between each channel relative to the distribution of these intensities over space. In combination, this results in colocalization being evaluated by not just the spatial co-occurrence of fluorescence channels in the image but also the proportion of fluorescence intensities between each channel at these regions of co-occurrence.

While co-occurrence may be sufficient for visual evaluation of possible colocalization events it is an unreliable indicator of colocalization within the specimen as a qualitative measure of correlation is unintuitive and difficult. For this reason, there exist a number of algorithms which measure colocalization within the image region or a selected sub-region as true colocalization cannot be determined due to imaging limitations and the number of spatial characteristics describing colocalization being difficult to fully capture. Due to this many of these algorithms provide pros and cons based on the application. These pros and cons can range from intermediate values (between 1 and 0) being uncertain in interpretation, being sensitive to noise, or time-expensive iterations to converge to a reliable result [20]. The standard colocalization algorithms are Pearson's correlation coefficient, Manders' overlap coefficient, Manders' correlation coefficient, Costes' approach, Van Steensel's approach, and Li's approach [20, 29].

These methods provide a numeric evaluation of the colocalization present in the analyzed region but due to the constraints present within these methods, they are typically applied to selected sub-regions as opposed to the full specimen image. When these regions contain multiple colocalization events these methods cannot distribute the calculated colocalization score between the events thus precise region is of greater importance to evaluate the colocalization measure of each region of co-occurrence. A recently developed method is that of Regression adjusted colocalisation colour mapping (RACC) [30] which calculates the colocalization throughout the region (which can be the entire image) and outputs the specimen image (with both images overlaid) with a colocalization channel overlaid. This colocalization channel presents values proportional to the colocalization strength which is zero for regions without overlap but across regions of co-occurrence the provided value allows the colocalization strength to be spatially and visually mapped to these individual event locations within the specimen.

2.4. Image preprocessing

With the specimen images acquired through microscopy, some of which are detailed in Section 2.2, the analysis of the features described by the image would ideally commence. As was introduced in Section 1.1.3, it is not uncommon for the specimen representation captured in the image to deviate from the actual specimen due to the presence of noise, distortion or artefacts. In the context of noise, the quality of the image can be approximated as the similarity between the actual specimen and the structures of interest that may be

labelled and what is described in the captured image.

This actual specimen representation refers to the ideal visualisation of the specimen where stained structures and biomolecules of interest to the analysis are accurately differentiated as well as the size, shape and position of these captured in the image precisely reflect the matching objects in the real specimen. When there are components in the image that are not reflexive of the image then the image diverges from the specimen and the image degrades thus elements of the image causing this are referred to as degradation factors.

2.4.1. Sources of interference in image quality

As mentioned prior, the correction of the image quality degradation is necessary for the accurate analysis of these microscopy images. The factors contributing to the degradation have been listed as distortion, noise and artefacts although there is a tentative degree of overlap between these. For clarity, the definition of distortion, noise and artefacts in the context of this research will be presented before future use.

Distortion is the spatial deformation of the image signal such that the shape of the image signal and described features differ from the real subject. Noise can be visually described as perturbations of the captured structures in the image that reduces their clarity without presenting distinct structures (where distinct structures have shape and size similar to the structures of interest). Since noise is typically thermal or electrical in origin (from the microscope) the noise factor is applied to the image signal when captured and is applied across the entire image with the severity of this noise described as the ratio between the image signal power and the power of the noise termed as the signal-to-noise ratio (SNR). Artefacts are the last component and describe any aberration in the captured image signal such that distortion and noise could be categorized as types of artefacts. For this research artefacts will refer to visual aberrations in the image signal that are irregular and present with sizes and shapes closer to the structures of interest but typically are lacking in overall signal intensity at that focal plane. The severity of these artefacts is dependent on the similarity between them and the structures of interest and how close the average intensity of an artefact is to the average structure intensity. These artefacts originate from fluorescence that was not intended for capture (due to the occurrence of bleed-through); the fluorophore could be binding to biomolecules outside of expectation (e.g. binds to specific proteins but unexpectedly these proteins are found in both the cytoplasm and the organelle with the former being unexpected); or out-of-focus fluorescence from a currently out of depth structure has not been sufficiently reduced appearing as an artefact. This sounds similar to that of noise but rather these artefacts can originate from specimen elements that are not intended for capture yet are described, even weakly, within the image.

With these research-specific definitions presented an overview of some of the origins of these image quality degradation factors is described below:

- **Microscope Noise:** This takes the form of Poisson noise and Gaussian noise contributed to the image signal through the photon detector by thermal and electrical noise originating from certain microscope components.
- **Photon Noise:** Due to the discrete particle nature of photons the quantity of photons emitted by excited fluorophores is not consistent. When this inconsistency in emission occurs there is in turn a variation quantity of photons detected by the microscope over a period of time. This culminates in the variance of the image signal strength, particularly for short imaging periods where a baseline expected photon count cannot be reliably determined to mitigate this variation. This variance visually appears as noise which is characterised in the image signal by a Poisson Distribution [31, 32].
- **Microscope Diffraction:** As described in Section 2.2.3, the fluorescence emitted by the specimen passes through an objective lens and a pinhole aperture both of which cause the photon signal, when viewed as a wave, to diffract. This diffraction induces a distortion of the image signal at the detector and can be characterized as a blurring radiating from each pixel. The severity of this distortion is relative to both the image signal conditions (e.g. emission wavelength) and the microscope apparatus (e.g. numerical aperture).
- **Background fluorescence:** As described in Section 2.2.3, this is a broad category of fluorescence which has been unexpectedly captured and does not necessarily originate from the selected fluorophores in focus. This can contain auto-fluorescence, out-of-focus fluorescence, and bleed-through among many other possible environmental fluorescence.

2.4.2. Standard pipeline for image pre-processing

Specimens are imaged using microscopy to analyse and measure properties, activities, and behaviours related to the specimen. These specimens are imaged such that the visual features describing phenomena and characteristics can be measured for the appropriate analysis. As detailed in Section 2.4.1, the reliability of these measurements is proportional to the image quality which is degraded by several factors. To restore image quality several methods can be applied which reduce the presence of these degradation factors. This phase of the image preparation before the analysis, but after imaging, during which these methods are applied is referred to as the pre-processing phase (introduced in Section 1.1.3) where a number of these methods are applied sequentially. This sequential application

of methods and techniques for image quality restoration is the pre-processing pipeline (introduced in Section ??) through which the output of each method feeds into the input of the next allowing the degradation reduction to compound along the pipeline.

There are numerous different techniques which can be applied each of which addresses specific degradation factors and have conditions under which they perform optimally. It is for this reason that sequential application of methods is ideal as these techniques address different degradation factors but in the process of reducing these factors the image signal can get diminished if too aggressive (e.g. blurring structure edges during noise removal) or the performance of said techniques are sensitive to the presence of other factors. By sequentially applying these techniques preceding techniques can reduce factors while improving the conditions for proceeding techniques while proceeding techniques can compensate for the limitations of the prior techniques. For this reason, it is important to plan the techniques used and the aggressiveness at which each technique is applied to balance between image signal preservation and noise removal. For colocalization analysis (overview in Section 2.3) in the context of the mitophagy research use case a high-level abstract pre-processing pipeline, based on the pre-processing employed by [10], is presented:

1. The PSF for the image is generated based on the image metadata and the microscope used. This was mentioned in Section 2.2.3 and will be discussed in further detail in Section 2.4.3.
2. Deconvolution is performed on the image utilizing the appropriate PSF. This will mitigate the distortion which occurs at the objective lens. The reason for deconvolution being applied so early will be discussed in Section 2.4.3.
3. Filters and image enhancement is employed to maximize the image signal strength while diminishing, or removing, noise. Frequently employed noise removal can diminish signal across the image although more so for noise thus image enhancement restores the strength of the image signal.
4. Foreground and background segmentation. The removal of artefacts induced by background signals and uneven illumination that is both irregular and prominent for standard denoising to effectively mitigate.

Further image segmentation techniques or analysis techniques would be employed after this to acquire the measurements from which to draw observations. For the mitophagy use case, these observations will typically relate to the conditions and behaviours of the imaged organelles concerning the mitophagy process. This is not definitive of all pre-processing pipelines but provides an example of the sequential order in which these methods are applied for effective image preparation.

2.4.3. Overview of the methods to restore image quality

As discussed in Section 2.4.2 prior, several techniques are applied sequentially to the image to restore the image quality. These techniques will be categorized (for this research) by the effect they have on the image. The groups will be deconvolution, denoising, image enhancement and thresholding.

Deconvolution

During the imaging process, there is some degree of expected blurring of the image due to diffraction. This was described in detail in both Sections 2.2.3 and 2.4.1 where diffraction is the phenomenon where light scatters after passing through an aperture leading to the fluorescence spreading from singular points. This spread is more visible when it originates from brighter points thus structures with strong image signals present more prominent spreads which appear as blurry regions surrounding the original structure. This is problematic as this distorts the actual spatial representation of the imaged objects as they may appear larger than they are and the overlapping of these blurred regions can obfuscate edges and fine details thereby reducing the image quality. A technique known as deconvolution can be applied to the image after imaging which works to reverse the diffraction phenomenon but this requires the diffraction pattern for the image to be known and it is for this reason that overlap between diffraction blurs can interfere with effective reconstruction.

The diffraction pattern can be determined in several ways but a convenient and effective method is to determine the theoretical diffraction pattern by calculating the PSF. The PSF is a mathematical model of the diffraction pattern that while lacking the nuance of the actual pattern is similar enough for effective deconvolution. The PSF calculation considers several imaging and fluorophore properties such as the relationship between the image pixel size to real distance in nm, the NA, the emission wavelength, the refractive index, the distance between lateral sections in 3D images and the microscope optical model among other properties [33] with an example of a generated PSF shown in Figure 2.6.

In the theoretical pipeline described in Section 2.4.2 the first step is to determine the PSF and the second is to apply deconvolution after which further processing can be applied. To justify this order first a simple algebraic representation of the captured image in relation to the blurring is shown [34, 35]:

$$G(x, y) = F(x, y) * S + \eta(x, y) \quad (2.1)$$

Where G is the captured distorted image, F is the original undistorted image signal, S is the representation of the diffraction and the additive white noise is represented by η . The additive white noise is usually Poisson/Gaussian and originates from electrical or thermal

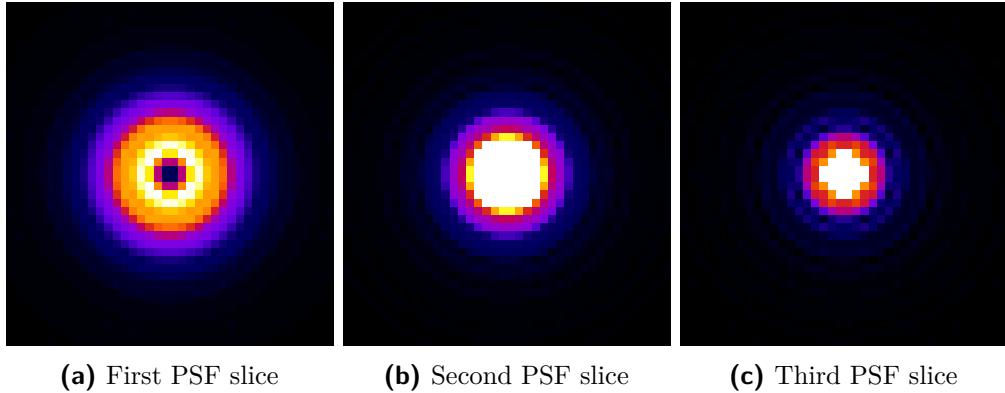


Figure 2.6: PSF example for a 6 slice 3D image. The PSF is mirrored between the first three and last three slices thus only the first three are shown and the order is inverted for the next three. The slices (a)-(c) are ordered by depth where the later in the order the deeper the slice.

energies in the microscope detector and apparatus. From 2.1, the culprit of the distortion can be seen as S which is also the PSF and the original image F is convolved with the PSF to result in the distorted captured image. To reconstruct the original image F a deconvolution operation needs to be applied with the appropriate PSF to ‘reverse’ the diffraction effect of the microscope pinhole.

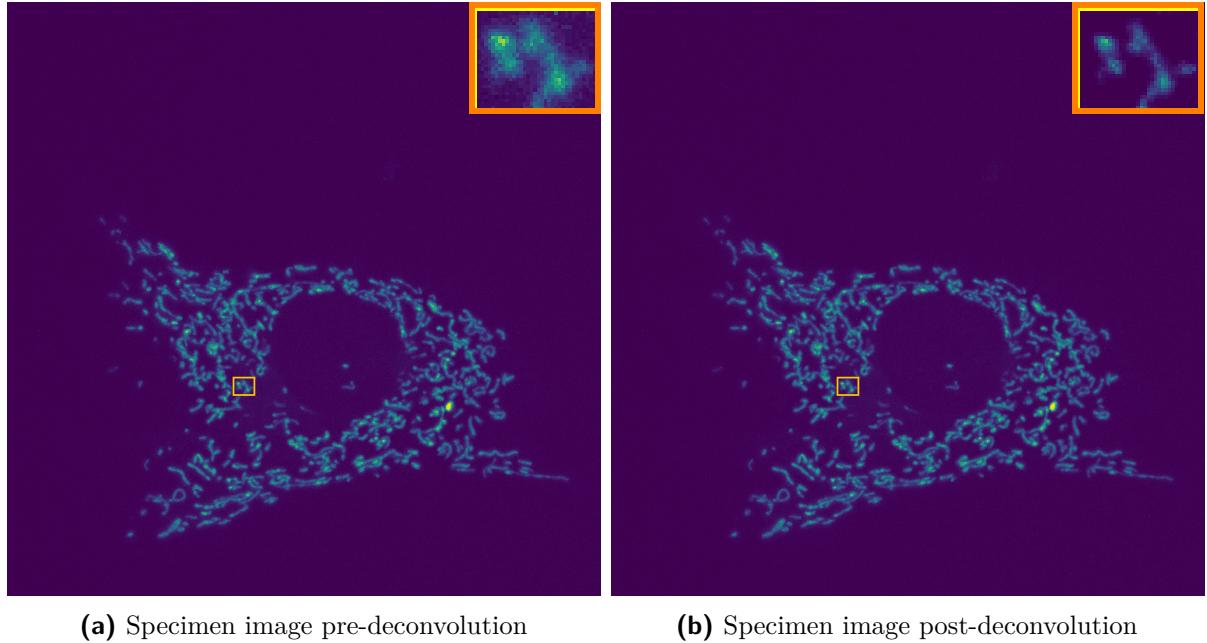
$$F(x, y) = G(x, y) * H(S^*) \quad (2.2)$$

Where $H()$ is the deconvolution function and S^* is the determined PSF for the image. Depending on the method by which the PSF is calculated, and when, the selection of what algorithms to apply to the image before deconvolution is significant. Operations that change the image signal, particularly in a non-linear or adaptive local approach such as Gaussian blurring, can lead to changes in the dispersion of the image signals within the image and thereby change the properties of the blur. If the PSF is generated so that these adjustments can be compensated for then this may be of little concern but if the PSF cannot do so then applying other algorithms before deconvolution could introduce unexpected outcomes and risks.

Regarding the additive white noise, it is not uncommon for deconvolution algorithms to also remove noise [10] and since the additive white noise is theoretically independent of the image signal distortion then denoising algorithms targeting solely the additive white noise could be applied with minimal issue.

There are multiple deconvolution algorithms which can be used such as Richardson-Lucy, Richardson-Lucy Total Variation, Tikhonov Regularization Inverse Filter, Tikhonov-Miller and Regularized Inverse Filter to name a few [35]. The two primary groups of deconvolution algorithms are iterative methods, such as Richardson-Lucy, and inverse methods, such as Tikhonov Regularization Inverse Filter. In Figure 2.7 the effects of deconvolution are shown with a comparison of before and after the application. The pre-deconvolution image

(Fig. 2.7a) shows how the diffraction induces a blurred region distorting the perceived volume and shape of structures in the image as was discussed prior.



(a) Specimen image pre-deconvolution **(b)** Specimen image post-deconvolution

Figure 2.7: A maximum intensity projection of a specimen image is shown to compare a) pre-deconvolution and b) post-deconvolution. A region of interest that is zoomed-in is shown in the top right corner of each respective figure.

Denoising

Denoising methods are transformations applied to the image to mitigate the presence of noise within the image. As detailed in Section 2.4.1, regarding the degradation factors definitions, noise is generally expressed in the image as a global presence throughout the image which reduces the clarity of the features but typically does not generate new distinctive features. To clarify this, noise induces fluctuations in the amplitude of the image signal centred around the original signal and while these fluctuations are randomly distributed throughout the image they still follow certain ‘rules’. The fluctuations of individual pixels occur independently of other pixels yet the sign (positive or negative) and amplitude of all noise-based fluctuations across the image are drawn from the same distribution [36]. There can be multiple noise fluctuations, with different origins, that compound in the resultant effect on the image pixels but the distributions from which these are drawn can be identical or different between different noise distributions. In an ideal situation, it may be possible to estimate the distribution of the noise from the image histogram but the compounding of other degradation factors on the image signal result in this being quite difficult for real image signals.

Due to this methods that can mitigate the impact of the noise independent of knowing the noise distribution are necessary. These techniques are composed prominently of filter

methods which are applied to pixel neighbourhoods and apply a filter operation. The filter operations can be standard mathematical operations, such as the median or mean of the neighbourhood, where the calculated value is applied to the pixel the neighbourhood is centred on. Irrespective of the exact filter method applied a consistent characteristic of denoising outputs is the smoothing of the image as can be seen in Figure 2.8 where the Sigma filter [37]. As mentioned prior, many of these methods apply an operation to a pixel which adjusts the value based on the surrounding pixels thus reducing the variance within these neighbourhoods.

While the discussion has been constrained to finite neighbourhoods there are also techniques which consider all pixels for the operation but the contribution of each pixel is relative to the distance from the pixel to be adjusted. This reduction in variance in image regions is fine for those already expressing a low variance but image features like edges are expressed by the relatively high variance localized around the edge. It is for this reason that excessive smoothing through denoising is detrimental for features expressed by high local variances limiting the quantity of denoising that can be applied [37]. Likewise, even image regions that express limited variance may describe a feature yet excessive smoothing can eventually remove the boundary between the background and the objects to be analyzed. For this reason, denoising does have an upper limit on its application of it based on the noise level being too strong or the image signal being too weak. There do exist denoising methods optimized for edge preservation such as the Sigma filter [37] but these, among the numerous other denoising methods, will not be explored in further detail for this work.

Image Enhancement

While denoising methods are optimal for removing consistently present random fluctuations in pixel intensity across the image they are not designed to improve image signal strength or correct uneven signal distributions throughout the image which do not describe image features such as object edges and the region of pixels composing the body of the object. Object bodies are typically composed of pixels within a bounded range (excluding outliers) with the lower bound being of a magnitude greater than the background with the brighter pixels closer to the object centre or clustered around it. What is important is that image enhancement is easier to characterise by two components: firstly, the goal of it is to amplify the presence of both objects in the image and features describing said objects, like edges, and, secondly, while all image enhancement methods aim to enhance the image how this is done less consistent than other method types as it improved clarity and differentiation of objects from the background.

Before exploring the behaviour of image enhancement techniques the clarity and differentiation regarding image objects will be detailed for this context. Image objects have a body, described prior, surrounded by an edge which is in contrast to the pixels

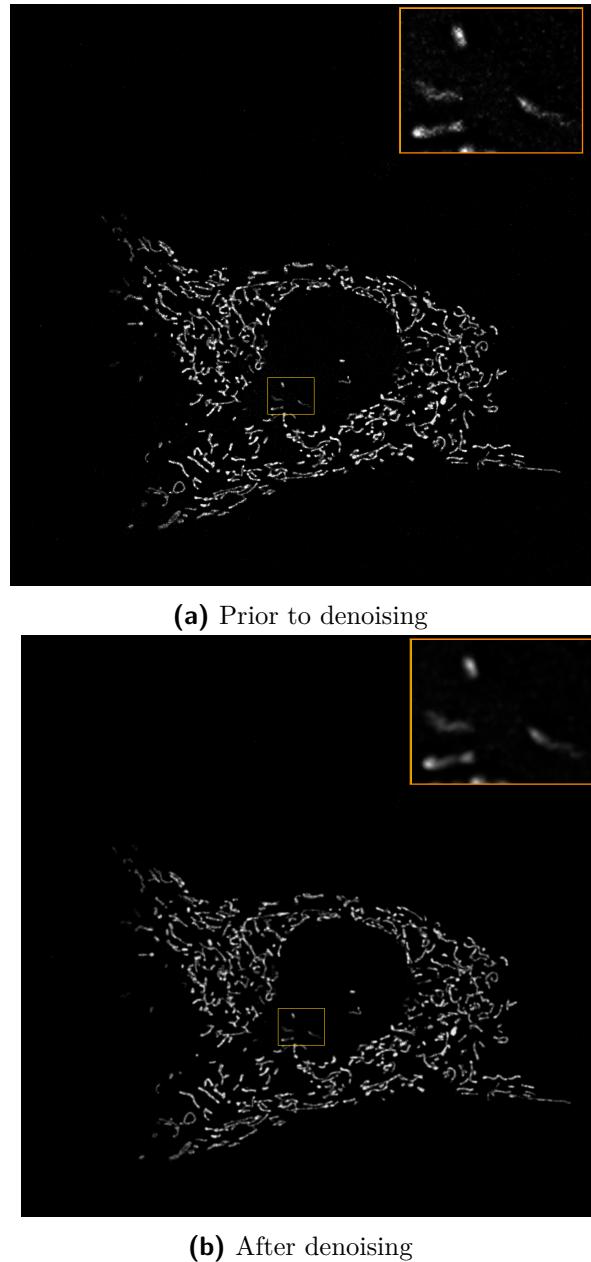


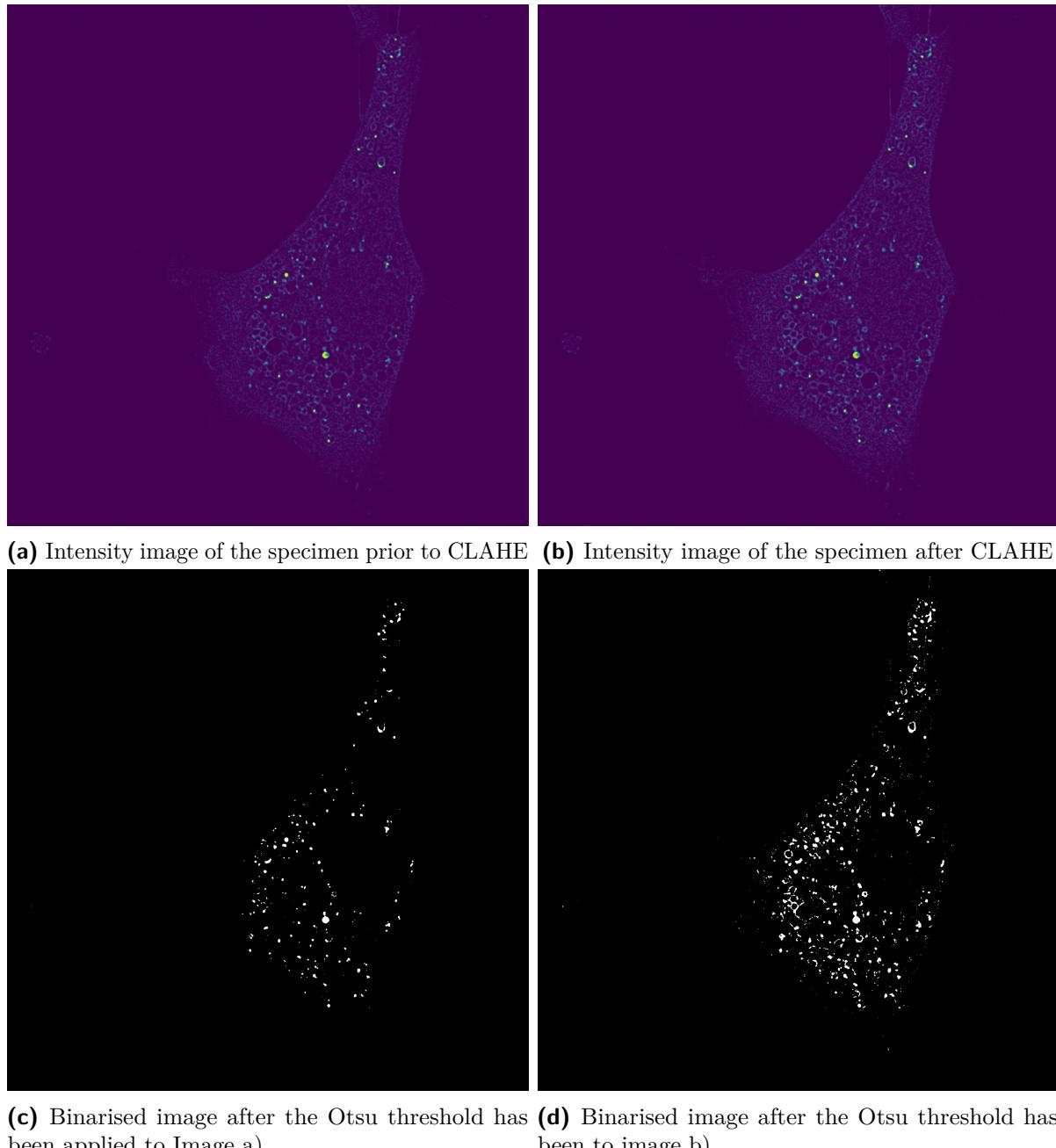
Figure 2.8: The effect of filtering on a mitochondria sample with before (a) and after (b). The sample images are maximum-intensity projects flattened from 3D with a region of interest bordered with an orange frame with a zoom-in at the top-right

surrounding it whether those belong to some other object or the background. The contrast of the object edges is crucial to determine the boundaries of objects particularly when there are overlapping objects as insufficient contrast may lead to objects being incorrectly joined. The object body contrast does not need to be as large as the edges but must still contrast with the background as background regions in the object body could be cavities. For this reason, object differentiation is dependent on the size of the gap between the brightness of pixels composing the object and those composing the background while, in this context, the clarity is dependent on the brightness resolution of the image which can be viewed as the minimum and maximum pixel values of the image excluding outliers (outer values with relatively few pixels attributed to them) as a lower brightness resolution provides a more limited range within which to separate the object and background pixels. This also requires the background to be consistent as it is supposed to be the lowest brightness level in the image as it describes pixels where no fluorescence was localised in the actual specimen. These methods can range from contrast stretching to correcting uneven illumination. Contrast stretching can target either the whole image or localised regions within which to enhance the contrast present which is beneficial for contrast-dependent features such as detecting the edges of objects. Uneven illumination across the image is characterized by changes in the brightness of the pixels across the image that do not describe any objects as object edges or bodies. This change can appear to be homogeneous relative to the object sizes and interfere with the contrast required for differentiation. Due to this, methods correcting uneven illumination attempt to normalize the brightness of regions based on how low the variance within that region is i.e. brightness is lowered for homogeneous regions with low contrast while high contrast regions remain. I struggled a bit with describing the correction of uneven illumination. While I feel my idea of it is correct the explanation feels a bit difficult. If you have any advice for simplifying it or what should be corrected in the explanation would be appreciated. An example of correcting uneven illumination across the background is the rolling ball algorithm [38] while an example of contrast enhancement is seen in “Contrast limited adaptive histogram equalization”(CLAHE) [39].

For the background subtraction, I have been struggling to find images to which I can apply background subtraction myself (that actually have uneven illumination). Do you think it is fine if I use this image from scikit-image? https://scikit-image.org/docs/stable/_images/sphx_glr_plot_rolling_ball_001.png

Thresholding

Thresholding is a method by which pixels are assigned a label dependent on the pixel value in relation to a threshold value (lesser or greater than). In a pre-processing context, this is typically applied to differentiate between foreground and background pixels. This functions under the assumption that objects of interest in the image will be composed of



(a) Intensity image of the specimen prior to CLAHE (b) Intensity image of the specimen after CLAHE
(c) Binarised image after the Otsu threshold has been applied to Image a) (d) Binarised image after the Otsu threshold has been applied to image b)

Figure 2.9: The impact of contrast enhancement on the image analysis process is presented regarding the automated thresholding between foreground and background structures. The specimen image is shown in (a) which has had background subtraction applied and CLAHE has been applied to this image to restore contrast as shown in (b). The impact of this on global thresholding (Otsu) is shown in (c) which is Otsu applied before CLAHE and (d) is the result of Otsu after CLAHE has been applied.

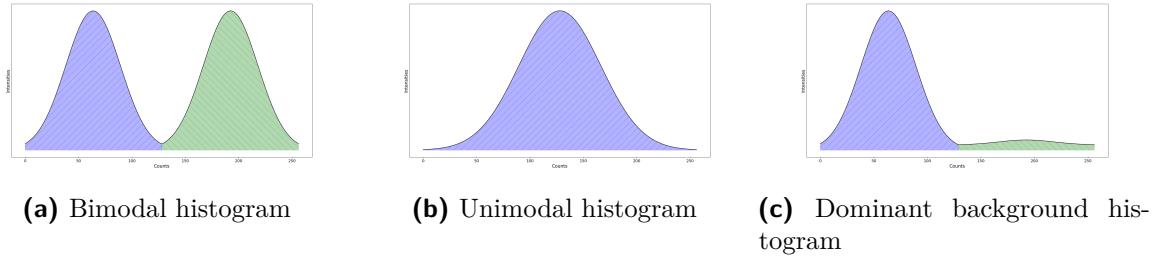


Figure 2.10: Three common histogram modalities with each mode being visually labelled with different colours and hatching (blue vs green and hatching direction).

pixels with a value greater than the threshold thus all objects of interest will be assigned to the foreground. Background pixels on the other hand are artefacts, noise or objects not of interest (appearing due to autofluorescence, out-of-focus fluorescence, or bleed-through). After threshold application, only the foreground pixels, those belonging to objects of interest, will be preserved.

Outside of pre-processing threshold methods technically fall under image segmentation which is a body of techniques that assign labels to pixels based on some criteria. Image segmentation has a myriad of methods depending on the field of application (microscopy, morphological analysis, object detection, etc.) where the number of labels is determined by the number of groups that an object, described by pixels, can belong to. This application is relative to pre-processing thus the focus is on thresholding images to get binary classifications (foreground or background).

With the binary classification of images to separate pixels composing objects of interest from ‘other’ pixels (where ‘other’ refers to noise, artefacts and inconsistent background illumination) a single threshold value is required to classify each pixel. This classifying behaviour is consistent across all threshold methods but the determination of the threshold value is what changes between methods. There are two groups of threshold methods which are global thresholding and local thresholding [4] with the former determining a threshold using and applying it across the entire image while the latter determines and applies the thresholds across localised regions of the image.

Global thresholding Global methods, like *Otsu thresholding* [40], determine the threshold value using the image histogram by selecting an intensity value separating the histogram values into two groups. In ideal situations with a distinct separation between the grey-level intensities of the background and foreground will be characterized in the histogram with two peaks where the clusters centred around each peak correspond to the pixels for the foreground (higher intensity) and background (lower intensity) respectively [4, p.163]. An example of this can be seen in Figure 2.10 where the different potential histogram shapes are displayed. When there are distinctly two peaking crests with limited overlaps of their bodies in the trough (as indicated in Figure 2.10) then the histogram is referred to as

bimodal. The further the histogram shape diverges from bimodal the less effective global thresholding becomes as the threshold value to separate the foreground and background becomes less clear. When this threshold value becomes uncertain then there is a greater risk of false classification of either background or foreground pixels. The greater the quantity of noise and artefacts, particularly of intensities close to the foreground pixel intensities, the worse global thresholds perform.

Local thresholding These threshold methods address this sensitivity to the histogram shape deviating from bimodal (which can be caused by noise, artefacts, and uneven illumination among others) by evaluating the image itself. This is achieved by performing calculations within sub-regions of the image and applying the threshold value derived by these calculations to this sub-region itself [36, p. 72]. A notable local thresholding method is *Adaptive thresholding* [4, p.162] which calculates some threshold value by applying an operation across the pixel values within a square neighbourhood of shape $n \times n$ where n specifies the size of the neighbourhood. The neighbourhood is of an odd size so that the centre of the neighbourhood can be on the currently thresholded pixel. The operation applied to the neighbourhood to calculate said threshold, in *adaptive thresholding* [41], typically being that of the mean or the weighted (Gaussian) mean of the neighbourhood and potentially further tuning of the threshold strictness by the subtraction of some constant.

Some adaptive thresholding approaches employ the variance or standard deviation of pixel neighbourhoods as a thresholding criterion but there is no definitive adaptive thresholding approach as each was developed under a specific use case. Of note is that Adaptive thresholding is sensitive to neighbourhood size as the threshold is determined by the local mean and/or variance. The threshold employed is basically that the greater the magnitude of the central neighbourhood pixel, currently being thresholded, than the surrounding pixels then the higher the likelihood it will be preserved yet if the central pixel magnitude is quite similar to, or less than, the surroundings then it is removed. This functions well for high-contrast object edges and neighbourhoods large enough to capture both background and object pixels since neighbourhoods that are too small (fit entirely within the object with minimal background) can lead to loss of object pixels if the pixels composing the object are relatively homogeneous which is more typically inside object bodies with the greatest contrast typically being between objects and the background. Likewise, overly large neighbourhoods risk approximating the shortcomings of global thresholds when applied to this data which is that across the image the distribution of intensities between background and foreground is uneven. Adaptive thresholding methods have been implemented in *OpenCV* [42] and *imageJ* [41] by the method described above. Examples of alternative methods have been presented by *A New Local Adaptive Thresholding Technique in Binarization* [43] and *Adaptive Document Binarization* [44].

2.5. Further investigation into prominent deconvolution and image quality restoration methods

2.5.1. Otsu thresholding

Otsu Thresholding is an automatic thresholding method that determines a threshold to separate two classes of intensities within an image. These classes each occupy continuous ranges of grey levels (to be referred to as intensity values) across the image histogram such that the lower class will occupy a range of 1 to k and the upper class will occupy a range of $k + 1$ to L where L is some maximum intensity for an image and k is some determined intensity threshold for the image [40]. This threshold (k) is selected such that the variance within each class, the variance between both classes and the variance across the image is maximized. The Otsu threshold is not without shortcomings as a prominent issue is it requires the image to be bimodal (or at least approximately bimodal) such that it can determine an optimized threshold value.

Images with multi-modal histograms or images with noise that disrupts the desired modality of the image result in inaccurate threshold values. Over the years Otsu thresholding has been built upon in multiple papers to either improve the performance or to address multiple shortcomings. One such development resulted in an Otsu thresholding adaptation that can determine multiple thresholds for multi-modal images [45] but requires the expected number of classes to be provided before calculation. This is challenging in cases where the number of classes is unknown or the imaging conditions are such that the variance criteria are unrelated to legitimate classes creating confusion in the determined thresholds. In a theoretical and ideal example (illustrated in Figure 2.11), the threshold of a bimodal image will be within the valley between two peaks on the histogram where the between-class variance is maximal but in a case where the perceived modality is greater than one and a class expectation of two is provided, such that there are three peaks and two valleys on the image histogram, then the determined threshold may fall within either valley. In summary, Otsu thresholding performs best when the image histogram modality is two or more, the number of classes is known (the modality) and there are no aberrations across the histogram due to noise or other imaging conditions such that the modality of the image histogram is uncertain.

Do you think the bimodal distribution (Fig. 2.11) is sufficient for the explanation of the behaviour of Otsu with an ideal modality?

2.5.2. Hysteresis thresholding

Hysteresis thresholding technique that utilizes two thresholds and evaluates pixels/voxels, for the remainder of this description pixels/voxels will be referred to as pixels, relative to

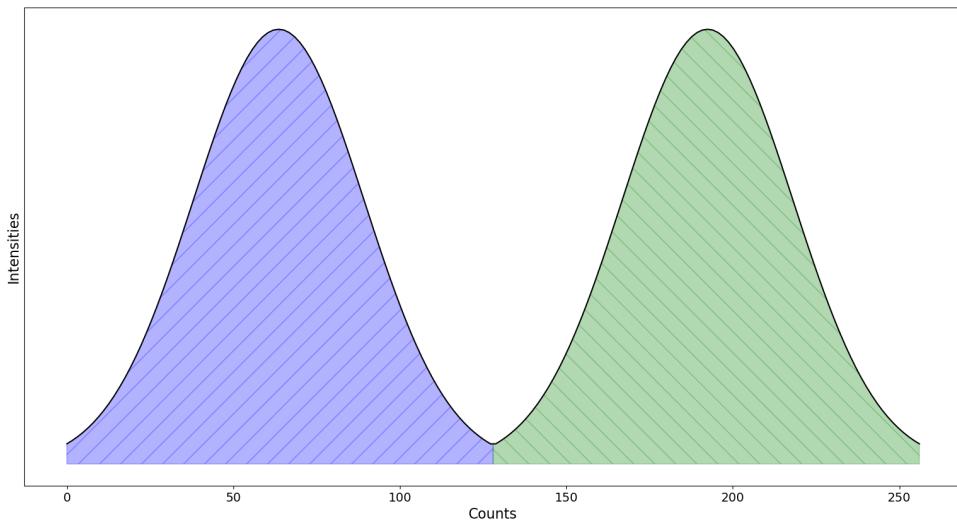
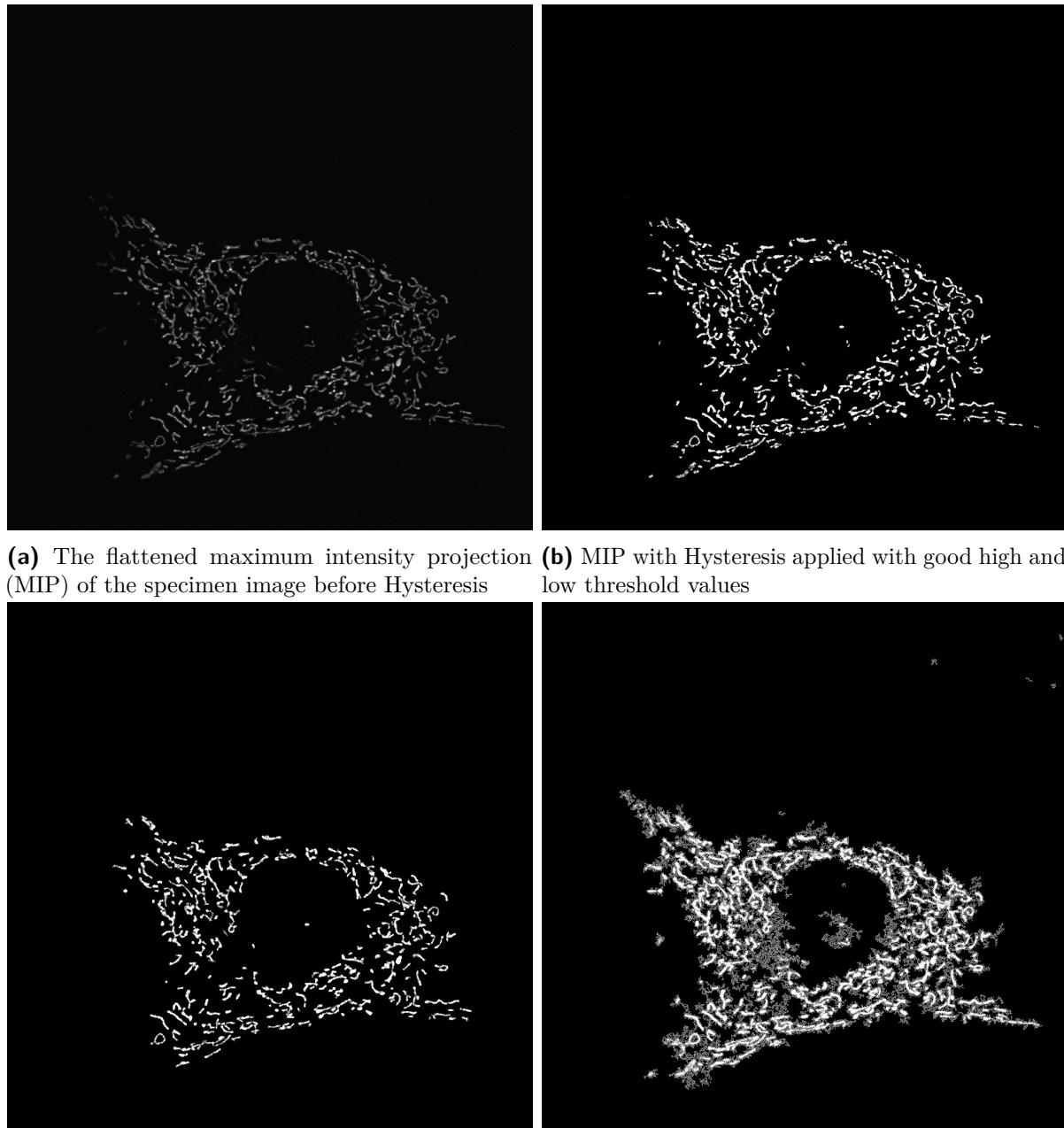


Figure 2.11: A hypothetical and ideal bimodal distribution with the two modes separated in the centre. Each mode is defined by different colours and hatching styles with the separation point being at the ideal threshold value

their surrounding pixels for threshold application. The application of hysteresis thresholding is such that there are two provided thresholds where one can be designated as the high threshold T_{Upper} and the other is to be designated T_{Lower} where $T_{Upper} \geq T_{Lower}$ and $T_{Upper}, T_{Lower} \in \Re$. Across the image, any pixels greater than the T_{Upper} than those pixels are immediately retained and any pixels below the T_{Lower} are removed. For pixels that have a value between the T_{Upper} and T_{Lower} , if it has a neighbouring pixel that has been retained (such as pixels with a value above T_{Upper}) then that pixel will also be retained. Thus structures containing pixels with continuous segments of pixels with values above the T_{Lower} which connect to pixels with values above the T_{Upper} are retained. It can be approximated that T_{Upper} has significant control over the number of structures and that T_{Lower} determines the possible volume of these structures where structures refer to continuous segments of pixels that represent some structure of interest in the image specimen/subject. A caveat of the Hysteresis threshold is the bias towards erroneously joining spatially close segments for values of T_{Lower} that are too low thus indirectly affecting the structures count. One of the earliest descriptions of this method is in *A Computational Approach to Edge Detection* [46] where the method was applied to address streaking in images. where edge contours in images where streaking are “*the breaking up of an edge contour caused by the operator output fluctuating above and below the threshold along the length of the contour*” [46, p. 689-690] which was described to be a problem more prevalent with single threshold value methods. An example of Hysteresis thresholding and the influence of erroneously selected threshold parameter values are shown in Figure 2.12.



(a) The flattened maximum intensity projection (MIP) of the specimen image before Hysteresis
(b) MIP with Hysteresis applied with good high and low threshold values
(c) MIP with Hysteresis applied with an overly strict (high) high threshold value retaining less individual structures
(d) MIP with Hysteresis applied with an overly lax (low) low threshold retaining volume that is outside of the expected structures

Figure 2.12: Comparison of the effect of Hysteresis threshold parameters with the: a) original image before thresholding, b) sufficient low and high thresholds, c) overly strict high threshold, and d) overly lenient low threshold shown

2.5.3. Triangle thresholding

Triangle thresholding is a method by which a triangle's geometric properties are applied to the histogram to determine the threshold. This method was employed in *Automatic Measurement of Sister Chromatid Exchange* [47] and a provided figure, shown in Figure 2.13, is used to visualise the logic employed. The peak intensity count value of the histogram is used as the height of the triangle and the range of intensity bins from the tail end of the histogram to the bins corresponding to the peak intensity is used as the triangle width. The tail end will always be in the direction of the histogram that is further from the peak value bin. The tail end is the longer length of the histogram bin range originating from the bin of the peak intensity count and either the upper or lower bounds. The lower bound is generally the zero bin and the upper bound is usually the bin corresponding to the maximum intensity value of the image. This can be seen in the third line of equation 2.3 where N is the maximum number of histogram bins; $h(n)$ is the bin count for the corresponding bin n ; n_{peak} is the bin with the maximum number of bin counts; h_{peak} is the maximum number of counts for a bin in the histogram; and W is the width from the peak bin n_{peak} to the further bound which is either the lower or upper bin range bound.

$$\begin{aligned} h(n) &= \text{hist}(image) \text{ where } n \in N \\ h(n_{peak}) &= h_{peak} = \max(\mathbf{h}) \\ W &= \max(N - n_{peak}, n_{peak}) \end{aligned} \quad (2.3)$$

With the triangle height and width established the hypotenuse of the triangle can be determined. Using the hypotenuse, a perpendicular line is measured from the hypotenuse to the histogram count value at a corresponding bin n where the bin n which provides the maximum perpendicular length is used as the threshold intensity.

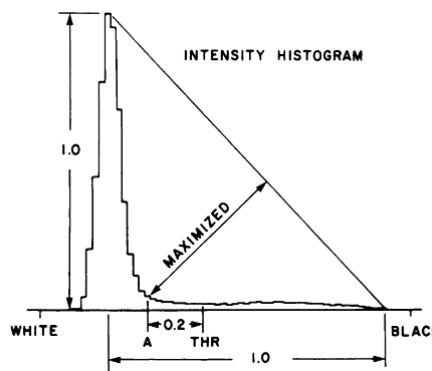


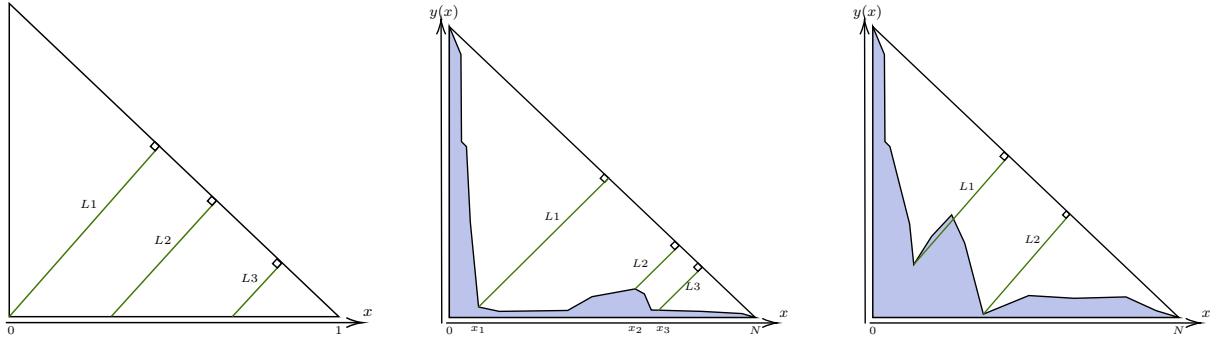
Figure 2.13: Depiction of normalized histogram with annotated measurements for threshold selection [47, p.742]

This method functions well when the threshold being sought is situated close to the peak within the histogram. When this peak represents noise then the method is more

effective when the angle from the peak value to the desired threshold value is steeper. If there are multiple modes in the histogram due to different noise sources or imaging properties then it can be more challenging to select an effective threshold value. This is due to the likelihood of the maximal perpendicular lengths being jointly dependent on the observed bin's closeness to the peak bin and the bin count value where the likelihood increases as the observed bin tends to the peak bin and the bin count value is lower. Figure 2.14 shows a series of hypothetical triangles to illustrate the impact of histogram shape on the threshold value calculations. In Fig. 2.14a the perpendicular lengths range from L_1 to L_3 with the chosen thresholds following the same order (to say that if L_1 was ignored then L_2 would be the next best threshold) where these distances are a limited subset for illustration, all bin count values will have a corresponding length. In Fig. 2.14b and 2.14c the effect of the histogram shape can be seen where the former depicts either a bimodal or dominant background bimodal histogram while the latter depicts a multimodal histogram. The bimodal histogram still presents the lowest illustrated bin value corresponding to L_1 as the best threshold since the separation between the two modes allows a large length to that bin. The multimodal histogram shows that the multiple modes, or crests, increase the difficulty in selecting the correct threshold as there is no clear separation and both L_1 and L_2 are similar in value (Fig. 2.14c). Image contrast can stretch the histogram providing greater resolution to separate these modes and false modes induced by noise in the image (making it appear multimodal) can impair the effectiveness of this threshold. Based on this Triangle thresholding is effective when there are bimodal or dominant background histograms but is sensitive to noise and inappropriate adjustments to the histogram shape. The main references are towards the beginning of this subsection which also informs later explanations but I did not explicitly reference them as part of this is from my inference of the method. Do you think I should cite those same sources again at this later point for credibility?

2.5.4. Richardson-Lucy deconvolution

The Richardson-Lucy (RL) deconvolution algorithm is based on the combination of two independently developed algorithms by W. Richardson [48] and L. Lucy [49]. For this exploration, the logic behind the algorithm will be described from W. Richardson's paper [48] as it was more general focusing on image restoration as opposed to L. Lucy's [49] application in the astronomical sciences field. RL deconvolution is an iterative algorithm that utilizes successive iterations to approximate a solution to minimize image diffraction. The development of this algorithm relied on a set of assumptions such as the distorted image can be represented as $H = W * S$ where H is the distorted image, W is the original image before distortion and S is the PSF representing the diffraction which distorts the image. The second assumption is that W , S and H are probability-frequency functions



(a) Depiction of the perpendicular lengths from the triangle hypotenuse to the x -axis

(b) A slope shape that is steep but shows the impact of a crest at higher bin values due to L_1 and L_2 and L_3 being similar

(c) A different slope shape with a secondary, smaller peak branching off of the decline from the first peak.

Figure 2.14: A depiction of the typical trend in perpendicular distances L relative to the slope shape. a) The histogram with no slope visible showing the lower x -axis (bin) values result in larger perpendicular distances. The impact of the slope shape is evident in (b) and (c) where the steepness in (b) results in L_1 being the largest distance while L_2 and L_3 are similar despite L_2 having a lower bin value. c) Shows the impact of a secondary peak branching off of the slope can lead to two similar lengths at different bin values. The bin (x -axis) value resulting in the largest perpendicular distance is used as the threshold

such that if i were some observed outcome of an event and there are I total possible outcomes where $i \in I$ then there will be multiple observations N where each outcome will be observed n times where $n \in N$. As this is a probability-frequency function the probability of each event outcome comes to be measured from the frequency of said outcome relative to the total number of observed outcomes e.g. $P(W = W_i) = \frac{W_i}{\sum^I W_i} = \frac{n_i}{N}$. From this $W = \sum_I W_i$ where each of these function probabilities can be reduced to the observance count per outcome divided by the total number of observations across all possible outcomes. Using a single dimension i is merely for descriptive simplicity and this can be expanded to multiple dimensions such that $W = \sum_{i,j} W_{i,j}$ and both i and j can be employed as the axis of an array or the coordinates of the pixels in an image where the outcome would be the intensity value of the pixel.

The crux of the method relies on two equations. The first equation is the application of Baye's theorem regarding the conditional probability of an event W_i given an event H_k where k is some observed outcome independent of i . The resulting formula (Eq. 2.4) [48, Eq. 1]:

$$P(W_i|H_k) = \frac{P(H_k|W_i)P(W_i)}{\sum_j P(H_k|W_j)P(W_j)}; i = \{1, I\}, j = \{1, J\}, k = \{1, K\} \quad (2.4)$$

where H_k is described to be some arbitrary cell k of H . This relationship is important as the probability of determining the correct value of W_i depends on the value of H_k as H is the distorted image from W must be restored while W is not known. Since S is the function of the distortion applied to each event i of W resulting in H for some k value(s)

thus $P(H_k|W_i) = P(S_{i,k}) = \frac{S_{i,k}}{S}$. A second consideration is that since W_i is not known then the best solution is to use an estimated W_i where a better estimate will result in a more accurate W_i . To approach this accurate W_i the process is performed iteratively where each W_i determined in a prior iteration is used as an estimate for the current W_i estimation where r is used to refer to the current iteration [48, Eq. 4]. Using both of these considerations an equation(Eq. 2.5) [48, Eq. 5] is determined:

$$W_{i,r+1} = W_{i,r} \sum_k \frac{S_{i,k} H_k}{\sum_j S_{j,k} W_{j,r}} \quad (2.5)$$

One final consideration is made due to the finite size of arrays in computers and equation 2.5 is changed in terms of the event summation bounds. The rewritten equation [48, Eq. 6] is:

$$W_{i,r+1} = W_{i,r} \sum_{k=i}^c \frac{S_{k-i+1} H_k}{\sum_{j=a}^b S_{k-j+1} W_{j,r}} \quad (2.6)$$

Where in equation 2.6, $a = (1, k - J + 1)_{max}$, $b = (k, I)_{min}$, and $c = i + J - 1$. For equation 2.6, W. Richardson describes that the summation over k and the summation with the bounds a and b in the denominator appear to be a corrective measure to adjust the estimated W_i over successive iterations where the denominator can compensate for escalating in value and encouraging the estimated W_i to converge to some value. This compensation in one estimated element $W_{i,r}$ may be further from the final converged $W_{i,R}$ value at some iteration to improve the convergence of some other element $W_{m,r}$ where $m \in I$. The event element i is one dimensional but the element variables can be expanded to account for multiple dimension arrays. A shortcoming of this algorithm is the potential for noise amplification of successive iterations although an algorithm developed from Richardson-Lucy is *Richardson-Lucy with total-variation regularization* which implements a regularization term to compensate for the noise amplification [35, p.33]. In this subsection I initially prefaced that the explanation will be based on W. Richardson's paper and I cite the individual equations I mention from his paper. Despite this do you think I need to cite him more frequently across this section?

2.6. Other techniques implemented within this research

2.6.1. Kneedle algorithm

Kneedle is an approach to determine ‘knees’ in a distribution [50]. Traditionally the knee, or elbow, of a distribution is described as some approximate point by which there is a change in the trend of the distribution. This is similar to the concept of the inflection point of the distribution but there is no standardised definition of a ‘knee’ rather it is the visual estimation of the point performed heuristically by a human viewing the visualisation

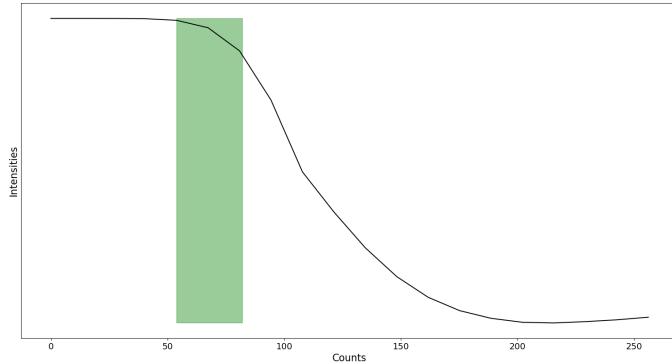


Figure 2.15: Approximate knees in a graph by heuristic evaluation. The highlighted region designates the likely region within which a knee point would be chosen. A range is displayed as the exact point is dependent on heuristic bias.

of the data. Due to this, it can be loosely defined as the point most significant to the change in the trend of the data. A visualised distribution with an estimated region in which a ‘knee’ would be selected is shown in Figure 2.15.

One method of interest that attempts to address this is *Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior* [50] which both presents a mathematical definition of a ‘knee’ and the method by which this ‘knee’ is determined.

Knee Definition Firstly, it was stated prior that the knee of a distribution is visually similar to that of the inflection point yet this point does not capture the point at which trends change but rather a change in the sign of the curvature of the distribution. The chosen definition is based on the curvature of a continuous definition where the point at which maximum curvature is expressed is the closest to what would be the knee point [50, p.2]. The equation describing curvature ($K_f(x)$) is shown in Equation 2.7 where $f(x)$ is the distribution for which the curvature is being described.

$$K_f(x) = \frac{f''(x)}{(1 + f'(x)^2)^{1.5}} \quad (2.7)$$

This curvature equation performs well with continuous functions yet is challenged by discrete distributions. The causes behind these challenges are related to the incompatibility encountered when fitting a continuous function to a finite distribution as a minimum of 3 data points are required to calculate the curvature (centred on the second data point) thus the curvature for the first and final point in the discrete set cannot be determined. Another major challenge is that determination of the maximum curvature within the distribution requires this calculation to be performed across the whole discrete set and it is possible for the point of maximum curvature to exist outside the range of the finite set.

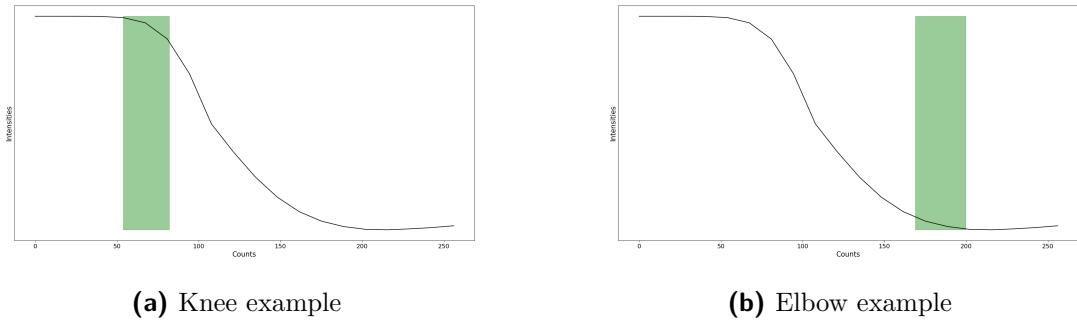


Figure 2.16: Knee and elbow approximation regions are shown for some given distribution

Knee method It is explicitly stated that “Kneedle”, the method by which the knees are determined, is predicated on the assumption that the knee is the point expressing the maximum curvature but there can be multiple knee points within a distribution or set. For multiple points to be approximated it is the local maxima of the curvature across the set that describes a knee point. These local maxima of the curvature are determined by “*that are local maxima if the curve is rotated θ degrees clockwise about (x_{min}, y_{min}) through the line formed by the points (x_{min}, y_{min}) and (x_{max}, y_{max})* ” [50, p.3]. Thus far the designed method presumes that the knee point is for an increasing curve that is of negative concavity but this can be performed for curves with a positive concavity (an “elbow” as opposed to a knee) by inverting the distribution with examples for each shown in Figure 2.16. The method through which the Kneedle algorithm determines the knee points can be separated into six steps given some data set D which is a finite discrete set of paired values (x_i, y_i) :

1. Smooth the discrete data set (D) to reduce the noise in the data. The retention of the shape of the set must be maximised.
 2. Normalize the finite data set points as the algorithm behaviour is independent of the point magnitudes. Both x and y values are normalized to a unit square while preserving their shape, as shown (Equation 2.8):

$$D_{sn} = \{(x_{sn}, y_{sn})\}, \text{ where}$$

$$x_{sn_i} = (x_{s_i} - \min\{x_s\}) / (\max\{x_s\} - \min\{x_s\}) \quad (2.8)$$

$$y_{sn_i} = (y_{s_i} - \min\{y_s\}) / (\max\{y_s\} - \min\{y_s\})$$

3. A difference set (D_d) represents the differences between the $x-$ and $y-$ value pairs which are mapped as $(x_{sn}, y_{sn} - x_{sn})$. This constructs a difference curve where the increases in the $y-$ values are relative to the magnitude of the paired $x-$ values. The insight provided by this difference curve is the behaviour it describes. When the curve transitions from horizontal to decreasing at a point indicates that said point is a potential knee.

4. As mentioned prior, it is where the difference curve transitions from flattening to decreasing denoting the presence of a knee. Under the assumption that the points of the set are not decreasing, it is the local maxima of the difference curve that embody positions around which the rate of increase decreases noticeably. Each of these local maxima of the difference set is a candidate knee point.
5. Each candidate knee occupies a local maxima where a certain deceleration in the smoothed data increase is perceived but the quantity of deceleration over a number of points defines what is declared as knees. At each candidate knee of the difference curve, a threshold value is calculated with a tunable sensitivity parameter (S). This threshold (T_{lmx_i}) is calculated by calculating the mean of the forward difference, with a spacing of 1, $(x_{sn_{i+1}} - x_{sn_i})$ which is multiplied by the sensitivity value (S) which is finally subtracted from the candidate knee value (y_{lmx_i}) for some candidate knee (i) shown in Equation 2.9.

$$T_{lmx_i} = y_{lmx_i} - S \cdot \frac{\sum_{i=1}^{n-1} (x_{sn_{i+1}} - x_{sn_i})}{n - 1} \quad (2.9)$$

6. Each calculated threshold value is unique to each candidate knee (difference curve local maxima) and it is the decrease in the difference curve, after a candidate knee, relative to this threshold that determines whether it is a valid knee. The criteria are that the difference curve values, after a candidate knee, must decrease below the associated knee threshold ($y < T_{lmx_i}$) value prior to either increasing from a local minimum or encountering another candidate knee. If a local minimum of the difference curve is encountered, after which the curve increases, then the next candidate knee is evaluated. The smaller the sensitivity parameter (S) the closer the threshold to the candidate knee value (local maxima) and the easier it is to pass the criteria.

2.7. Prominent literature used in this research

Throughout this chapter, various pieces of literature have been used in developing the system itself (detailed in Chapter 3) or to develop an understanding of the biology and image processing that informed future decisions indirectly. For this breakdown, the literature will be separated by those that provided a foundational understanding of the fields involved with this research, those involved with the pre-processing of the image data prior to the thresholding system, and last will be the literature that was involved in the development of the automated thresholding system itself.

Foundational understanding

Biology The foundation of this came from a high-level understanding of the biological phenomena that describe the organelles of interest and the mitophagy process which is the process that will be analyzed after complete pre-processing. The largest contributor to understanding mitophagy and the related organelles was *Human Physiology: From Cells to Systems* [1] which covered information regarding autophagy, the mitochondria and, to a lesser degree, mitophagy. More detailed information relating to autophagosomes, lysosomes and the relationship between the two is detailed in *The discovery of lysosomes and autophagy* [11]. Detailed insight on mitophagy and its relationship with the fission and fusion of mitochondria is covered in *Mitochondrial fission, fusion, and stress* [12] and *Mechanisms of mitophagy* [13].

Microscopy The fundamentals provided by microscopy are to recognise the causes of reduced image quality or errant phenomena observed that originate from the imaging process. A particular impact on this understanding of confocal microscopy is *Fluorescence microscopy* [15] but it is the constraints listed as bleed-through, photobleaching and diffraction that are the most valuable as these understanding of these constraints can inform undesired behaviours or details seen in the image during the analysis that might not relate to the actual specimen. Bleed-through was detailed by *Fluorescence microscopy* [15] and *A guided tour into subcellular colocalization analysis in light microscopy* [20]; photobleaching was informed by *Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy* [22] and *Loss of image quality in photobleaching during microscopic imaging of fluorescent probes bound to chromatin* [23]; and diffraction was informed by *Diffraction-unlimited optical microscopy* [24] and *Diffraction-unlimited optical microscopy* [25].

Pre-processing understanding

Since this research is exploring and developing an automated high-throughput thresholding system it is important to understand other pre-processing techniques that are applied to the image prior to the application of the system. Likewise understanding the influence of these techniques regarding the manner in which they improve image quality (noise removal, signal enhancement, etc.) is crucial as each of these methods can impart an undue effect on the image as a byproduct of their process and mitigate these byproducts using subsequent pre-processing techniques, including the automated thresholding. An overview of deconvolution, the generation of the PSF, and a high-level mathematical representation of the diffraction effect and its removal via deconvolution is described by *3D PSF models for fluorescence microscopy in ImageJ* [33] for the generation of synthetic PSFs, and by both *Blind deconvolution of gaussian blurred images containing additive white gaussian noise* [34]

and *Deconvolutionlab2: An open-source software for deconvolution microscopy* [35] in terms of the mathematical representation. For denoising, a statistical understanding of noise (not artefacts or bleed-through) is provided by *Analyzing fluorescence microscopy images with ImageJ* [36] while the behaviour of standard denoising filters is informed by median and Gaussian filters but also by the Sigma filter [37]. For image enhancement, the primary insight developed was through analysis of two such methods for background subtraction and contrast enhancement described by *Biomedical image processing* [38] and *Contrast Limited Adaptive Histogram Equalization* [39] respectively. The last group of methods were those relating to thresholding which is separated into global and local methods that segment the image into foreground and background thus binarizing it. Thresholding as a whole was described by *Chapter 9 - image segmentation* [4] in the context of image segmentation while further detail regarding global and local methods was informed by [40] and *Analyzing fluorescence microscopy images with ImageJ* [36]. These pre-processing methods were all to be applied sequentially within a pipeline and a paper which stimulated the exploration into why these methods are applied sequentially was provided by *A pipeline for multidimensional confocal analysis of mitochondrial morphology, function, and dynamics in pancreatic β -cells* [10].

Literature that directly impacted system design

The research that directly featured in the automated thresholding system being developed or that was compared against for evaluation consists of a number of thresholding techniques and other algorithms. These thresholding techniques are the Otsu threshold [40]) and Triangle threshold [47] as they are both automated threshold methods reliant on image histogram being bimodal or close to. Hysteresis thresholding, proposed by [46], is important as it forms the backbone of the system design while the Kneedle algorithm [50] is an unconventional approach for low threshold approximation detailed in Chapter 3.

2.8. Chapter conclusions

This chapter conveyed theory and concepts relating to the techniques that will be implemented or are to be explored throughout this work. This theory extended from providing context around the approaches dictated by the mitophagy research use case to those being explored for implementation in the high-throughput thresholding system design. The mitophagy research use case determines the objects to be imaged (Section 2.1), the microscopy technique to be used (Section 2.2), and the analysis approach (colocalisation) that the images must be of suitable quality for (Section 2.3).

The thresholding system being designed is not expected to function in isolation with a number of adjacent approaches applied before or after the system to maximise the restored

image quality. For this reason, the context was provided regarding factors of image quality degradation (Section 2.4.1), a high-level overview of the sequential application of approaches to improve the image quality (Section 2.4.2), and a high-level overview of these approaches grouping them by the aspect of image quality that is addressed (Section 2.4.3). There is also a deeper discussion on the theory surrounding specific approaches of interest to this work (Section 2.5).

This chapter serves to consolidate and organise all of the theoretical backgrounds of phenomena, techniques and approaches that inform design choices and observed behaviours of the high-throughput thresholding system. The proceeding chapters (Chapter 3 and Chapter 4) both reference this chapter in terms of both the specimens and imaging techniques used for image acquisition, and the techniques implemented in the system design to develop a high-throughput thresholding system.

Chapter 3

Methodology

With Chapter 2 providing background context for the

3.1. Dataset acquisition process and preparation

The image data set that was acquired for this research project is composed of microscopy images that can be used in mitophagy analysis. This is due to the primary intended application of this automated binarization system being in the high-throughput pre-processing of microscopy images in mitophagy research.

3.1.1. Sample preparation prior to imaging

Cell culturing

MEF and GT1-7 cell lines were used and were maintained in Dulbecco's Modified Eagle Medium (DMEM) (ThermoFisher, #41965062) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich, #F0679) and 1% Penstrep (Sigma-Aldrich, #P4333) and were kept in a humidified incubator at 37°C and 5% CO₂. Cells were subcultured using trypsin (Sigma-Aldrich, #T4049) to detach cells from flasks. Afterwards, DMEM was added to the cells at a 2:1 ratio and cells were collected into 15 mL Falcon tubes (Bio-Smart Scientific, #50050). Cells were centrifuged at 1500 rpm at room temperature for 3 minutes (Centrifuge, Eppendorf, #5804R). Media was discarded and cells were resuspended in fresh DMEM. Cells were seeded to T25 flasks (Bio-Smart Scientific, #70025) 8-chamber dished (ThermoFisher, #Z734853).

Treatments applied

All treatment interventions were made up using 1x phosphate-buffered saline. The treatments applied were memantine hydrochloride (Sigma-Aldrich, #PHR1886), carbonyl cyanide m-chlorophenylhydrazone (CCCP) (Sigma-Aldrich, #857815), and baflomycin A1 (Baf) (Sigma-Aldrich, #B1793). Only the GT1-7 cells were treated with Memantine hydrochloride which was done by incubating the treatment with the cells at working concentrations of either 50 μM (LM) or 100 μM (HM), low and high concentration dosage

respectively, for 48 hours prior to any imaging and while refreshing any memantine-treated media every 24 hours. The CCCP treatment was incubated 6 hours prior to imaging at $10 \mu\text{M}$ causing mitochondrial depolarization with the goal of increasing the frequency of mitophagy events to observe them. The Baf treatment was incubated with the cells 4 hours prior to imaging at 400 nM to inhibit the completion of mitophagy by preventing the fusion of lysosomes and autophagosomes (described in Section 2.1.2) resulting in damaged mitochondria to remain contained in autophagosomes. Different combinations of these treatments formed treatment groups, with no treatments forming a control group, so that observations from the analysis could be better correlated with the appropriate treatments.

I feel that this, due to being basically copied from Sholto, should reference the same paper in case it is flagged for plagiarism

3.1.2. Image acquisition via microscopy

The mitochondria, autophagosomes, and lysosomes were imaged for the image data set used in this research as they are the organelles associated with mitophagy. To image these intra-cellular organelles with specificity and easy differentiation confocal microscopy was used as fluorophores can be selected specific to the organelles of interest. The imaging was performed using a Carl Zeiss LSM780 ELYRA PS.1 Super-resolution platform using an LCI Plan Apochromat 100x/1.4 Oil DIC M27 objective at 100x magnification with the lateral slices composing the 3D z-stack having a step-size of $0.35 \mu\text{m}$ between each slice with the number of slices ranging from 6 to 13 depending on the sample.

Green fluorescent protein - light chain 3 (GFP-LC3) plasmid (Addgene, #21073), LysoTracker Red (ThermoFisher, #L7528), and MitoTracker DeepRed (ThermoFisher, #M22426) were used to fluorescently stain for autophagosomes, lysosomes and mitochondria respectively. For GFP-LC3, $5 \mu\text{g}$ were transfected into the cells using the NEON Electroporation System (Invitrogen, MPK5000) and NEON Transfection tool kit (Invitrogen, MPK10096) according to the manufacturer's instructions. These transfected cells were seeded onto 8-chamber dishes 24 hours before the treatment protocol. The LysoTracker Red and MitoTracker DeepRed were combined into a master mix solution at 75 nM each and were incubated with the cells for 20 minutes prior to imaging. The GFP, Lysotracker and Mitotracker used were excited during fluorescence microscopy by lasers with a wavelength of 488nm, 561 nm and 633 nm respectively with 561 nm and 633 nm kept on separate tracks to avoid cross-talk and bleed-through between them.

3.1.3. Preprocessing prior to thresholding

After the cell images have been acquired they are pre-processed to improve the image quality before any analysis is undertaken. The pre-processing pipeline that was implemented entailed:

1. Deconvolution: Richardson-Lucy Total Variation (RLTV) Deconvolution [35, 51]
2. Upscaling: Bicubic interpolation
3. Background subtraction: Rolling ball algorithm [38]
4. Denoising: Sigma filter plus [37]
5. Image enhancement: Contrast Limited Adaptive Histogram Equalisation (CLAHE) [39]
6. Gamma correction

For each image in the batch the number of lateral z-slices (z-resolution), the NA, the channel wavelength, the refractive index immersion, the pixel resolution, the pixel size and z-step size would be used to generate a theoretical PSF for the said channel using software [33]. Since the channels were associated with specific fluorophores and all other parameters, except for the z-resolution (z-stack size), the PSF could be generated prior and applied to all images matching those parameters.

With the PSF generated deconvolution is performed using a Fiji [9] plugin called DeconvolutionLab2 [35] and the applied techniques is RLTV [51]. This differs from the Richardson-Lucy deconvolution described prior (Section 2.5.4) where the addition of the total variation regularization reduces noise amplification while still preserving the edges of objects. Bicubic interpolation is used to upscale the deconvolved image prior to further processing. Once upscaled, background subtraction is applied to correct any uneven illumination that may be present in the image followed by denoising. With the illumination balanced throughout the image, the sigma filter can be applied to remove noise. This filter is similar to a mean filter which is with the added benefit of better edge preservation and robustness to outliers within the pixel neighbourhood. Image enhancement through CLAHE is applied which compensates for some of the signal loss due to denoising and the localised application of it to pixel neighbourhoods means that the contrast enhancement is adapted to the specific pixel regions. This results in a contrast adjustment that is not sensitive to localised fluctuations in the contrast providing a better end result given that the parameters are suitable for the image. Lastly, gamma correction is applied to the image and is typically followed by binarization to produce a mask of the objects of interest depending on the analysis. Since this research project entails the development of a binarized mask designating the foreground and background objects in the image all images will be pre-processed as described above.

3.2. Automatic thresholding method

In order to achieve the objectives discussed in Section 1.3, a new image binarization method was developed. Initially, both automated global and local thresholding methods were experimented with where it was heuristically determined that each had caveats preventing them from meeting these objectives. While automated global thresholding methods require no mandatory parameter selections the results achieved are highly sensitive to both the imaging conditions and the image quality. Automated local thresholding methods are more robust to image quality due to the ability to evaluate the threshold based on pixel attributes that are significant in localized regions (such as the contrast of an object edge) but are dependent on the correct parameter selection and tuning to produce optimal results. Due to this a new method was developed with the intent of adopting the benefits of each of both global and local being the lack of parameter tuning and the robustness to poor image quality.

3.2.1. Choice of threshold method

Rather than attempt to establish an entirely new binarization method the improvement of an established method was deemed prudent. Since such a method would have already been tested prior benefits and drawbacks of such a method were already known and could potentially be addressed through this method.

Hysteresis thresholding (Section 2.5.2) was chosen as the results were similar to that of local thresholding methods where the outcome was the intensity of pixels (or voxels) within an object where considered and allowed individual objects to be validated. This was also similar to global thresholding methods as the parameters to be selected were intensities which could be inferred from the image histogram or be determined by other image intensity-based criteria. While methods like Adaptive Thresholding may provide more robust results, the necessity of visual analysis in the approximation of the parameters including the tuning of said parameters makes it, along with most other automated local thresholding methods, either inappropriate or less appropriate for automated high-throughput thresholding. Using the showcase of the Hysteresis application in Figure 3.1, an analogy can be presented to both summarize the logic behind Hysteresis thresholding and the role each parameter plays in it. Assume that the pixels of the image described a terrain composed of a landmass with heights relative to the pixel intensities displayed in Figure 3.1b. Certain regions of this landmass are of interest typically those that would be islands possessing mountains of a certain height. The sea level and mountain threshold are each visualised by a #blue and #yellow plane respectively where regions of land above the ‘water’ are considered independent regions, or continuous contours, with each being a

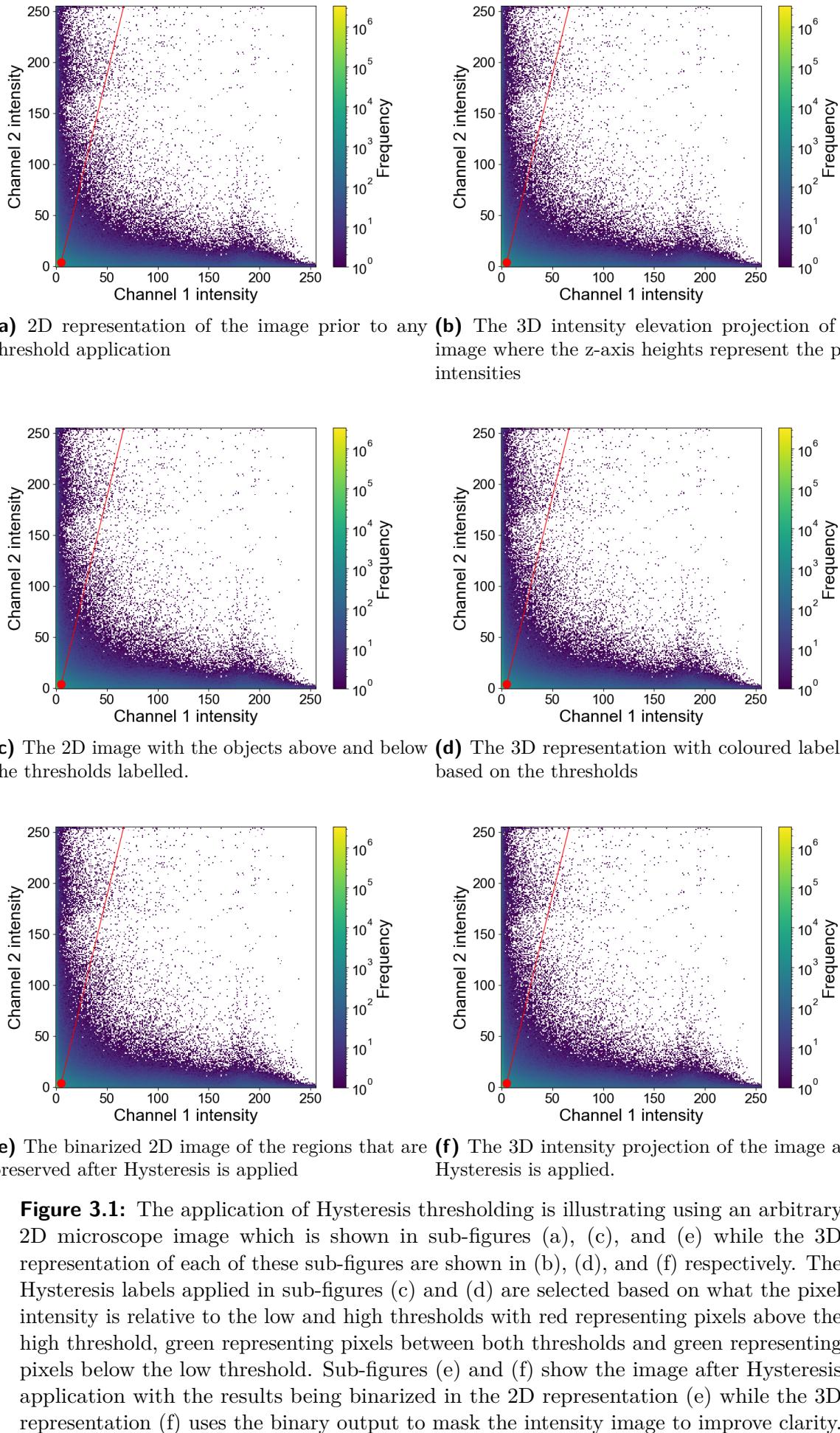
potential island of interest as seen in Figure 3.1d. Islands that intersect with the #yellow plane are those that can be deemed to be mountains and said islands are then deemed to be of interest. A representation of this selection can be shown in Figure 3.1f where only the islands meeting said criteria remain. In this analogy, the sea level represented the low threshold, the mountain height threshold represented the high threshold while the continuous regions above the low threshold represented potential foreground regions.

Considerations due to this were determined early in the system prototyping when postulations regarding how image properties or behaviours could be related to the thresholding and be quantified using an algorithm removing human involvement. This is performed in automated local thresholding methods like Sauvola [44] where a number of parameters can be left as their defaults but these only function well under specific circumstances and can collapse if the image differs too much from the image around which the defaults were approximated. A number of these methods were used earlier for pre-processing and it was developed method was to be a medium built upon Hysteresis thresholding and what components of Hysteresis thresholding could be addressed through this method. The primary components are the binarization logic the method employs and the parameters which need to be selected to execute this logic which typically involve some level of visual analysis to approximate appropriate parameters. It is the latter component is the parameter selection which prevents high-throughput application yet provides most of the robustness as the low and high thresholds can be tuned to reduce the impact of image degradation on the binarization result.

The system design based on this will target the determination parameter selection in a manner that is both automatic and adapts to each image by allowing them to be automatically "tuned" to each image by considering features related to the image pixel intensities. Based on the analogy described prior, the low threshold must first be acquired as it determines what pixels compose potential foreground regions while the acquisition of the high threshold will follow to filter which of said regions actually belong to the image foreground.

3.2.2. Acquisition of the low threshold

As has been described in Section 2.5.2, the low threshold is a crucial step in determining the quality of the Hysteresis thresholding outcome since it is the first threshold value deciding which voxels are background. This is important as the high threshold only acts on the remaining non-background voxels in the image thus a low threshold that is too low or too high results in either image noise that increases the difficulty of the high threshold selection or removes valuable regions of objects. Despite this, thresholding is a subjective operation frequently relying on the judgement of an expert to evaluate what is currently



designated as background voxels in the image and whether they belong to an object of interest to the analysis.

Initial steps

In the beginning, an investigation into robust and accurate low threshold estimators was performed but it was found that the established methods were lacking and for this reason established automated global threshold methods were considered. Automated global threshold methods had the benefit of not requiring parameter setting or tuning (aligning with a design objective of the system this research was developing) and there was a robust body of literature on these methods regarding their development. Many of these methods also were implemented in Python libraries thus the methods ‘Li’, ‘Mean’, ‘Otsu’, ‘Triangle’, and ‘Yen’ [6] were tested. These methods were tested by application to the 25 sample images with the threshold results compared to the low threshold values selected by the experts.

Initial findings were that values chosen by ‘Li’ or ‘Mean’ were often far lower than the experts while ‘Yen’ resulted in threshold values that diverged so greatly from all of the experts that Yen was omitted from further testing. Despite thresholding being a subjective evaluation, as mentioned prior, it is still ideal to achieve a closer value to the experts although how close, unfortunately, is image dependent. The remaining two methods, ‘Otsu’ and ‘Triangle’, were generally closer to the expert values although Otsu was generally closer to the experts. *The table is empty for now as I need to determine if formatting functions well for horizontally long tables using tikz or tabular. For now, this is a placeholder and the boxplots visually describe this better (perhaps this should be moved to appendices?)* This is visually described in Figure 3.2 where the overlaid box plot and strip plots represent the difference between each thresholding method and the experts for each sample. Although this is not a perfect comparison as each sample is technically unique to each other and this comparison is reductionist in that it does not consider the conditions of each image that lead to certain results but this can serve as an estimation of the sensitivity of the performance of the thresholding methods and how robust they are to a range of image conditions. As these comparisons regard the difference between the threshold method and the expert results tending towards zero are better as they are more similar to the expert with Figures 3.2a and 3.2b describing the difference between the threshold value and the number of foreground voxels respectively. Although it was already established that Otsu and Triangle were the best-performing threshold methods there is confirmation of this with the threshold value difference (Fig. 3.2a) where Otsu has larger quartile ranges than that of Triangle but with the distribution of sample values more closely centred on the zero line. What this means is that Otsu experiences greater swing in performance across samples yet when it does perform well it performs better

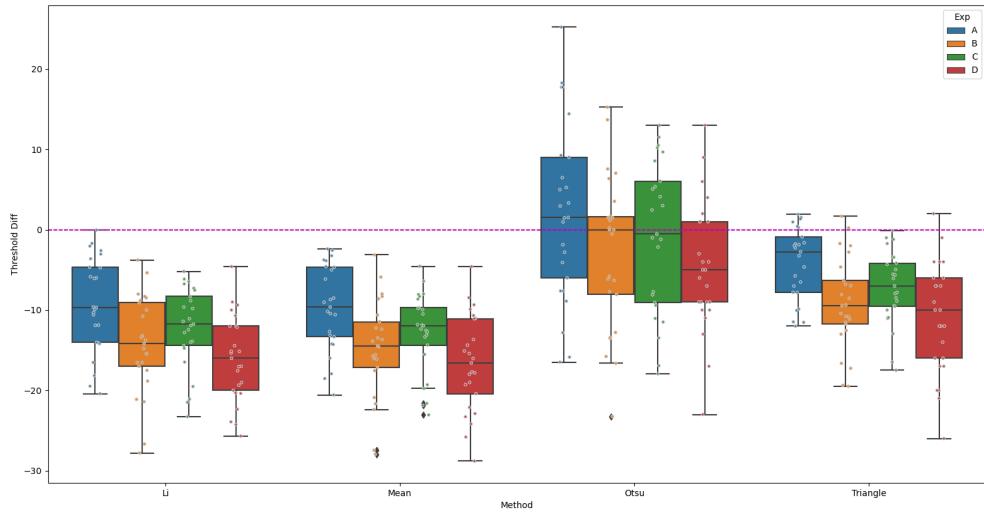
that Triangle yet Triangle is more consistent in outcome although this outcome is slightly worse. Image thresholding remains a highly subjective and image-dependent evaluation and for this reason, although still an unsatisfactory measure of overall performance, an evaluation in the binarization output per sample (Fig. 3.2b) is to be explored. This difference in foreground voxel counts shows that save for Expert A, Otsu has greater consistent similarity to the experts although the shortcoming is that when it deviates it can deviate strongly. When this analysis was conducted there was also an evaluation of the binarization outcome on the samples between the thresholding methods and the experts although only these reductive comparisons have been shown as they serve as effective approximations of the binarization results across a large body of samples. What was determined is that although these methods could provide passable results, the impact of poorly selected low thresholds can render a binarization outcome useless for analysis, if not detrimental if uncaught, thus an alternative thresholding method is required.

The alternative thresholding method that was decided on was an application of the ‘Kneedle’ algorithm (Section 2.6.1) where elbows within the image histogram are to be used as the low threshold. The motivation for this decision was based on comparisons between the expert-selected low thresholds and where they are positioned on the image histogram as it was hypothesised that there is some low-intensity region of the histogram that would represent the image background. The challenge surrounding this was, once again, that image binarization is a subjective process where typically some value can be estimated and based on its impact on the binarization it is further tuned. Despite this, it was hypothesised that there must be some baseline which would generally be agreed upon by experts regardless of the results. This baseline was formulated from four notions being:

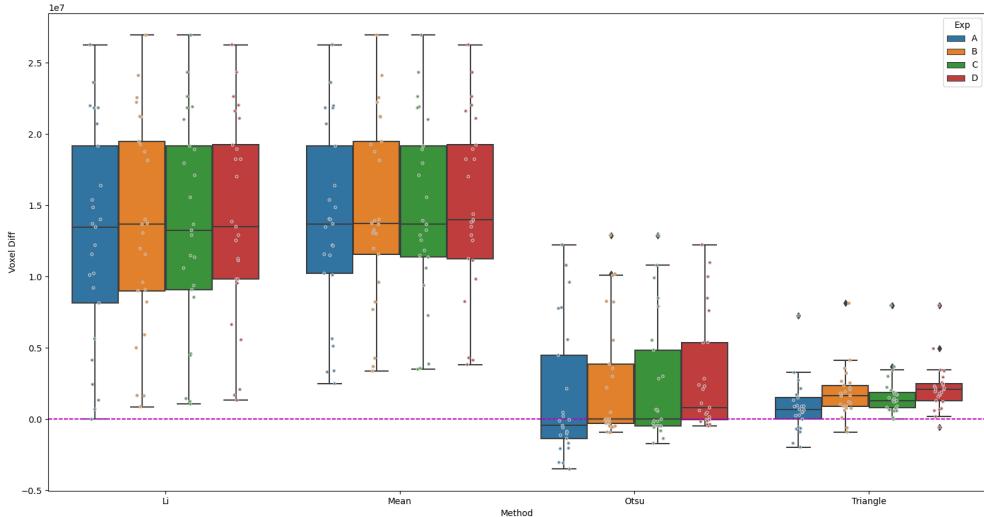
1. That there will always be a background within the image, as there must be two states to differentiate between with binarization, where the background will be comprised of voxels with intensities ranging from the lowest in the image up to, or below, the lowest foreground intensity.
2. While the upper bound of the background intensity range is uncertain initially, the lower bound is well-defined. Given this, there is complete certainty that the lowest intensity belongs to the background but this does not hold true for consecutively greater intensity values.

Note: This background upper bound is uncertain in “lower quality” images where there is no clear separation between the background and foreground modes due to either noise or other irregularities.

3. The background is composed of two potential “groups” of the background of which the lower intensity is referred to as the ‘empty’ background and the high intensity is



(a) Difference between the low threshold determined by the automated method and the experts



(b) Difference in non-background voxel count between the automated methods and the experts

Figure 3.2: Difference between expert values and automated thresholding results in terms of the threshold value and binarization outcome depicted using box plots overlaid with strip plots. These plots describe the difference between the thresholding method and the experts for each sample where (a) represents the difference in the numeric threshold value while (b) represents the difference in the count of the foreground image pixels. As greater similarity is preferred the goal is a value of zero denoted by a dashed magenta line on both sets of plots.

the ‘irregular’ background. The term potential was used as the ‘empty’ background is always present and this is the background voxels that do not describe any objects within the image but rather is the absence of any observable objects or phenomena which typically exist in a small near homogeneous intensity range. This composes the space between the foreground objects in the image but there is another group of backgrounds which can appear during unwanted imaging conditions which is the ‘irregular’ background.

Note: This ‘irregular’ background is an umbrella term for the background elements in the image composed of fluorescence signals which are unwanted for the analysis or were not intended to be captured, e.g. a small amount of a biomolecules autofluorescence emission is within the filtered spectra and thus is included appearing as a faint shadow. This fluorescence is referred to as irregular as it does not follow noise models nor is it distinct enough to typically present as foreground objects thus it is irregular with respect to what is supposed to, or expected to, compose the image background.

4. With pre-processing applied it is expected that background induced by noise will be near homogeneous and describe some “empty background” where no actual or false objects are depicted and exhibit among the lowest intensity values in the image. From this, our certainty is highest among the lowest intensities where the “empty background” is situated and diminishes as the voxel intensities originate from irregular background conditions.

Note: This view was motivated by evaluations of the image histograms and the impact of the low threshold values on the image by image analysis. This was depicted by a high peak spanning a narrow range of intensity values centred at the lowest histogram values.

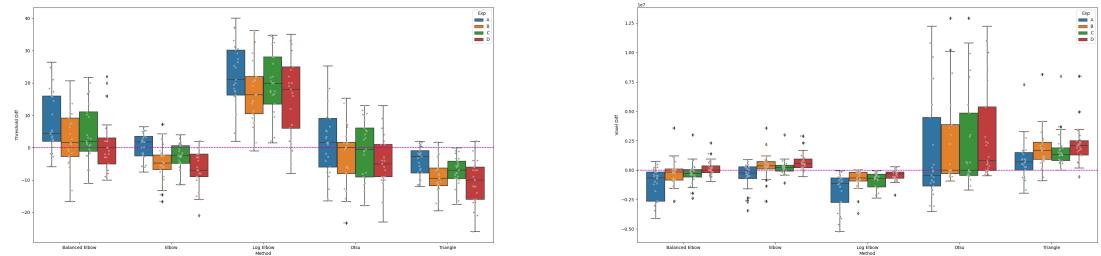
Based on these notions a theoretical framework was established which can be referred to as the “low-intensity background confidence” where the confidence that an intensity value belongs to the background is relative to how low the intensity is and that the confidence for an intensity value cannot be greater than any lower intensity values e.g. if we are A confident that voxels of intensity x belongs to the background and B confident that voxels of an intensity $x + 1$ belong to the background then $A \geq B$ must hold true for the confidence. Despite this algebraic example, the confidence is closer to that of an informed intuition instead of some known distribution of function. Regardless, our most confident background selection lay from where the low-intensity peak significantly flattened in slope and thus a method to approximate this best point of plateau became necessary. This does not imply that all voxels of an intensity greater than this point do not belong to the background but rather were deemed to be a point which would sufficiently

capture the majority of the noise which would be most impactful to the binarization as the low threshold determined the volume and joining of potentially foreground objects and since the “empty background” constituted the space between the foreground, insufficient binarization could induce the erroneous joining of foreground structures. **Should I include some (say two) example histograms where I annotate where the low-intensity peak in the histogram lies?**

Using Kneedle for low threshold approximation

The “Kneedle” algorithm (Section 2.6.1) is applied to the image histogram with the goal of approximating a sufficiently low threshold such that at least all of the ‘empty’ background is captured within its range. This relies on the fact that Kneedle can determine either the knees or elbows within a function which can be characterised as the point of most sudden change in the gradient of the function after either increasing or decreasing respectively. For this application, the ‘empty’ background is assumed to always be described by some crest among the lowest intensity values and it is the fall from this crest that is of interest. It was only after the application of this method and the evaluation of the binarized regions, regarding what it was certain was background from what might be foreground, that the effectiveness was observed. While the method at times fell short of the Expert values it frequently excluded the ‘empty’ background values at least and was found to generally perform well. This can be seen in Figure 3.3 which compares the performance of the Kneedle low threshold, as well as Otsu and Triangle for reference as the prior best performers, to the Experts. While this comparison is reductive as it does not evaluate what is being binarized in the image (objects, shapes or patterns typically are passed through for evaluation by the High threshold) since it is a linear threshold it is known that if two thresholds are identical then their impact on the image binarization will be identical. From the results, it is seen that the results acquired by Kneedle are closer to the Experts across the span of samples and that the variance in the performance is tighter describing a more reliable outcome which is easier to predict. What may be noticed is that there is also a mention of “Log Kneedle” in both figures as well as “Chosen Kneedle”.

Log Kneedle is the low threshold elbow that is acquired using the log of the histogram as opposed to the normal histogram. This application was hypothesised based on the notion that taking the log of the histogram will normalize and rescale it allowing the higher-intensity histogram components to be discerned if they have been drowned out by a huge ‘empty’ background peak which can result in low threshold values that are too low. This is applicable as this log normalization allows large changes in the ‘empty’ background, irregular background, and potential foreground intensity range to be evaluated in the histogram typically resulting in a higher elbow for the low threshold. Looking at Figure 3.3a and 3.3b it can be seen that the results of the Log Kneedle are highly



(a) Difference in threshold metric to be applied as the Hysteresis low threshold **(b)** Difference in voxel count post-binariization using the threshold value as the image would appear after low threshold application

Figure 3.3: Comparison between the thresholding method and the Experts for Otsu, Triangle, normal Elbow, log Elbow, and the chosen Elbow. This comparison is depicted using box plots to describe the variance in the difference between the experts across multiple samples and the simplified distribution of these values.

varied in performance, particularly so for Expert A which also performed the worst, but its perspective on irregular background presence that the ‘empty’ background might overshadow was believed to hold merit but not in the current approach.

The Chosen Kneedle was the approach which was established to tune whether to get the normal or log-transformed low threshold under some criteria. The criteria that were chosen were that if the Otsu threshold exceeded that of the Kneedle elbow then the log Kneedle elbow would be used instead. This was under the assumption that Otsu will be a poor performer given that the image quality resulted in the foreground and background modes being difficult to separate. The primary motivation behind this was that there were times when both the normal Kneedle elbow value and the Otsu value were too low leading to a greater amount of ‘empty’ background being retained and diminishing the robustness of the high threshold approximation. This was a stopgap solution though and it cannot be reduced that the log Kneedle elbow value will be necessarily closer to the Experts but rather there is a significantly lower risk of ‘empty’ background retention and the value still remains constrained. I am slightly uncertain how to end this section. I feel that there is a distinct difference in the impact of both the Chosen and Normal Kneedle value although based on the similarity with the experts Normal appears to perform better but for cases where there is too much noise the normal incorrectly retains the chosen Kneedle is better. I could place a reference to Chapter 4 where the impact of each on the overall performance is evaluated?

3.2.3. High threshold acquisition

With the low threshold for a given sample image selected the final step is to approximate a suitable high threshold value. While the low threshold approximation was also challenging the high threshold was a more complex problem due to the image histogram being comparatively less useful. While images with clearly depicted foreground and background modes in the histogram allow for the high threshold to be selected, relative to the foreground mode, this is not the case in images where the foreground mode is unclear. This is typically due to either out-of-focus fluorescence, higher-intensity noise components not removed by denoising, unexpectedly stained biomolecules (e.g. free-floating proteins in the Cytoplasm), or image artefacts. The former two are addressed by deconvolution and denoising respectively during pre-processing although when the presence is too strong there may be interferences remaining, particularly with noise as overly aggressive denoising can degrade the image signal, yet the latter two cannot be addressed as easily with most methods although appropriate image thresholding is one of the means to address this. The reason for this is that many thresholding methods can be tuned to preserve the actual foreground objects in the image while both rejecting the remainder as background and minimizing image signal degradation. The images worked with in this research exhibit all four of the aforementioned obstructions, in different severities, to effectively separate the foreground and background modes thus a means of estimating an optimal labelling of foreground objects.

Design considerations made for foreground acquisition were important to make as this method could not rely on some known perfect output but rather had to find some result which provided the best-estimated outcome. There were three primary considerations that were made which are:

1. The system must adapt to each image to get an image-specific outcome but apply identical logic to all of them
2. Image attributes used in the system must relate to the intensities of the potential foreground voxels
3. If the system makes an error then it must be done in a manner that is the least detrimental to future analysis

To address these considerations intensity-related attributes such as the pixel/voxel intensity distribution visualised by the histogram and intensity distributions across pixel neighbourhoods were raised. The primary problem with these is that utilizing the voxel intensity distribution across the image encounters the same problems of mode separation which confronts global threshold methods while neighbourhood-based approaches where

the intensity distribution is aggregated encounter the parameter selection obstacles of local threshold methods where the size of the neighbourhood, among other parameters, directly affects performance. Since the low threshold had already been resolved a method that evaluates only the impact of the high threshold was required and led to a hypothesis that if the impact of the high threshold on the image across a threshold range could create a distribution then perhaps that could describe other image features?

3.2.4. Iterative hysteresis histogram (IHH)

This method was termed iterative hysteresis histogram (IHH), which is a distribution relating to the high threshold hysteresis application but also the name of the method by which it is acquired. This distribution is the total count of potential foreground voxels at each high threshold where the high threshold ranges from 1 intensity above the low threshold up to the maximum image intensity. An example illustrating the approach used to construct the IHH is shown in Figure 3.4 where a microscope maximum intensity projection (MIP) image is used for illustrative purposes. This culminates in a distribution that begins at one above the low threshold intensity with some maximum count X which is a non-increasing function $f(x)$ which can either remain the same or decrease as x increases (where x is the high threshold value). Due to this non-increasing nature of the function, the largest magnitude of $f(x)$ will be at the lowest possible x value (one above the low threshold) and the shape of $f(x)$ across the intensity range can describe certain image properties. This distribution across the high threshold range can describe the foreground ‘confidence’ of structures within the image as higher intensities are more likely to compose the foreground as the intensity tends towards the maximum intensity (similar to the background confidence baseline described at the end of Section 3.2.2). The reason for this confidence is that voxel intensities within the image are chiefly dependent on how in-focus the emitting fluorophores are implying that it is highly likely that the depth at which the captured fluorescence originates aligns with the depth described in the image and that there is a reasonably high density of fluorophores at that position such that the increased density of fluorescence originating from said point result in a higher combined intensity value in the image. This assumption functions as a heuristically based justification for this confidence but is not the sole reason and only functions under the assumption that the quality at which the image is captured is sufficient such that foreground and background separation can be performed using visual analysis performed by a human. As illustrated in Figure 3.5 an image with a well-separated foreground and background has a more rectangular IHH (Fig. 3.5a. & b. for MIP and IHH respectively) where the high peak in count in the highest intensities depicts that a large proportion of the structures persist from a strict high threshold and only a few additional structures appear at the lower intensities. The opposite is seen in the poorly separated foreground

and background (Fig. 3.5c. & d. for MIP and IHH respectively) where the sloping nature of the IHH reflects the dispersion of the structures across the range of intensity values increasing the difficulty in evaluating what is and is not of the foreground in the image. The challenge is that there are few algorithmic ways in which to quantify the shape of the IHH and consistent logic by which to evaluate it compared to some metric such as standard deviation or variance for example.

Using the IHH

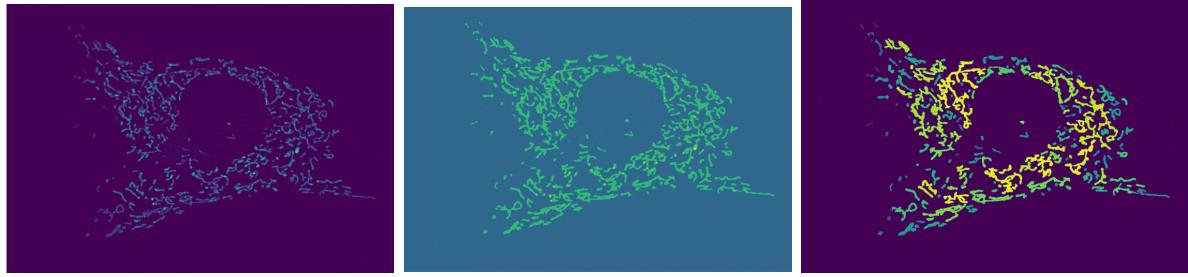
While it was realized that the IHH captured information regarding the foreground and background separation of an image in a non-increasing distribution the challenge remained in reliably exposing features to quantify without muddling the understanding of what the distribution describes. After evaluating the IHH shape and its implications regarding the distribution of potential foreground over a range of high threshold strictness it was hypothesized that desired thresholds are likely to situate around “flat” regions of the IHH. The reason that “flat” IHH regions were of interest is based on what the IHH itself describes which is the distribution of the maximal intensity of each structure in the image (after low threshold application) with structures persisting at larger high thresholds being more within focus in the image. These flat regions were typically selected at lower high threshold intensities since we want to include as many potentially viable foreground structures in future analysis but this was based on an observation made with visual analysis. Regardless, this lead to the investigation of two fronts:

1. Is a desirable threshold typically located on or adjacent to a relatively “flat” region of the IHH?
2. How can this be quantified in a manner which can be fed into an algorithm?

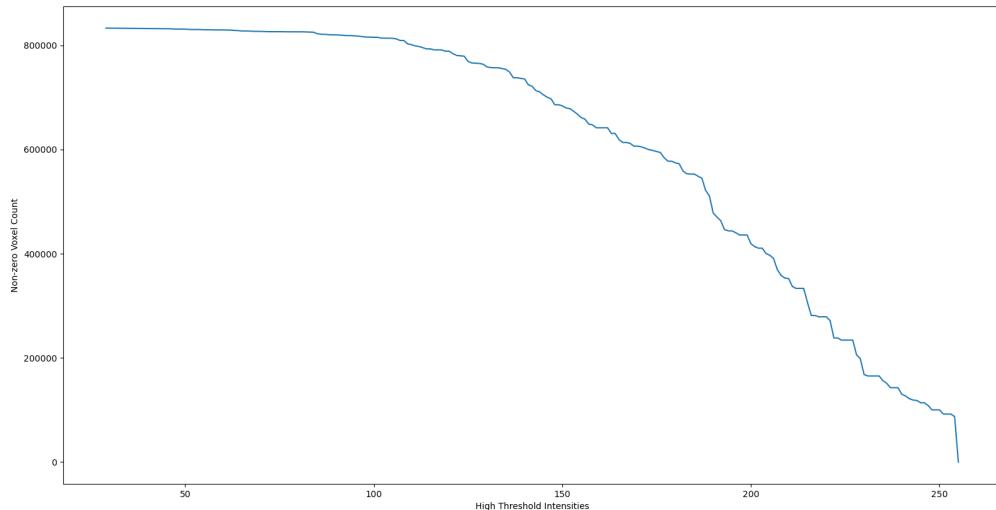
Do you think I should include a few examples showing that the relatively flat regions of the IHH correlate with being near a threshold that would render a viable binarization outcome? With the objectives made clear experimentation began where the first objective was heuristically confirmed, although partially, that although there was no definitive proof there was a near pattern where ‘viable’ high thresholds were situated near relatively flat IHH regions with a low margin of error. The real challenge came into the quantification of this with regard to detecting a ‘flat’ region (how flat must it be for consideration?) and if there are multiple flat regions how should the selection be made between them?

IHH gradient

The IHH gradient was believed to be the most effective solution to this where a gradient instead of a derivative would be used. The reason that either of these would be used is that if the slope of the IHH can be seen as representing the change in the IHH value then

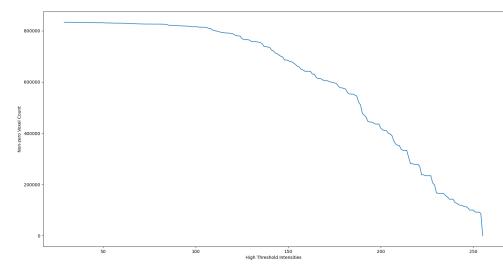
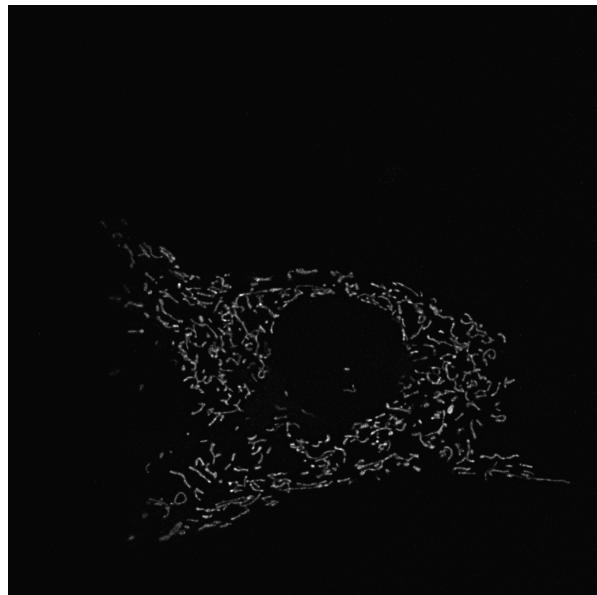


(a) Raw MIP of a microscopy image
(b) Structures labelled using unique colours for differentiation
(c) Structures coloured based on brightest contained pixel intensity

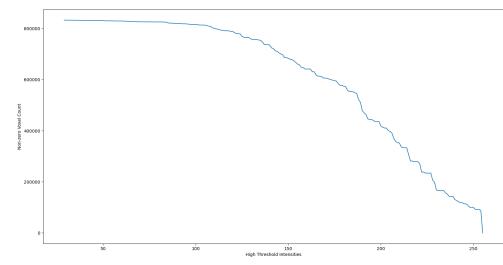
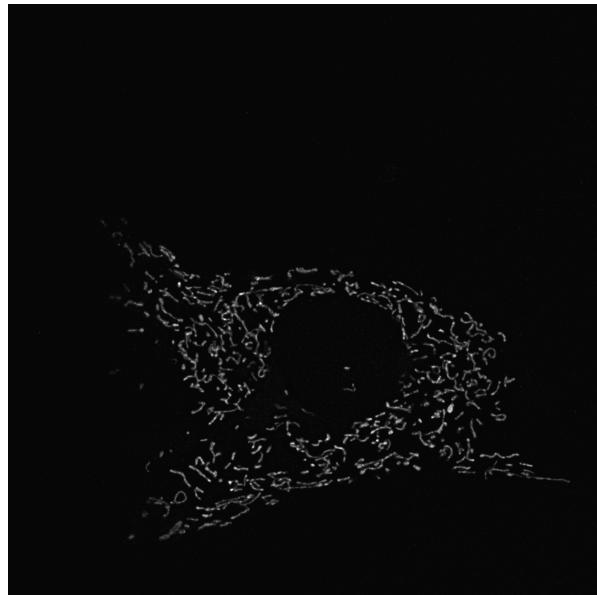


(d) IHH of image with annotations regarding the highest intensity possessed by each structure

Figure 3.4: The IHH in Figure 3.4d is supposed to have colour coding that matches the structure colours in 3.4cAn illustration of the sequence by which the IHH is determined for a microscopy cell image. The image used is a 2D maximum intensity projection of an original 3D image for illustrative purposes. This depicts how **a)** the raw image appears initially; **b)** the labelling of the independent structures in the image which are labelled by both colour and a numeric annotation; **c)** the structures are now re-coloured based on the highest pixel intensity contained within it which is also reflected in the annotated labels; **d)** lastly, the IHH that is constructed from the sum of the volumes of each structure from the lowest intensity up until their respective maximum.



(a) MIP of well separated foreground and background



(c) MIP of poorly separated foreground and background

(b) IHH of a)

Figure 3.5: Examples of two opposing IHH's from images with good and bad foreground to background separability.

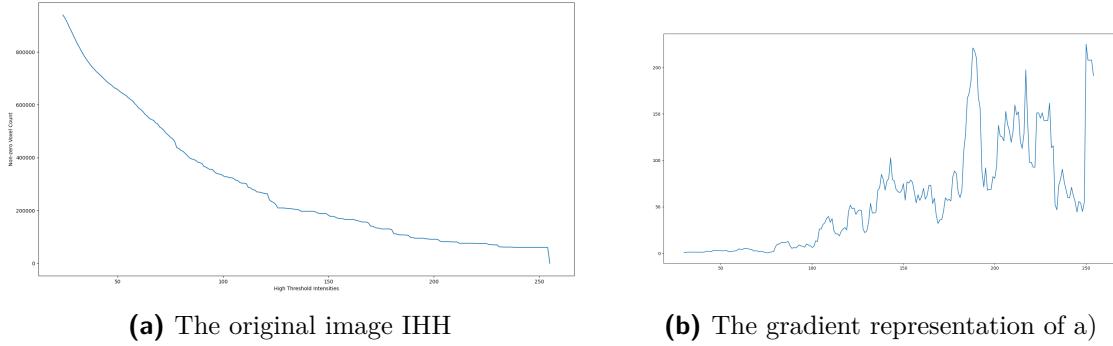


Figure 3.6: Comparison between the original IHH (a) representation and the IHH gradient (b) of Figure 3.5c

a flat IHH region would have no slope where the steepness of the slope is described by the gradient or the derivative. The gradient is what was decided to be used as the derivative requires some generalization of the IHH distribution function but said function cannot be inferred from the IHH thus the gradient evaluated against the distribution itself is more viable where the change in the IHH magnitudes ($y(n)$) between sequentially increasing intensity values ($x(n)$) can be measured as shown in Equation 3.1. An illustration of an IHH and the gradient representation ($y'(n)$) can be seen in Figure 3.6.

$$m(n) = \left| \frac{\Delta y(n)}{\Delta x(n)} \right| = \left| \frac{y(n) - y(n-1)}{x(n) - x(n-1)} \right| \text{ where } N \geq n \geq 1 \quad (3.1)$$

I am also considering having a figure where I highlight a region of the gradient representation with a high magnitude and match it to the same intensity window in the normal IHH and then colour code an MIP accordingly. The colouration in both graphs will be based on the intensity (the high threshold value that the pop out after) and this colouring will be mapped to the MIP with all other structures remaining grayscale. This would illustrate how an intensity range in flux could be composed of uncertainty. The same could be done with flat regions. Let me know your thoughts.

The problem is that the gradient that we calculated does express the change in total structure volumes at increasing high threshold strictness but it is quite discrete as it only measures this change between consecutive intensities and it cannot capture changes expressed across an extended range of intensities and it is not uncommon for trends to be expressed over a range of values based on the shape of the IHH. A second problem found from this is that the resulting gradient distribution can have a noisy appearance due to the discrete approach used to calculate the gradient representation, as can be seen in Figure 3.6b, which increases the difficulty in numerically evaluating trends in the distribution to estimate where might be a potential point for high threshold selection. An example of this would be if, hypothetically, the ideal point is the minimum value within the distribution or some minima going from the lowest point of the high threshold range

to the maximum. It could be caught in a sub-optimal local minima. These examples are stated while ignoring the potential challenges or problems those methods raise but regardless this noisy representation obscures the trends in the gradient representation that actually describe the high threshold strictness (or sensitivity) behaviour for an image and how the overall binarization responds to said behaviour.

The solution proposed was to apply a smoothing operation using a rolling average where an intensity window (of size W) is designated and centred on a gradient representation magnitude $y'(n)$ at an intensity n . This centred window is then rolled across the full range of intensities N where the average $\bar{y}'(n)$ is used as the new smoothed gradient value ($\bar{y}'(n)$) at n . This is seen in Equation 3.2 with m_1 and m_2 being the bounded upper and lower window ranges as described in Equation 3.3 where m_1 and m_2 are both $W/2$ save for when the centre of the window (n) is too close to the bottom ($n \rightarrow 0$) or end ($n \rightarrow N$) of the range. When the centre is too close the lower (m_1) or upper (m_2) bound of the window is shortened to fit in the reduced range and so too is the length of the window by which the value is divided by, given that the window contains less values to average.

$$\bar{y}'(n) = \frac{\sum_{t=n-m_1}^{n+m_2} y'(t)}{m_1 + m_2} \quad (3.2)$$

$$\text{where } m_1 = \begin{cases} 0, & \text{if } n = 0 \\ \leq W/2, & \text{if } n \leq W/2 \text{ and } m_2 = \\ w/2, & \text{if } n \geq W/2 \end{cases} \quad (3.3)$$

$$\begin{cases} 0, & \text{if } n = N \\ \leq W/2, & \text{if } n \geq N - W/2 \\ w/2, & \text{if } n \leq N - W/2 \end{cases}$$

Here I state that the window size will be 8 for use in the system development. Should I in Chapter 4 briefly explore the impact of the window size on the system results or should I append that to this subsection or chapter? Through heuristic testing, a window size of 8 was decided for use as it did not overly smooth the distribution via diminishing the trends but noisy behaviours were sufficiently suppressed. This was evaluated by visual analysis of the smoothed gradient representation of the IHH were it was observed that a smoothing window of 8 functioned optimally for a subset of samples but further testing will be performed accompanied by numerical evaluations of the impact on performance the window size has.

IHH gradient inversion

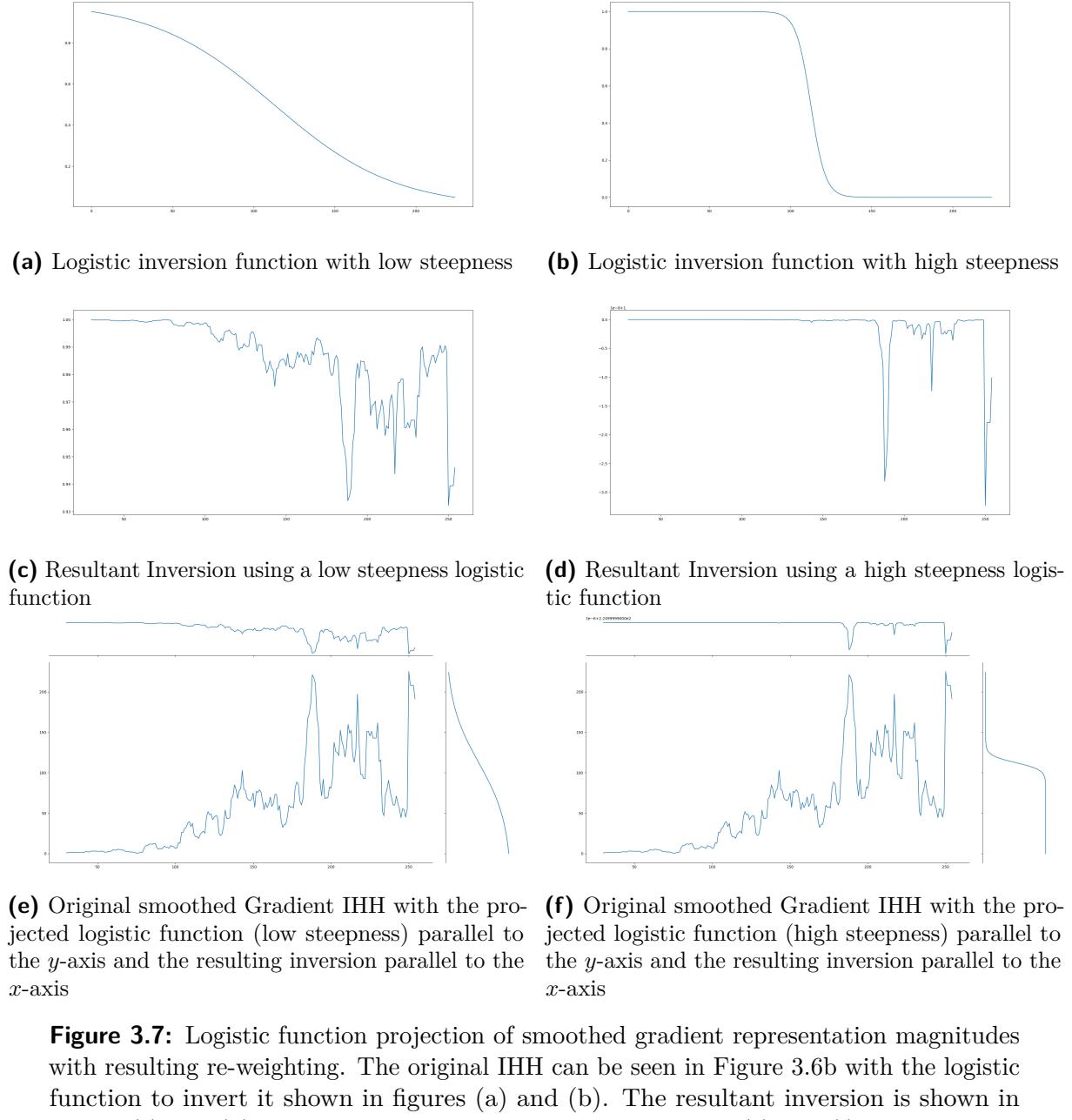
With the discrete noisy behaviour smoothed down while preserving the trends, a method to identify ideal regions was still required although the basis that flat IHH regions will inform this was still maintained. As comparatively flatter regions of the gradient IHH contribute more weight to the high threshold selection an inversion of the gradient representation magnitude is performed such that the values used in the distribution better reflect the

intended use of them in further algorithms. There were two methods by which the representation was to be inverted with the first being a sigmoid-based approach while the latter was a linear inversion. This sigmoid-based approach would employ the use of the logistic function to invert the smoothed gradient distribution to invert and scale the weight contributions of flatter regions. The motivations around this were initially that the occurrence of completely “flat” regions in the gradient distribution was ideal but also naive and there was no certainty that it would be present within any of the distributions. Due to this flatness limitation, a softer inversion method using the logistic function that could be tuned was first approached. This application projected the gradient representation magnitudes (y -axis) along the logistic function dependent variable (x -axis) with the resulting new magnitudes being based on the steepness parameter of the logistic function. What is consistent is that the lowest gradient magnitude will become 1 after re-weighting while the maximum gradient magnitude will become 0 but a steepness parameter with a resulting *S*-shaped curve, as shown in Figure 3.7, will result in a clustering of n of the bottom and top of the gradient magnitude. This results in a new distribution that essentially states “values close to the peak are essentially all experiencing too much change while values within a low enough margin of magnitude of change are approximately the same” reducing nuance but also emphasizing broad trends of the extreme gradient representation values. This introduced a new tunable parameter and although a value of 6 was settled on, for the purposes of developing subsequent features, further rigorous testing would be required to fully discern the impact of the logistic function shape, based on the steepness parameter, on the behaviour of the high threshold determination. Due to this, a more clear linear inversion was introduced in Equation 3.4 where the magnitudes of the smoothed gradient distribution ($\bar{y}'(n)$) are subtracted from and divided by the maximum of the distribution ($\bar{Y}'_{\text{maximum}}$) for all intensity values (n) resulting in a normalized and flipped smoothed gradient distribution ($\bar{y}'_{\text{new}}(n)$). The benefit of this is that the inversion will perform relatively consistently for gradient distributions regardless of the source image and that there will be no added tunable parameters while still achieving the desired outcome of emphasizing lower values in smoothed gradient distribution numerically.

$$\bar{y}'_{\text{new}}(n) = \frac{\bar{Y}'_{\text{maximum}} - \bar{y}'(n)}{\bar{Y}'_{\text{maximum}}} \quad (3.4)$$

3.2.5. Selecting a high threshold

With the smoothed gradient IHH distribution inverted such that values of lower change in IHH values are emphasized a high threshold value needs to be selected. The key challenge regarding high threshold selection using the gradient distribution is the criteria used to determine what region is best fit as there can be multiple regions of varied degrees of flatness in the IHH which are expressed in the gradient representation with the result



needing to be a single value. There were many attempts to determine a value similar to a mean of the intensities based on the inverted gradient distribution as weights and after experimentation, a method to achieve this was produced as shown in Equation 3.5. In this function, there are N sequential intensity values $x(n)$ that make up the x -axis of the inverted gradient distribution with the inverted gradient values $\bar{y}'_{\text{new}}(n)$ is used to weight the intensity value for each corresponding intensity n and averaged by the total count of values composing the distribution x -axis.

$$C = \frac{\sum_0^N \bar{y}'_{\text{new}}(n) \circ x(n)}{N} \quad (3.5)$$

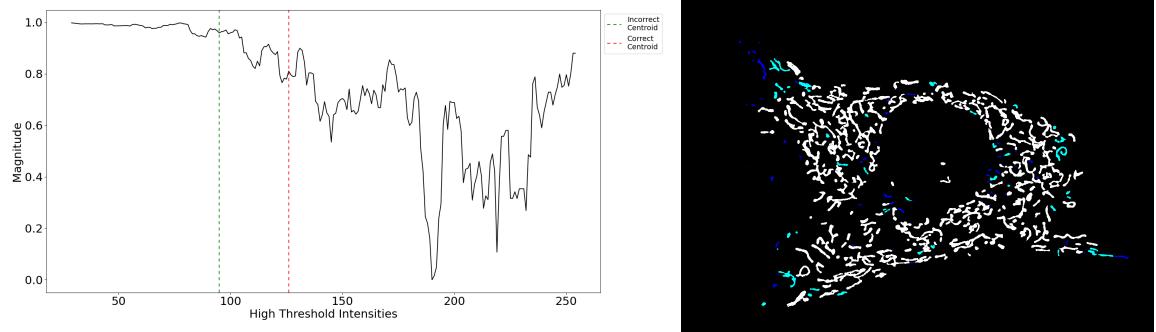
The result of this is a value that acts similar to a centre of mass (C) for an object given that the inverted gradient magnitudes represent the mass contributions of each intensity value in the distribution thus some ‘best fit’ result is achieved that takes all values across the distribution into account. After it was confirmed that this resulted in an intensity value calculated based on the gradient magnitudes the spectral centroid was discovered. Due to the greater context of how the algorithm should behave, based on Eq. 3.5, the centroid calculation was discovered where a crucial difference is that the value is divided by the sum of the inverted gradient weights across the intensity range ($\sum_0^N \bar{y}'_{\text{new}}(n)$) as seen in Equation 3.6. The added benefit of this approach is that the divisor is the sum of intensity weights ($\sum_0^N \bar{y}'_{\text{new}}(n)$) resulting in a more accurate intensity centroid (C_x) using a method that is more mathematically robust which is more applicable to this use case.

$$C_x = \frac{\sum_0^N \bar{y}'_{\text{new}} \circ x(n)}{\sum_0^N \bar{y}'_{\text{new}}(n)} \quad (3.6)$$

This final intensity centroid considers the distribution of ‘flatness’ throughout the IHH which is expressed as values tending towards 1 in the inverted gradient representation magnitude and selects a best-fit value but it was found that at times that the higher contrast resulted in very flat high intensities that contained far less object information than desired and the general approach is that if the system outcome is going to err then rather it err more lenient and preserve more voxels as foreground as these can be addressed with further processing but any voxels labelled as background may be wholly excluded from any further processing or evaluation risking information being lost.

3.2.6. Applying biases to the centroid result

As it was found that at times the centroid result returned a high threshold that was too strict biases were developed to balance this. The reason for this is speculated to be due to the reduction in the context of the objects depicted in the image due to the quantification of image properties. Primarily is this IHH distribution which describes the response



(a) Distribution showing the inverted gradient representation with the two centroid outcomes annotated by vertical, dashed lines

(b) The effect of these outcomes on the Hysteresis thresholding results are displayed using a MIP image. Blue describes the voxels above the low threshold; green describes all structures containing a voxel intensity above the incorrect centroid (3.5); and red describes all structures containing a voxel intensity above the correct centroid (3.6)

Figure 3.8: A comparison between the initial centroid method (incorrect centroid) and the robust method that is used (correct centroid). The difference in the centroid value outcome is illustrated on an inverted gradient distribution line plot where dashed vertical lines are used to illustrate the centroids while the difference in the impact on the Hysteresis thresholding results are displayed using a MIP image. The image is colour-coded to differentiate and aid in comparison between the threshold outcomes

behaviour of the image to varied high threshold strictness but does not provide the nuance to know when it is relegating a large number of image structures to the background as opposed to the foreground nor can it consider properties such as the shape or volume of the individual potential foreground structures in an effective manner. To address this two biases were applied.

IHH total volume bias

So as to add some bias to the high threshold determined using the centroid a second weighting metric was added which was based on the normalized IHH distribution for the image. This is applied to the numerator and denominator within the summation as shown in Equation 3.7. By doing so we introduce the context of the distribution of foreground information which the IHH describes as the y -axis representing the number of foreground voxels for each x -axis value representing the high threshold (therefore the strictness of the high threshold) resulting in an IHH weighting $v(n)$.

$$C_x = \frac{\sum_0^{N-1} v(n) \circ \bar{y}'_{\text{new}}(n) \circ x(n)}{\sum_0^N \bar{y}'_{\text{new}}(n) \circ v(n)} \quad (3.7)$$

The effect of this can be seen in Figure 3.9 where the difference between the centroid determined using the inverted gradient representation of the IHH can be greatly skewed

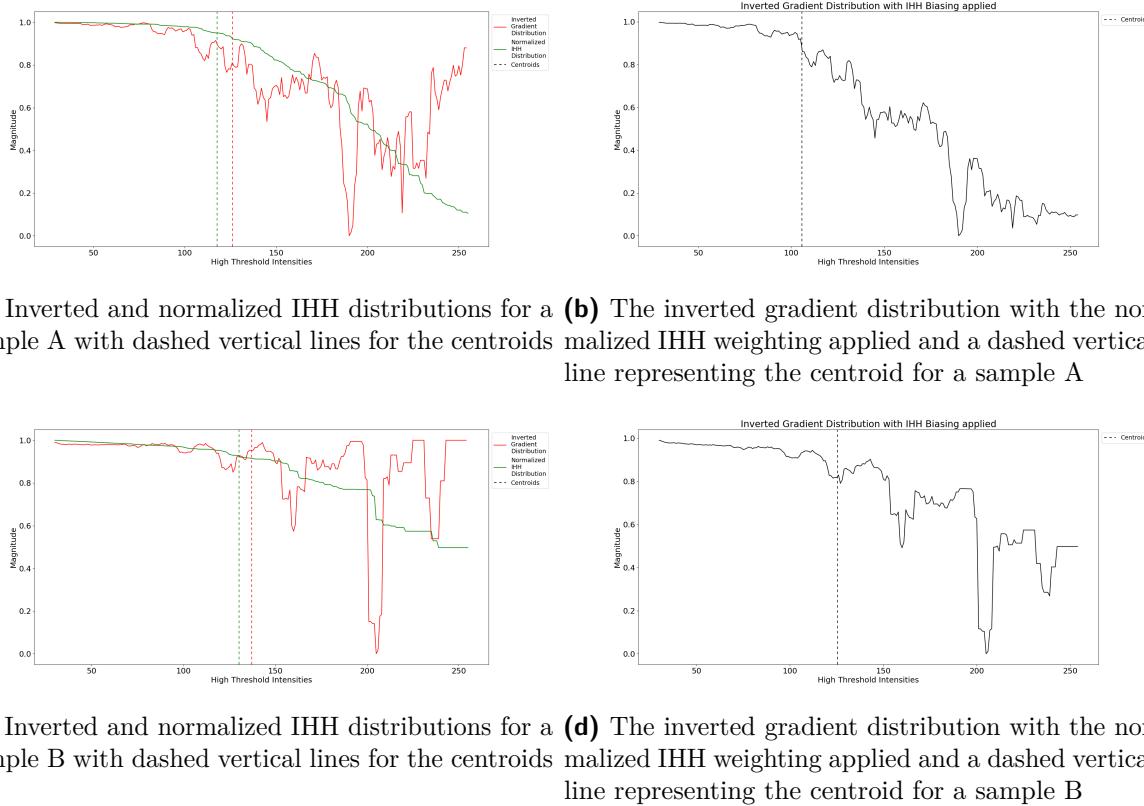


Figure 3.9: A showcase comparing the influences of the inverted gradient and normalized IHH distributions on the centroids separately and when combined. In the figures illustrating both the separate distributions in red and green (Fig. 3.9a & 3.9c) and the new distribution produced by multiplying said distributions and determining a new centroid (Fig. 3.9b & 3.9d). By comparing the change in the determined centroids when the distributions are combined for Samples A (Fig. 3.9a & 3.9b) and B (Fig. 3.9c & 3.9d) it can be observed that the new centroid determined in the former sample is at a lower value than either of the separate centroids while for the latter sample the new centroid is far closer to the separate centroids of the original distributions. This highlights the impact the shape of the distribution has on the calculated centroid value.

by the application of the normalized IHH distribution. In this figure, the distributions (separated and combined) for two sample images deviate further from the separate centroids the faster the IHH distribution decays relative to the increasing high threshold intensity along the x-axis. This means that although the centroid of two IHH distributions can be close between two samples their influence on an identical inverted gradient distribution is less predictable. The benefit of this is that the strength of this bias is dependent on the distribution of the potential foreground structure information across the high threshold intensity span. Better separated foreground structures with a higher contrast will typically result in distinctly rectangular distributions or ones that decay in wider steps and are more robust to fluctuations in the high threshold value.

The reason for this is the leniency towards false positives over false negatives for foreground selection as erroneously labelled foreground can be removed with further

processing yet the background erroneously labelled background is typically lost to further analysis. Since the high threshold selection is expected to be a best-fit approximation, tending towards a precise value, it is supposed that there will be some level of error and despite a goal being to minimize this error it was decided that it would rather err to a lower high threshold value as opposed to a higher value. A fixed shift in the approximated towards the lower intensity values was deemed inapplicable as it is naive to the imaging conditions and uniqueness presented by each sample image and no single value would work, this shift would need to adapt to the image. This is the reason that the IHH was chosen since tending the high threshold range towards the lower end typically results in a plateau or increase in IHH magnitude implying an increase in potential foreground information. As we tend to a lower high threshold value the risk of the potential foreground voxels being that of noise increases relatively speaking thus this is used in conjunction with the inverted gradient representation of the IHH with the intention that each compensates for the other. Figure 3.10 illustrates the effect of these different centroid results on the hysteresis thresholding outcome. The distributions shown in Figure 3.9 (a) and (b) are for this same sample image and in the red and green centroids shown in (a) match to the red and green colour channels shown in Figure 3.10c (red and green overlap appears as yellow). Despite the green channel threshold (based on the IHH distribution centroid) being lower than that of the red channel the final centroid calculated from the IHH weighted inverted gradient distribution (Fig. 3.9b) contains more structures than even the green channel as the centroid value is of an even lower value. This will provide a bias which will provide a lower high threshold intensity than the inverted gradient distribution along with the magnitude of this bias being image dependent meaning that the bias adapts to the sample image as opposed to some fixed value that needs to be entered for each sample.

Shrinking window bias

A second kind of bias that was implemented was the sliding window bias which affects the distribution for which the spectral centroid is calculated. This base premise is that within the intensity range above the low threshold, the lowest intensity value is supposed to be to the left while the highest intensity is to the right where the window initially spans from the furthest left with the upper range of the window (I) iterating across the range $I = [x_{min} + 1, X]$. This range begins at $x_{min} + 1$ which is 1 value greater than the smallest value within the original range after the low threshold application thus the lowest intensity value is not necessarily 0. This is described in Equation 3.8, and illustrated in Figure 3.11, which shows how the centroid (C_I) would be calculated for a window with an intensity upper bound of I where in this figure there are three windows shown, with the respective centroids, with varied window sizes to illustrate the effect on the centroid value a narrower distribution has. This can be applied in conjunction with the IHH total volume bias with

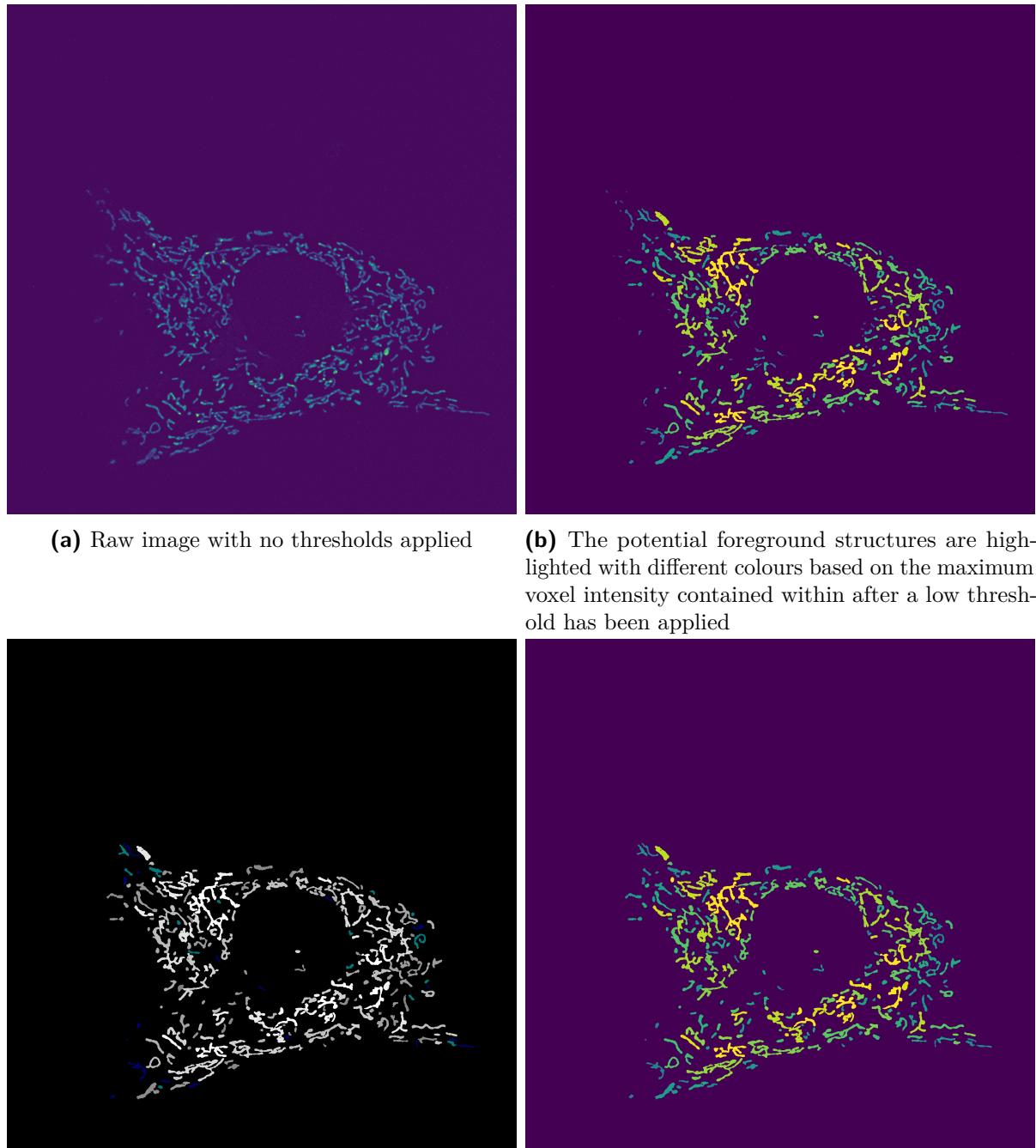


Figure 3.10: The images shown are MIP images of a sample image where the differences of the different centroids, determined using the relevant distributions, are illustrated. **I feel the caption of the RGB subfigure is a bit too long, do you have any suggestions? Perhaps I should summarize the subfigure captions and move the descriptions to the main caption?**

The images shown are MIP images of a sample image where the differences of the different centroids, determined using the relevant distributions, are illustrated. **I feel the caption of the RGB subfigure is a bit too long, do you have any suggestions? Perhaps I should summarize the subfigure captions and move the descriptions to the main caption?**

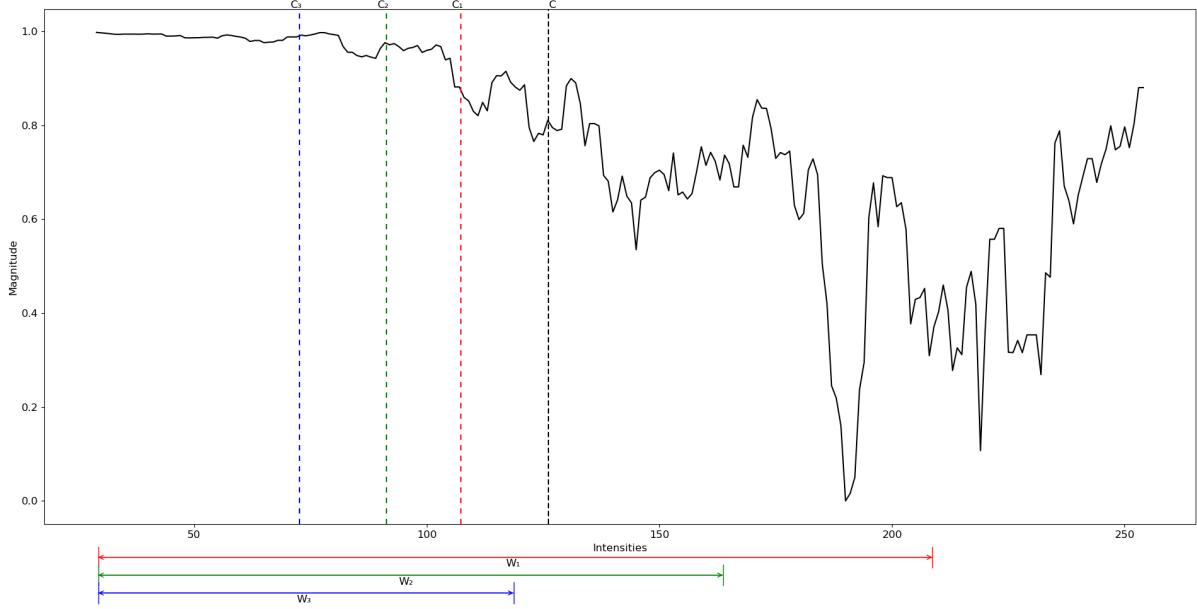


Figure 3.11: An example of how the window weighting approaches are applied to the distributions where in this example there are three bounded windows shown. The spans of these windows are annotated below the x-axis using coloured, horizontal lines with an accompanied label where window 1 (W_1) is shown with red, window 2 (W_2) is shown with green, and window 3 (W_3) is shown with blue. Each of these windows also has a respective centroid calculated for the distribution that falls within the window range which is annotated by a dashed vertical line with the appropriate colouring to match the windows and a label C_n where n is for the respective window. There is a black dashed vertical line with the label C which represents the distribution centroid for the whole distribution without the windows

\bar{y}'_{new} substituted for $v_w(i) \circ \bar{y}'_{\text{new}}(i)$.

$$C_I(d) = \frac{\sum_{i=x_{\min}}^d \bar{y}'_{\text{new}}(i) \circ x(i)}{\sum_{i=x_{\min}}^d \bar{y}'_{\text{new}}(i)} \text{ where } d \leq X \text{ and } d \subset X \quad (3.8)$$

As mentioned, this function will be performed for each iteration of I resulting in numerous windowed centroids $C_I(I)$ across the window width range which will tend toward the lower intensity as the window width shrinks and contain less higher intensity components but the degree to which they veer to the lower intensities is also subject to the distribution contained within the window. What is crucial is the manner by which these windowed centroids are consolidated to a singular value to act as the high threshold. Taking the straight average was initially considered but due to the windows reducing the total context provided by the image distributions (IHH or inverted gradient), it was believed that the contributions of each window could not be received equally towards the final value. It is for this reason that two consolidation approaches were developed with the goal of effectively weighting the contribution of each window centroid to the final high threshold value settled on. The consolidated window weighted centroid (C_W) is shown in Equation 3.9 where the weighted average of the window centroids (C_I) using the window weights (w) but

these weights are dependent on the weighting functions described in the consolidation approaches.

$$C_W = \frac{\sum_{t=x_{min}}^X C_I(t) \times w(t)}{\sum_{t=x_{min}}^X w(t)} \quad (3.9)$$

The window weighting consolidation approaches are the **window width weighting** approach and the **window mass weighing** approach.

Window width weighting is an approach where the weighting coefficient (w) that is applied for the consolidation is based on the current size of the window relative to the total width of the distribution. This is shown in Equation 3.10 where t is the current upper bound of the window and serves as the dependent variable, $w(t)$ is the window weighting to be used for consolidation, X is the maximum intensity of distribution range, and x_{min} is the minimum intensity of the distribution range as all intensities less than this have been removed by the low threshold.

$$w(t) = \frac{t - x_{min}}{X - x_{min}} \text{ where } X \geq t > x_{min} \quad (3.10)$$

This weighting approach provides a linearly decreasing weight function proportional to the loss of information provided by the distribution as the window shrinks. This results in a predictable weighting outcome for each window but it is believed that it is still somewhat naive to the information loss.

Window mass weighting is the approach where the weighting of each window centroid is weighed by the contained mass within the window relative to the total mass across the distribution. In this context, the "mass" can be derived by the total area under the distribution curve and the window mass is the area within the window bounds. Both of these areas are calculated by the sum of magnitudes (\bar{y}'_{new}) within the respective bounds (t and X) and the division of the windowed area by the total area results in a window mass ratio as shown in Equation 3.11 which is used as the window consolidation weighting ($w(t)$).

$$w(t) = \frac{\sum_{i=x_{min}}^t \bar{y}'_{new}(i)}{\sum_{i=x_{min}}^X \bar{y}'_{new}(i)} \quad (3.11)$$

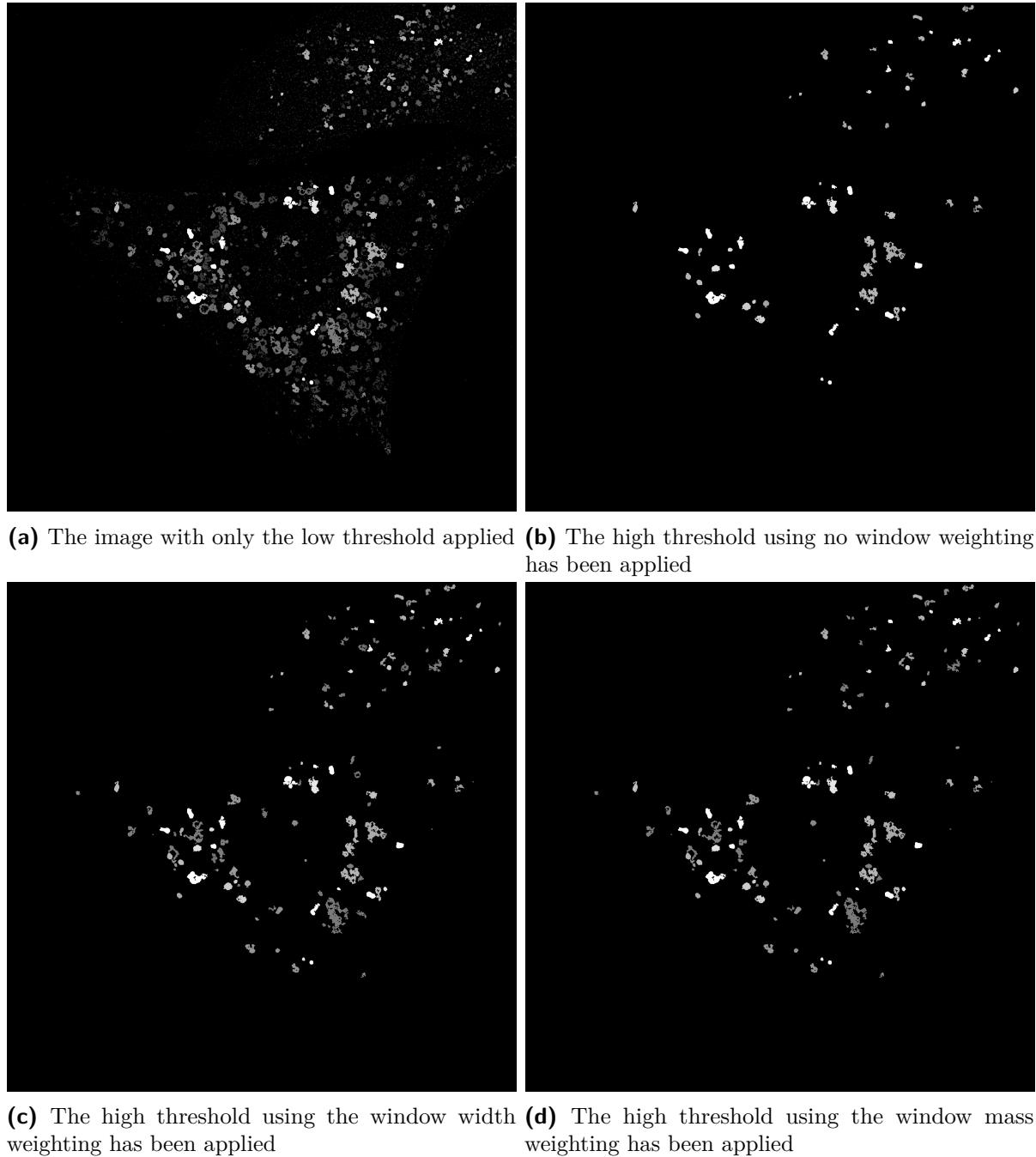
With the added consideration of the "mass" contained within the window relative to the whole distribution, it is believed that this should provide a better weight for consolidation. The reason for this is that it is known that the greater the magnitude of the distribution at specific regions the more it shifts the centroid towards it and the greater the cluster of magnitude (large magnitude over a large area) the greater the contributor. Suppose that there is a high-intensity region of low magnitude such that it contributes only 10% collectively to the centroid outcome, given this the loss of said region by windowing will be of little impact as even when included it was a negligible contributor. It is under this

reasoning that the weighting is derived where the weighting of a window is relative to the measure of influence said region of the distribution has over the centroid relative to the entire centroid. The caveat is that the window weighting has a non-linearly decreasing weighting function $w(t)$ but is believed to be less naive than the window width weighting.

Overall the impact of the window weighting approaches on the high threshold acquisition is displayed in Figure 3.12 where the contrast between no window methods being applied to either method being applied can be seen as the quantity of non-zero structures in Fig. 3.12b (high threshold is 155) is far less than either Fig. 3.12c or Fig. 3.12d (high threshold is 112 and 117 respectively). What is notable for this example is that the difference in high threshold results for the **window width** and **window mass** weightings are not that different but this is not necessarily the normality as the window weighting approaches are arguably more sensitive to the change in the shape of the distribution across the intensity range as it is a consolidation of numerous centroids thus compounding this "sensitivity" to the shape.

Weighting bias summary

From the exploration of both the IHH weighting bias and the windowing weighting bias approaches detailed thus far it has been observed that the impact the IHH weighting has on the final distribution shape (when combined) can greatly change the final result acquired. The window weighting bias increases the sensitivity to the distribution shape of the lower intensities of the distribution as the shrinking iterating window creates an emphasis on the distribution shape that is maximal at the lowest distribution intensity and diminishes as the intensity increases. Whether this diminishment is linear or non-linear depends on whether the **window width** weighting or **window mass** weighting approaches are used respectively. A showcase of this can be seen in Table 3.1 where the centroids to be used as the high thresholds for three random sample images are shown. In this table, the centroid values derived without any window weighting, with the window width weighting, and with the window mass weightings are displayed including the added involvement of the IHH bias. From the results shown it is observed that the biggest difference is the application of no window weighting versus some window weighting (as mentioned previously), as well as the difference between the two window weighting approaches not being terribly different although depending on the image, it may only be a single digit shift in value to determine if a structure is labelled as foreground. The biggest difference noted is when the IHH bias is applied but it is speculated that this is due to the IHH distribution being a non-increasing function and thus it is more likely that a greater proportion of the distribution is larger towards the lower intensities and the more aggressive this is the greater the impact the application has on the final centroid result. The expectation is that flatter IHH distributions will have a lesser impact on the centroid results while



(c) The high threshold using the window width weighting has been applied **(d)** The high threshold using the window mass weighting has been applied

Figure 3.12: This showcase shows the impact of the window weighting approaches on the determined centroids to be used for the high threshold. The structures shown are highlighted with a brightness proportional to the brightest voxel contained within and are flattened to a 2D MIP image from a 3D image. The original image with only the low threshold applied is shown in (a) with high threshold applications shown in (b), (c), and (d). No centroid window weighting biases are applied in (b) with a high threshold of 155 being derived; the window width weighting bias is applied in (c) with a high threshold of 112 being derived; and the window mass weighting bias is applied in (d) with a high threshold of 117

| Sample | IHH Bias | No Window Weighting | Window Width Weighting | Window Mass Weighting |
|--------|----------|---------------------|------------------------|-----------------------|
| A | False | 155 | 112 | 117 |
| | True | 107 | 87 | 85 |
| B | False | 126 | 95 | 93 |
| | True | 105 | 89 | 87 |
| C | False | 137 | 101 | 101 |
| | True | 125 | 97 | 96 |

Table 3.1: The different results of the weighting approaches (no window weighting, window width weighting, window mass weighting) on the final centroid result after consolidation for three random samples is shown. The effect of the IHH bias on the centroid outcome is included as it can have a drastic influence on the behaviour of the window weighting approaches

steeper (or more heavily decreasing distributions) will have a greater impact. This results in a redistribution of the "mass" of the distribution and can lead to low-intensity heavy windows impacting window mass weighting more so than window width weighting.

Chapter Conclusions

This chapter has detailed both the methodology behind the preparation of the data used in the development (and testing) of AHT, and the inner workings of the AHT itself.

Dataset preparation is discussed with a focus on the sample preparation prior to microscopy imaging, the microscopy imaging conditions (microscope model, configurations, etc.), and the preprocessing involved with the sample images prior to thresholding which AHT performs.

System design is detailed as the journey to select the thresholding method and motivations behind it, which was Hysteresis thresholding, and how the parameters for said method are to be approximated without any external human involvement. The parameter approximation was the primary focus with the low threshold and high threshold calculation being the two main components. The low threshold determination was the simpler component of the two as multiple methods were compared (including the Kneedle algorithm) with final method being chosen based on quantitative measures and qualitative evaluations of the behaviour of said methods.

The high threshold was more involved with a novel application of Hysteresis thresholding applied to generate a new distribution from which to analyze the thresholding response of the image in terms of the sensitivity of the total volume of structures to the high threshold

selection. Finally, two types of dynamic biasing methods are explored which can tune the high threshold results in a manner that is adaptive to the unique sample image. With the fundamental inner workings of AHT described the proceeding chapters will concern the performance of the AHT relative to other established methods, observations regarding the conditions favourable to AHT, and ending with a conclusion consolidating all of the findings and where that places this research with regard to future interest.

Chapter 4

System performance and observations

With the design and development of the AHT system covered in Chapter 3, the evaluation of this new method against established thresholding methods needs to be performed. This will entail a visual analysis where both the AHT method and established automatic thresholding methods, global and local, will be compared in binarization outcome against a threshold baseline. This baseline is generated through human-involved thresholding where thresholding parameters are manually tuned to achieve what is subjectively viewed as ‘suitable’ binarization. Both this baseline and the aforementioned thresholding methods will be subjectively evaluated by how they binarize the foreground structures perceived in the images. Additionally, synthetic data will be created to determine how robust the AHT and established thresholding methods are against noise. Lastly, observations made regarding AHT behaviors and thresholding outcomes will be discussed with particular focus on the different biases that can be applied. All images discussed in this chapter have been pre-processed using the pipeline described in Section 3.1.3.

4.1. Shortlisting viable thresholding methods for the evaluation

For this evaluation, the image processing package Fiji was used as it provides two built-in plugins for automated global and local thresholding methods called *Auto Threshold* and *Auto Local Threshold* respectively. These plug-ins provide access to several different thresholding methods, built into the plug-in, where either all methods or individually selected methods can be applied to an image through an accessible interface. Determination of the appropriate methods for an image and the tuning of parameters, for the local thresholding methods, still need to be performed by a human but these plug-ins allow for rapid evaluation of the method, and associated parameter selections, for each image. These threshold methods were applied to a small set of sample images for the this shortlisting process with the raw images, prior to any thresholding, shown in Figure 4.1.

A primary focus of evaluating the binarization outcomes, for the thresholding methods explored, are the loss, or exclusion, of foreground structures due to strict thresholding, and the increased inclusion of potentially background voxels in the foreground due to

lenient thresholding. Whether this binarization outcome is detrimental or beneficial to succeeding analyses can be approximated by a human with prior knowledge of the expected brightness, shape and size of the imaged structures. There are mild and major cases of this where mild cases depend on the bias of the human evaluator while major cases present such an aberrant outcome that it is conclusively a detriment to succeeding analysis using said binarization. Examples of these mild and major cases are shown in Figure 4.2 with reference to the original intensity image to illustrate the contrast in outcomes between mild and major cases.

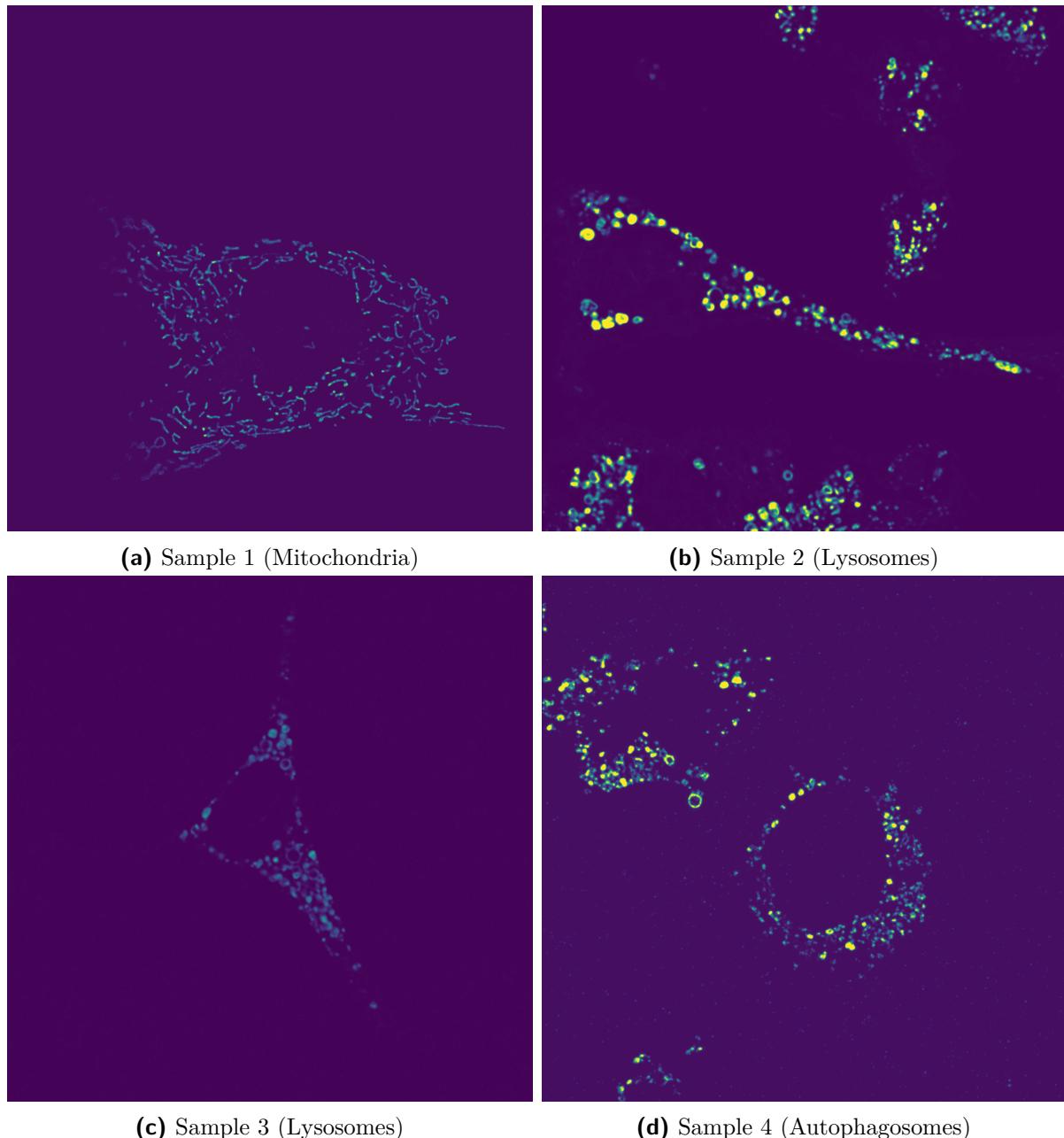


Figure 4.1: Set of samples used to shortlist the threshold methods. The centre slice of the images, rounded down where necessary, of the images is shown.

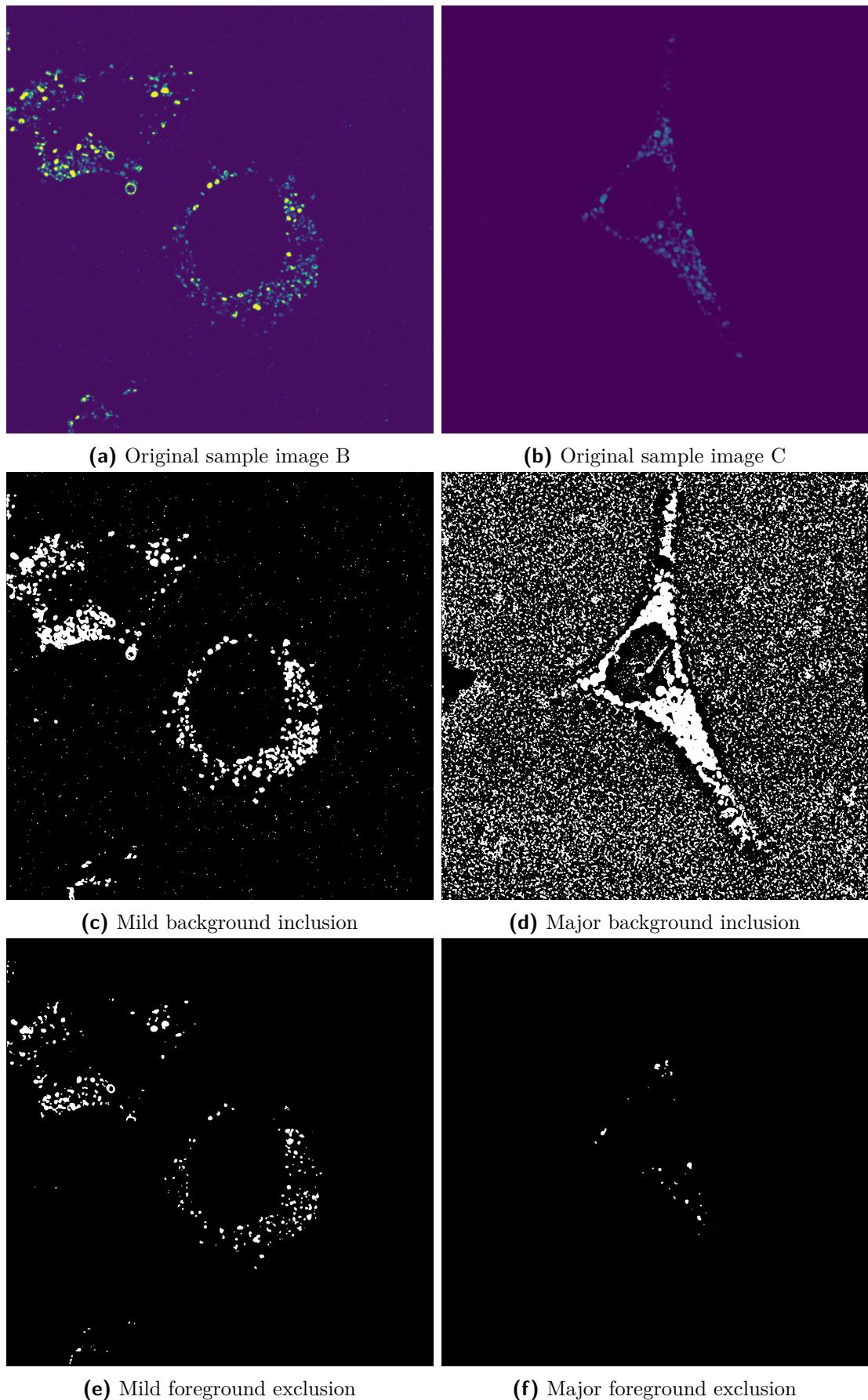


Figure 4.2: Examples of both mild and major cases of background inclusion and of foreground exclusion. The sample images presented are Samples 2 and 3, shown in 4.1, with (b), (d), and (f) representing the former and (a), (c), and (e) representing the latter.

4.1.1. Global threshold shortlisting

A total of 16 global threshold methods, provided by the *Auto Threshold* plugin, were tested across the shortlisting sample set:

- Huang
- Huang2
- Intermodes
- IsoData
- Li
- MaxEntropy
- Mean
- MinError(I)
- Minimum
- Moments
- Otsu
- Percentile
- RenyiEntropy
- Shanbhag
- Triangle
- Yen

Findings: Of the global thresholding methods explored, the Huang2, IsoData, Li, MaxEntropy, Moments, Otsu, RenyiEntropy, Triangle, and Yen methods were deemed viable for further evaluation due to the presence of mostly mild foreground exclusion and background inclusion and it is nuanced whether mild cases are detrimental and, if so, to what degree. While evaluating the severity of these is subjective, save for the extreme cases, it was observed that the viable methods only present at most two mild cases per method while only the Triangle method presents a singular instance of more severe background inclusion although this is still not comparable to the major case illustrated in Figure 4.2 (d). While this does not imply that the performance of these methods will be the same for evaluations involving different image data sets, it does flag methods expressing major foreground exclusion and background inclusion behaviours. Major occurrences of background inclusion can be seen in methods Huang (Figure ??), Mean (Figure ??), MinError(I) (Figure ??), and Percentile (Figure ??) while major occurrences of foreground exclusion can be seen in methods Intermodes (Figure ??), Minimum (Figure ??), and Shanbhag (Figure ??). Due to this, these methods have been deemed not worth further exploration in this study due to this poor performance and the showcasing of a singular case of either major foreground exclusion or background inclusion, dependent on the method, is shown in Figure 4.3 for the methods not deemed worth further exploration.

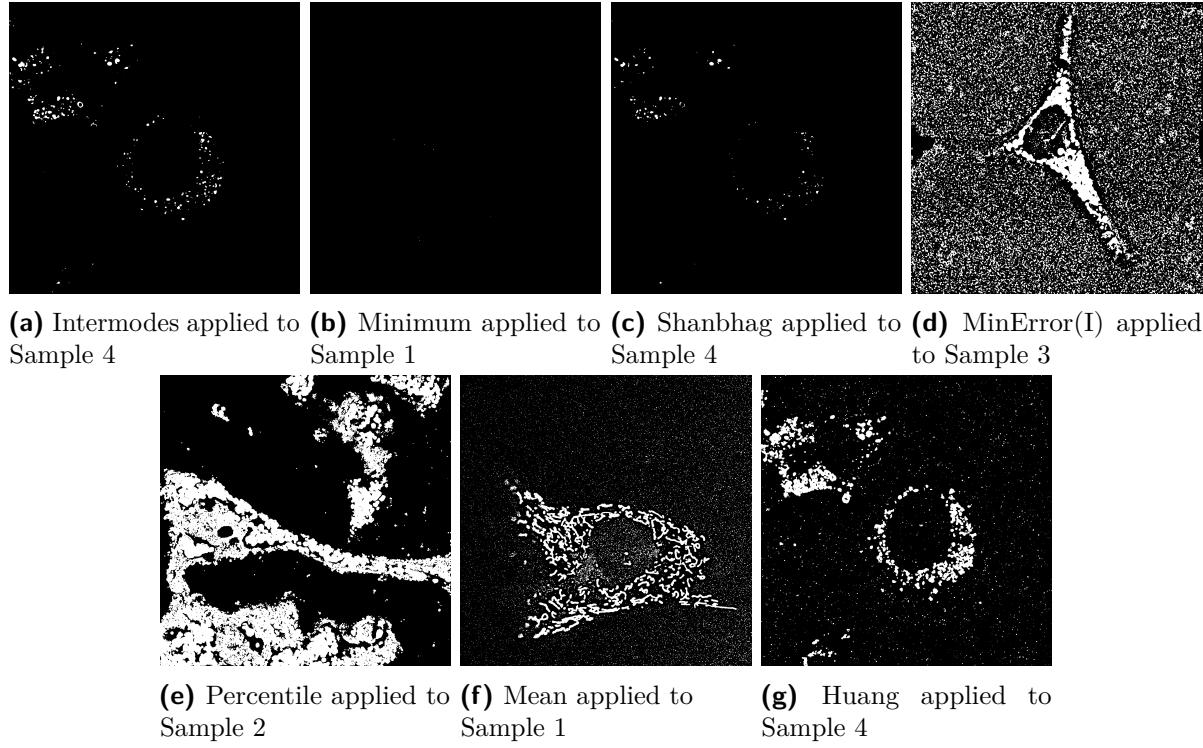


Figure 4.3: Showcase of non-viable global threshold outcomes for the shortlisting data set. The samples showcased are chosen to represent the shortcoming of these methods for the real sample images shown in Figure 4.1.

4.1.2. Local threshold shortlisting

A total of 9 local thresholding methods, provided by the “Auto Local Threshold” plug-in, were evaluated for viable application to the sample images. These methods are:

- Bernsen
- Contrast
- Mean
- Median
- MidGrey
- Niblack
- Otsu
- Phansalkar
- Sauvola

Compared to global thresholding methods, the shortlisting of local thresholding methods is a bit more involved due to the tuning of parameters being required for fair evaluation of these methods. All local thresholding methods binarize each pixel based on a pixel-specific threshold where this threshold is determined using a localized region, or window, surrounding said pixel. The size of this neighbourhood can greatly influence the binarization outcome thus selection of an appropriate window size for each image is crucial for every method. Additionally, all local thresholding methods being assessed, save for Contrast and Otsu, provide additional parameters that can be tuned to further influence the binarization. For rapid application of these local thresholds it is not uncommon to approximate parameters that best generalize across the sample set based on the binarization performance for a smaller subset. For this reason, the shortlisting will evaluate the local

thresholding methods with performance considered with respect to the parameter selections, per method, that generalize best across all of the shortlisting samples.

Findings: From observations of the local thresholding outcomes, the methods deemed viable for further use in evaluations were Bernsen, Mean, Median, and MidGrey. Furthermore, the optimal window size was selected as 15 as Bernsen and MidGrey both exhibit a growing risk of over-thresholding at higher window sizes, as seen in Figure 4.4, while Mean and Median still provide sufficient threshold outcomes. The effects of parameter tuning for these methods are shown in Figure 4.5 with RGB images with the different colour channels representing the thresholding outcome under different secondary parameters while a second row depicts the parameter selection deemed “optimal” with better clarity than the RGB images. For Bernsen, the colour channels represent the contrast threshold parameter values which are 10 for Red, 15 for Green, and 20 for Blue. The default value provided by Fiji was 15 thus the variations in the parameter value were centred around this. It was observed that the Red channel captured too much noise in the image (minor under-thresholding) while the Blue channel (only seen as white due to channel overlap) captured a few fewer structures than the Green and Red channels respectively the Green channel (15) was deemed as the optimal middle ground. For the Mean, Median, and MidGrey threshold methods, the colour channels represent the subtract constant parameter values which are 0 for Red, -5 for Green, and 5 for Blue. The colour channel to parameter mappings was selected to visualise with the best clarity as the Blue channel depicts gross under-thresholding, with the background almost completely blue, which the Red channel overlaps with as magenta. It was observed that for all of these methods both the Red and Blue channels depict under-thresholding of minor and major severities respectively while the Green channel (-5) provides the best outcome generalized across the shortlisting samples.

The outcomes for all shortlisting samples at the varied window size parameter and secondary parameter tunings can be viewed in Appendix A.2 for all methods (both those deemed viable and non-viable).

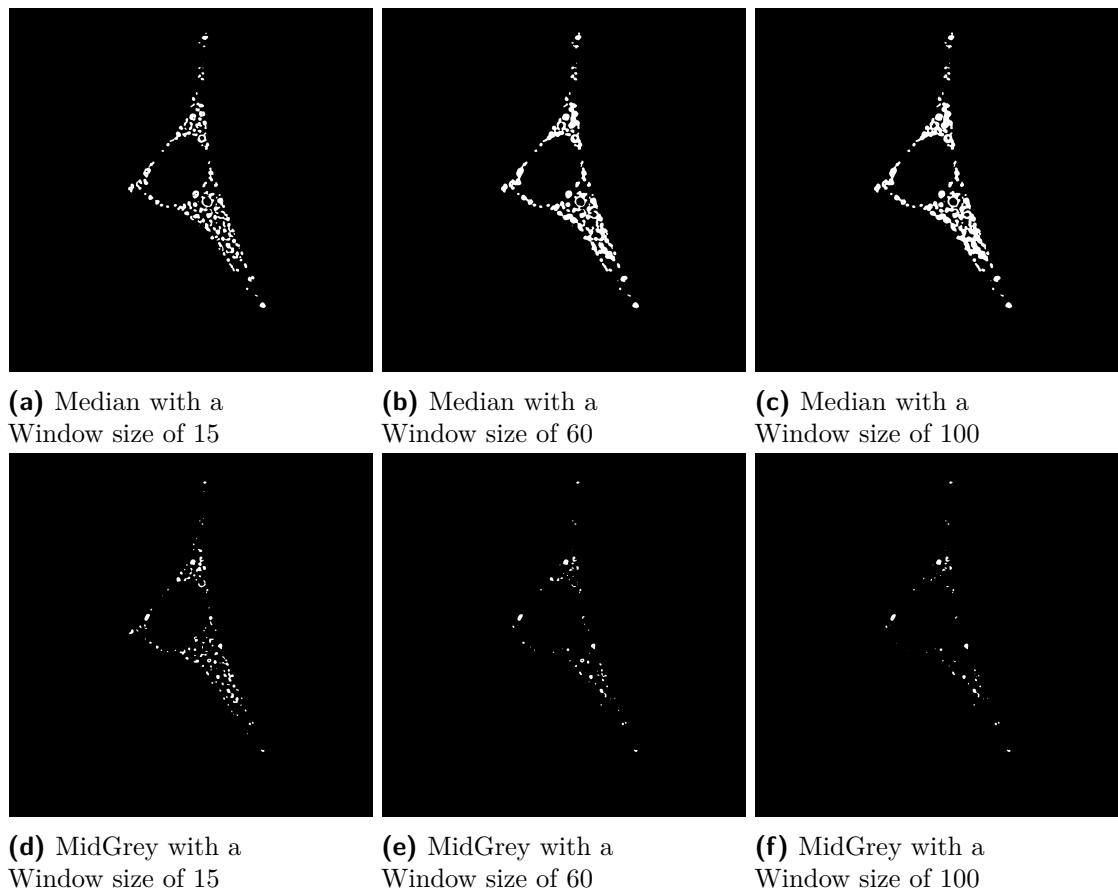


Figure 4.4: Showcasing the impact larger window sizes can have on local thresholding outcomes. For this example the Median and MidGrey thresholding methods have been applied to sample 3 with a subtraction constant of -5 .

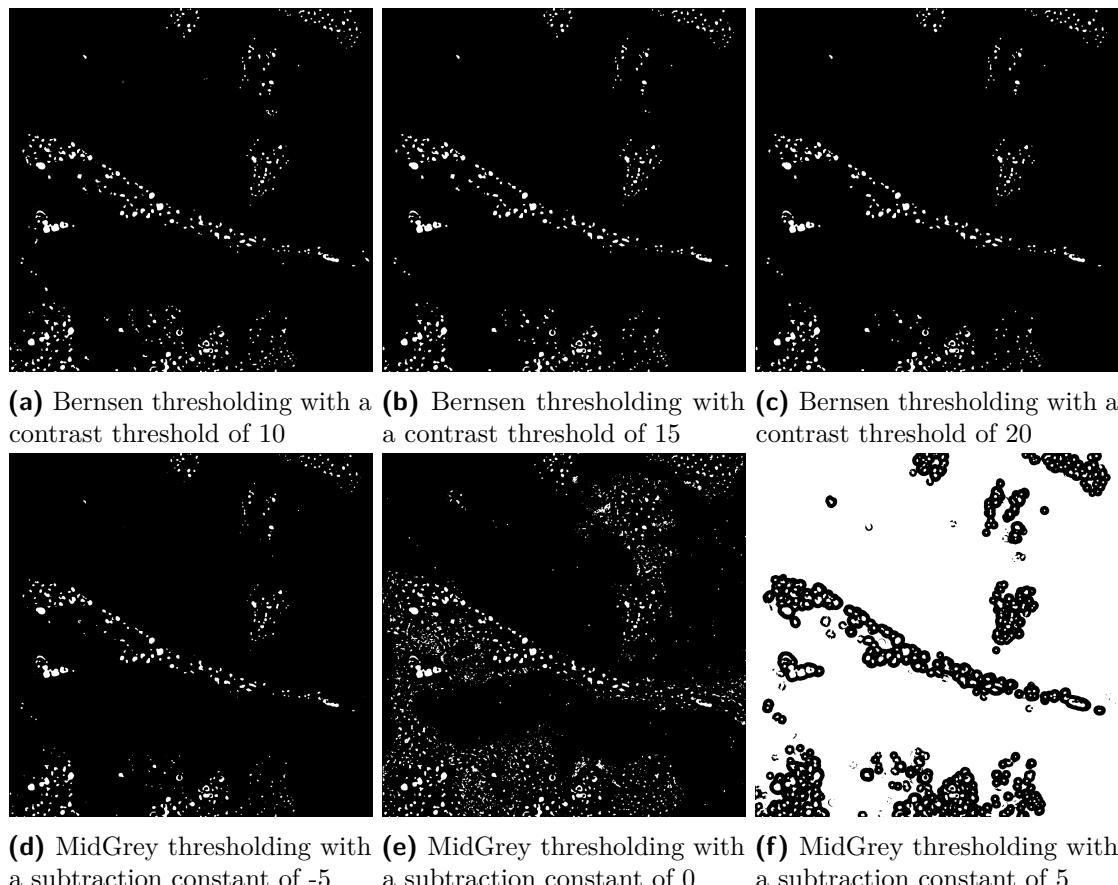


Figure 4.5: Showcase of the impact secondary parameter tuning can have on local thresholding methods. Bernsen and MidGrey are used in this showcase with both being applied to sample 2.

| Rank | Rank Description |
|------|---|
| 1 | The outcomes provided by this method are completely unusable |
| 2 | The outcomes present some under- or over-thresholding but the majority of structures are captured in the foreground |
| 3 | The outcomes are suitable for succeeding analysis although they still deviate from the manually tuned Hysteresis thresholding outcomes |
| 4 | The outcomes are equal to those achieved through manually tuned Hysteresis thresholding |
| 5 | The outcomes provided capture foreground structures while rejecting noise more effectively than the manually tuned Hysteresis thresholding (rejecting the false joining that Hysteresis is prone to is a contributor) |

Table 4.1: Table of the rankings with each ranks associated description

4.2. Visual analysis

With the global and local thresholding methods shortlisted, the performance of the AHT method can be evaluated against the shortlisted methods. This evaluation will be performed through visual analysis of the thresholding outcomes with comparison to manually tuned Hysteresis thresholding acting as the “ground truth”. The motivation is that humans are more capable of inferring if the correct structures have been labelled as foreground in the thresholding outcomes although this does introduce a measure of bias thus making this not a true ground truth but rather an approximation of the structures a humans might subjectively try to include as foreground in the image. The ranking of the thresholding methods performance, compared to the manually tuned Hysteresis thresholding, will be designated by a value ranging from 0 to 5 with the implication of each rank score detailed in Table 4.1.

Sample images: There is a total of three samples evaluated against with each presenting the lysosome, autophagosome and mitochondria organelle structures under different colour channels as shown in Figure 4.6. Red represents the mitochondria, green represents the autophagosomes, and blue represents the lysosomes but the overlaying of multiple colour channels can make differentiation between the structures challenging. For this reason, the performance of the thresholding measures will be explored for each organelle structure type where the rankings for method groups (global thresholds, local thresholds, and AHT bias variations) will be scored for each sample. An overall ranking for each thresholding method will be presented for each organelle type with the top three performing methods of each group explored further. All thresholding outcome images are present in Appendix B.

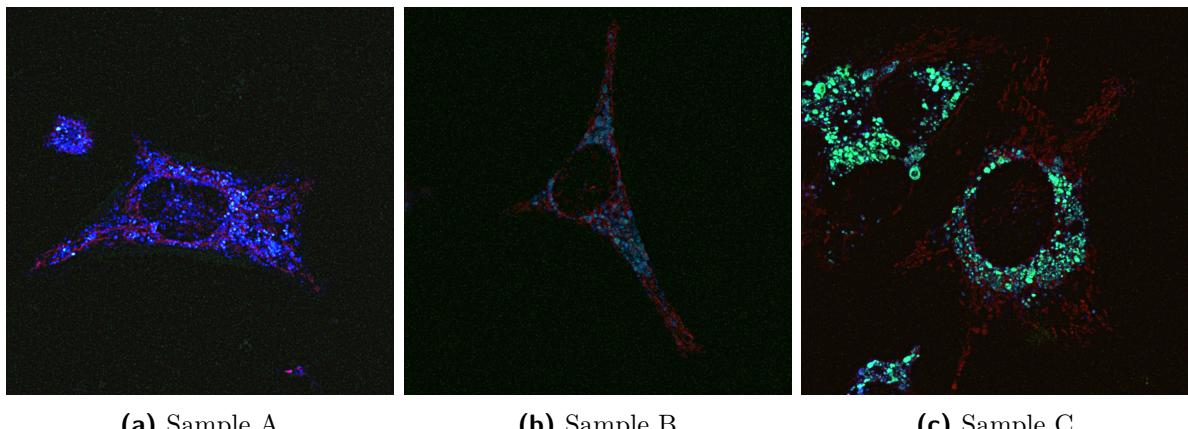


Figure 4.6: MIP sample images for the samples A-C where the organelle colour channels have been overlaid. The lysosomes are represented by the blue channel, the autophagosomes are represented by the green channel, and the mitochondria are represented by the red channel.

4.2.1. Lysosome structures

The lysosome structure images are presented in Figure 4.7 where the original MIP intensity images are shown along with the binary Hysteresis images for each sample. For each of these sample images, the accompanying Hysteresis threshold images are the foreground structures approximated through manual parameter tuning.

Sample A: Showcases foreground structures with high intensities that strongly contrast against the background but said background contains a large quantity of non-zero intensity noise and artefacts that could disrupt certain methods.

Sample B: Showcases foreground structures against a background containing little to no noise and, despite the maximum intensities for the majority of these structures being less than that of Sample A, the high contrast due to the low presence of noise could be ideal for many thresholding methods.

Sample C: Balanced between Sample A and Sample B, this sample possesses relatively high intensities for the foreground structures but with the presence of overlapping lower-intensity structure volumes, either belonging to out-of-focus structures or the structural volumes with low fluorophore concentrations, presents a challenge for confident differentiation for both manual and automatic thresholding methods.

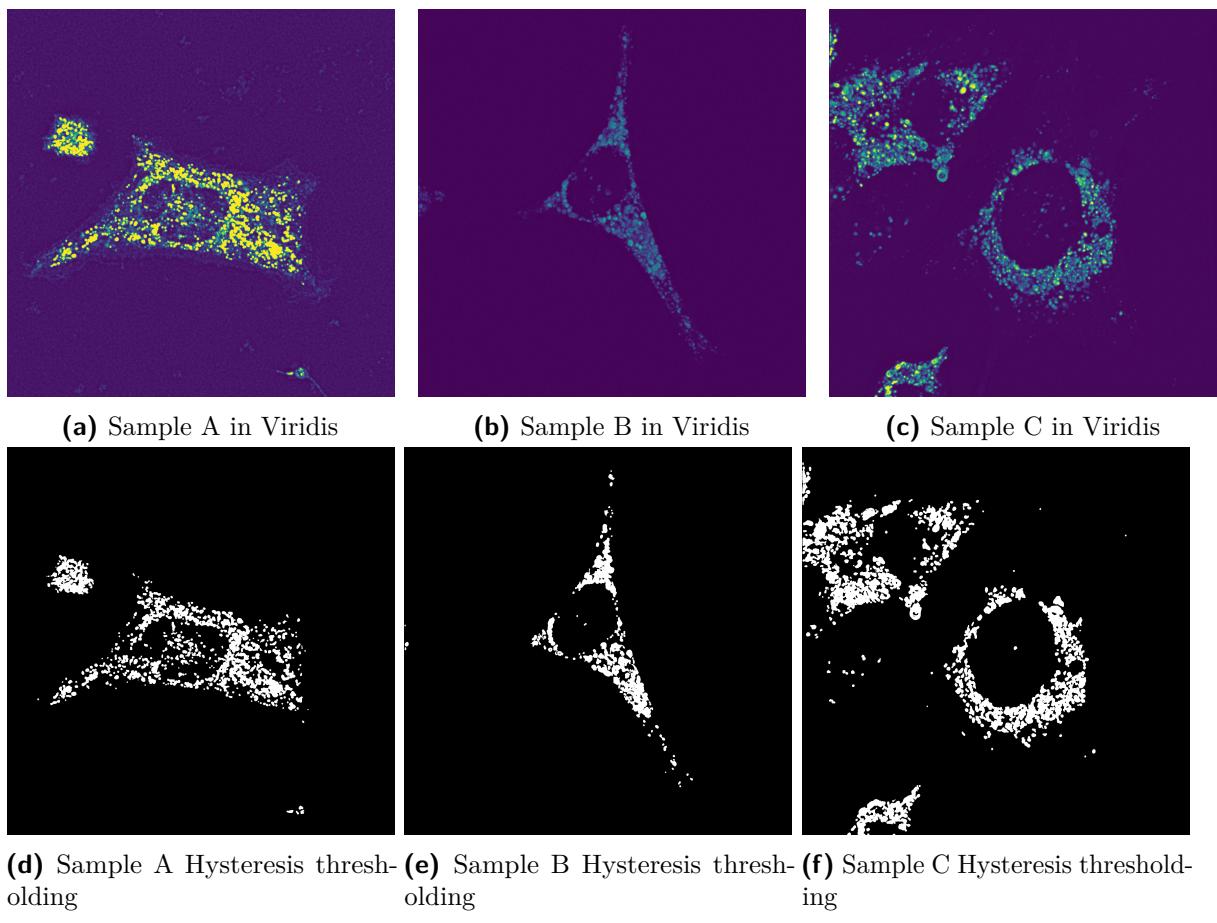


Figure 4.7: Showcase of the lysosome evaluation samples as MIP images. The first row shows the pre-processed images, prior to any thresholding, in the viridis colourmap. The second row shows the same samples after manually tuned Hysteresis thresholding has been applied.

Global thresholding methods

For the lysosome structures the top three global thresholding methods are Moments, IsoData and Otsu based on the mean rankings in Table 4.2.

| Method | Sample A | Sample B | Sample C | Mean Rank |
|--------------|----------|----------|----------|-----------|
| Huang2 | 1 | 2 | 2 | 1.6 |
| IsoData | 4 | 3 | 3 | 3.3 |
| Li | 1 | 1 | 1 | 1 |
| MaxEntropy | 1 | 3 | 2 | 2 |
| Moments | 3 | 3 | 5 | 3.6 |
| Otsu | 4 | 3 | 3 | 3.3 |
| RenyiEntropy | 1 | 3 | 5 | 3 |
| Triangle | 1 | 2 | 2 | 1.6 |
| Yen | 1 | 3 | 5 | 3 |

Table 4.2: Rankings for each of the global thresholding methods for the lysosome sample images

Moments The showcase of the Moments threshold outcomes is shown in Figure 4.8 with the Moments threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. There are also binarized images for each sample of only the Moments threshold outcomes which matches the red colour channel in the red-green images. From this figure, it is observed that the Moments threshold and Hysteresis threshold outcomes for both samples A and B are very similar save for Moments having a few more foreground structures. Considering the greater low-intensity noise present in sample A, compared to the other samples, and the low average structural intensities for sample B this difference of a few structures is negligible since the Hysteresis thresholding baseline compared against is an imperfect ground truth due to the bias involved in it. The biggest deviation is for sample C where Moments separates structures that are joined in the Hysteresis threshold outcome but this joining is likely false, a flaw of Hysteresis thresholding, based on the intensity image of sample C in Figure 4.7. For this reason, it is deemed that Moments thresholding outperforms Hysteresis thresholding for sample C.

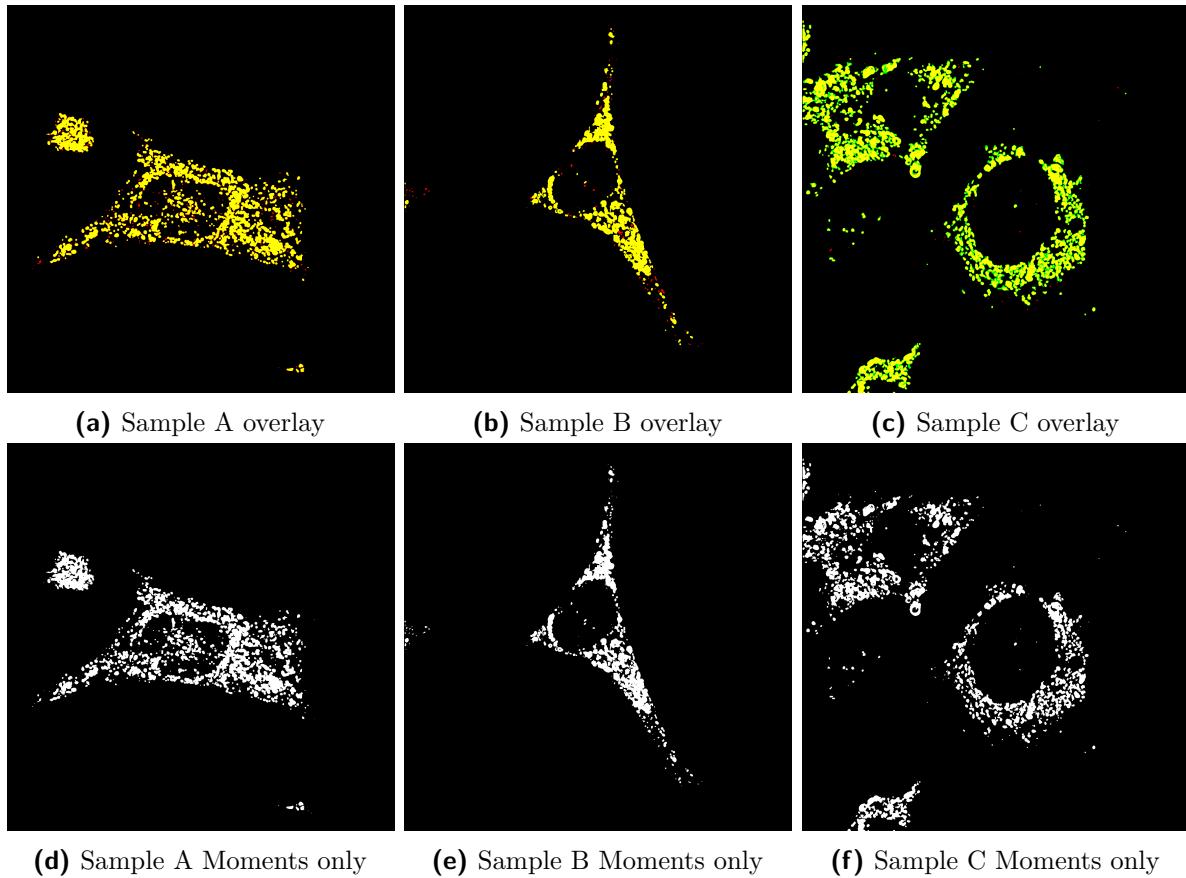


Figure 4.8: Showcase of the Moments thresholds outcomes with the top row being red-green overlay images with the Moments outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the Moments thresholding outcomes for improved clarity.

IsoData The showcase of the IsoData threshold outcomes is shown in Figure 4.9 with the IsoData threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. There are also binarized images for each sample of only the IsoData threshold outcomes which matches the red colour channel in the red-green images. The IsoData outcomes for all of the lysosome samples are quite similar to the Hysteresis thresholding outcomes although each expresses a minor degree of under-thresholding. This is observed to be a highly effective lysosome structure thresholding method as the extra structures captured by IsoData, that are not present in the Hysteresis threshold outcomes, are for structures with a mean intensity closer to the Hysteresis thresholding labelled foreground structures than the background.

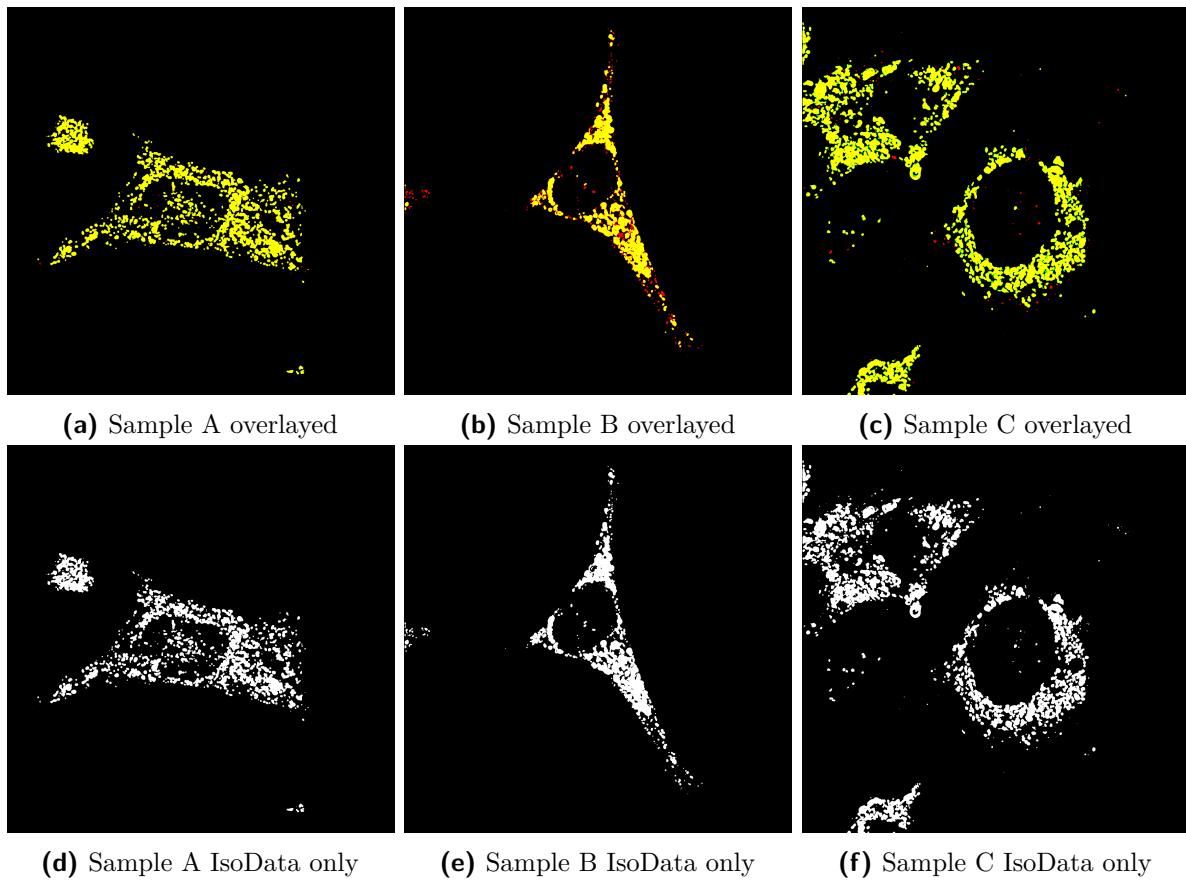


Figure 4.9: Showcase of the IsoData thresholds outcomes with the top row being red-green overlay images with the IsoData outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the IsoData thresholding outcomes for improved clarity.

Otsu The showcase of the Otsu threshold outcomes is shown in Figure 4.10 with the Otsu threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. There are also binarized images for each sample of only the Otsu threshold outcomes which matches the red colour channel in the red-green images. For sample A it is observed that the thresholding outcomes for both Otsu and Hysteresis

thresholding are similar with an arguably negligible number of extra structures present in the Otsu threshold outcome but not at a scale that could imply under-thresholding. Given the greater presence of noise in sample A than in the other samples this could imply that Otsu has some noise resilience but this might be dependent on the presence of high average intensities for foreground structures. Sample B presents structures with greater volume than in the Hysteresis threshold outcomes resulting in the increased joining of structures. While the lower average structural intensities in sample B reduces the definition of structure edges this is still quite problematic and shows that Otsu might be too lenient for lower contrast structures. Sample C presents a few extra structures for the Otsu thresholding but with some of these foreground labelled structures being quite dim, compared relative to the other structure intensities, there may be a minor degree of noise or artefacts being incorrectly labelled as foreground.

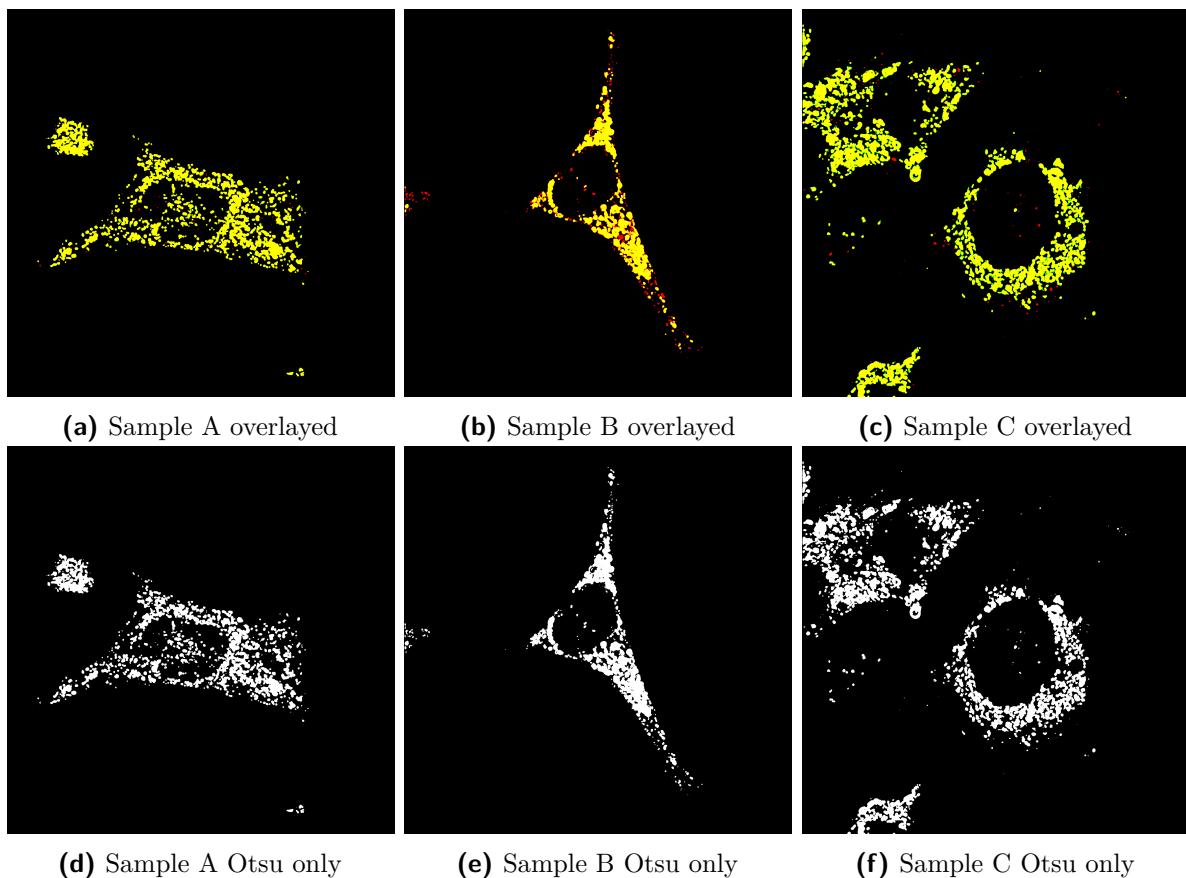


Figure 4.10: Showcase of the Otsu thresholds outcomes with the top row being red-green overlay images with the Otsu outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the Otsu thresholding outcomes for improved clarity.

Local thresholding methods

As there are only 4 local thresholding methods evaluated against only the top 2 performing methods will be explored further with these being the MidGrey and Bernsen thresold methods. An observation that can be made is that all of these thresholding outcomes are poor with all methods providing unusable results for sample A. This is likely due to the noise present in sample A being unfavourable to all of the local thresholding methods explored with performance varying for samples B and C.

| Method | Sample A | Sample B | Sample C | Mean Rank |
|---------|----------|----------|----------|-----------|
| Bernsen | 1 | 3 | 2 | 2 |
| Mean | 1 | 2 | 2 | 1.6 |
| Median | 1 | 2 | 2 | 1.6 |
| MidGrey | 1 | 4 | 3 | 2.6 |

Table 4.3: Rankings for each of the local thresholding methods for the lysosome sample images

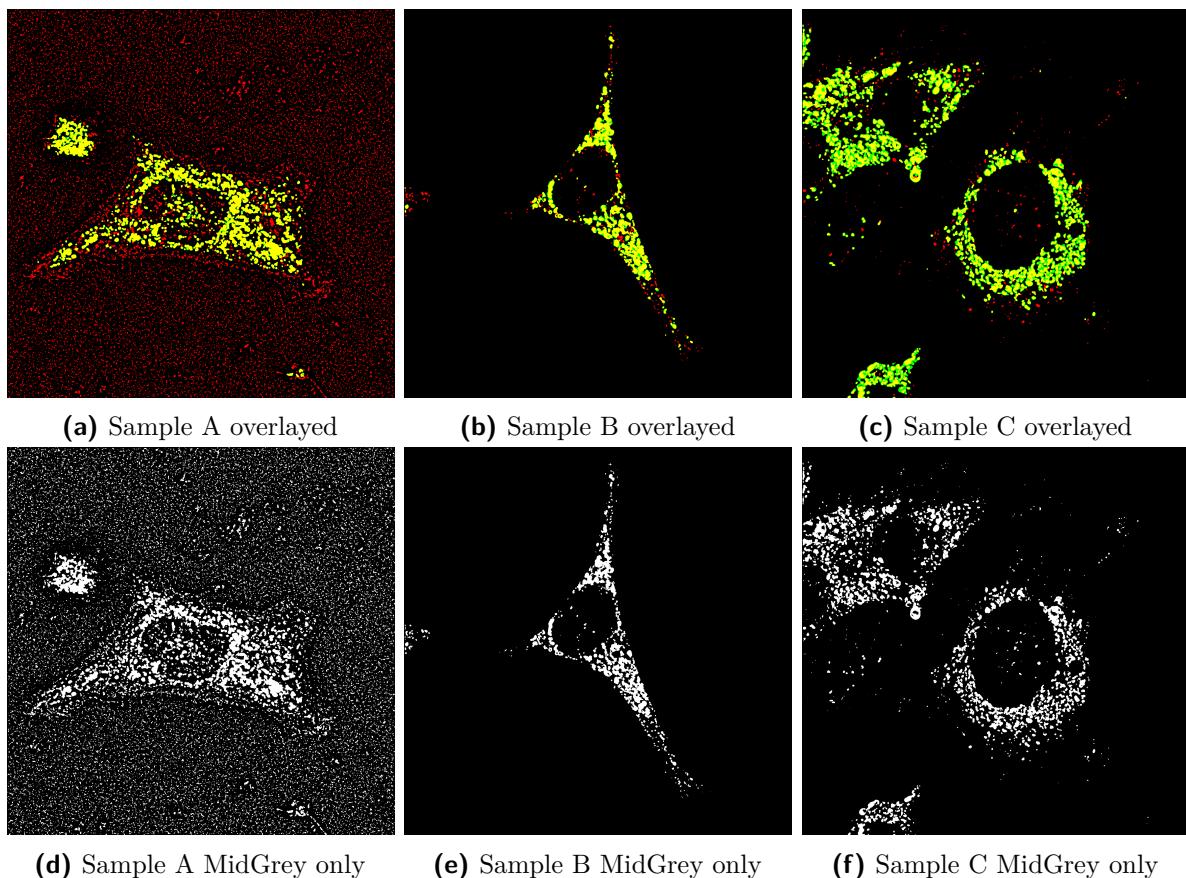


Figure 4.11: Showcase of the MidGrey thresholds outcomes with the top row being red-green overlay images with the MidGrey outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the MidGrey thresholding outcomes for improved clarity.

MidGrey The showcase of the MidGrey threshold outcomes is shown in Figure 4.11 with the MidGrey threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. There are also binarized images for each sample of only the MidGrey threshold outcomes which matches the red colour channel in the red-green images. Sample B for MidGrey is a relatively good outcome that is similar to the Hysteresis thresholding outcome, with a minor quantity of foreground structures which are subjectively viable depending on the evaluator's bias, and without any extra structure joining that is not present in the Hysteresis thresholding outcome. For sample C there are more foreground structures compared to the Hysteresis thresholding outcome that do not all map to structures with intensities closer to the foreground than the background thus there is some measure, although minor, of under-thresholding occurring.

Bernsen The showcase of the Bernsen threshold outcomes is shown in Figure 4.12 with the Bernsen threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. For sample B, while the primary foreground structures are the same as for the Hysteresis threshold outcome there is both an increased quantity of foreground structures and of structure joining. This makes the sample B outcome significantly less viable but minimal noise captured in the foreground renders this a weak but not wholly unusable result depending on the analysis. For sample C, there is an increase in the quantity of foreground structures, beyond what could be subjectively considered likely foreground, but there is an increase in structure separation. While not a great method for lysosome samples it could be applied for certain samples under specific image conditions but there are likely better methods to use.

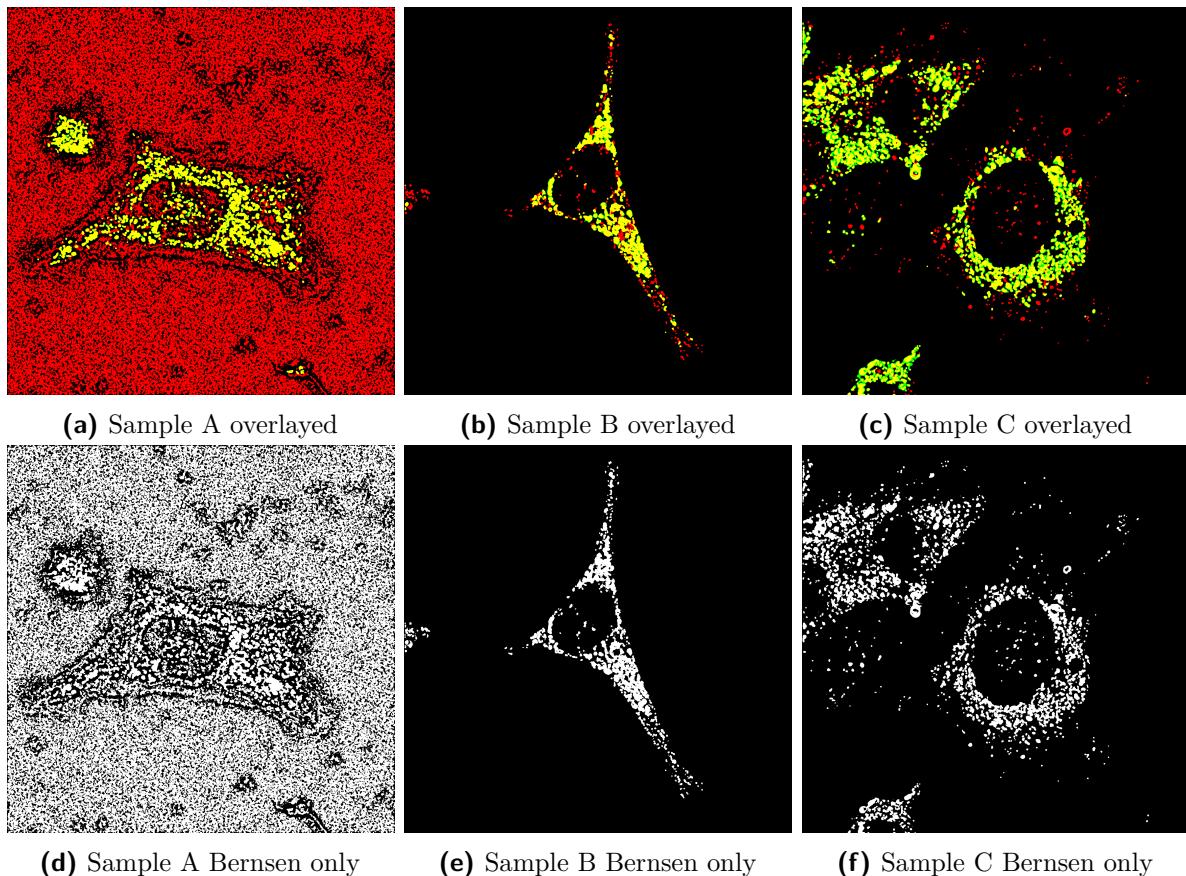


Figure 4.12: Showcase of the Bernsen thresholds outcomes with the top row being red-green overlay images with the Bernsen outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the Bernsen thresholding outcomes for improved clarity.

AHT method with varied bias applications

The rankings of the different AHT method biases are shown in Table 4.4 with most of these methods presenting mediocre performance overall. It was observed that a consistent detractor to their performance was due to the AHT approximated low threshold being low enough that it induced excessive structure joining. This is as far as I am currently. The current idea for the AHT is that since there isn't a clear top 3 I am considering taking the best, worst, and in-between? I am a bit uncertain. I am going to place three lineplots for the each of the lysosome samples which will have the IHH distribution, the inverted gradient distribution, vertical lines for the high threshold determined for each of the bias combos and potentially the distribution product (for when the IHH distribution is applied to the inverted gradient distribution). I plan to copy past this structure for autophagosomes and mitochondria.

| IHH bias applied | Window bias applied | Sample A | Sample B | Sample C | Mean Rank |
|------------------|---------------------|----------|----------|----------|-----------|
| No | None | 3 | 1 | 4 | 2.6 |
| No | Window width | 2 | 2 | 4 | 2.6 |
| No | Window mass | 2 | 1 | 4 | 2.3 |
| Yes | None | 2 | 3 | 4 | 3 |
| Yes | Window width | 2 | 3 | 3 | 2.6 |
| Yes | Window mass | 2 | 3 | 3 | 2.6 |

Table 4.4: Rankings of the AHT threshold outcomes for the three samples with different combinations of the IHH and Window biases

4.2.2. Autophagosome structures

The autophagosome structure images are presented in Figure 4.13 where the original MIP intensity images are shown along with the binary Hysteresis images for each sample. For each of these sample images, the accompanying Hysteresis threshold images are the foreground structures approximated through manual parameter tuning.

Sample A: Showcases foreground structures with high intensities that strongly contrast against the background but said background contains a large quantity of non-zero intensity noise and artefacts that could disrupt certain methods.

Sample B: Showcases foreground structures against a background containing little to no noise and, despite the maximum intensities for the majority of these structures being less than that of Sample A, the high contrast due to the low presence of noise could be ideal for many thresholding methods.

Sample C: Balanced between Sample A and Sample B, this sample possesses relatively high intensities for the foreground structures but with the presence of overlapping lower-intensity structure volumes, either belonging to out-of-focus structures or the structural volumes with low fluorophore concentrations, presents a challenge for confident differentiation for both manual and automatic thresholding methods.

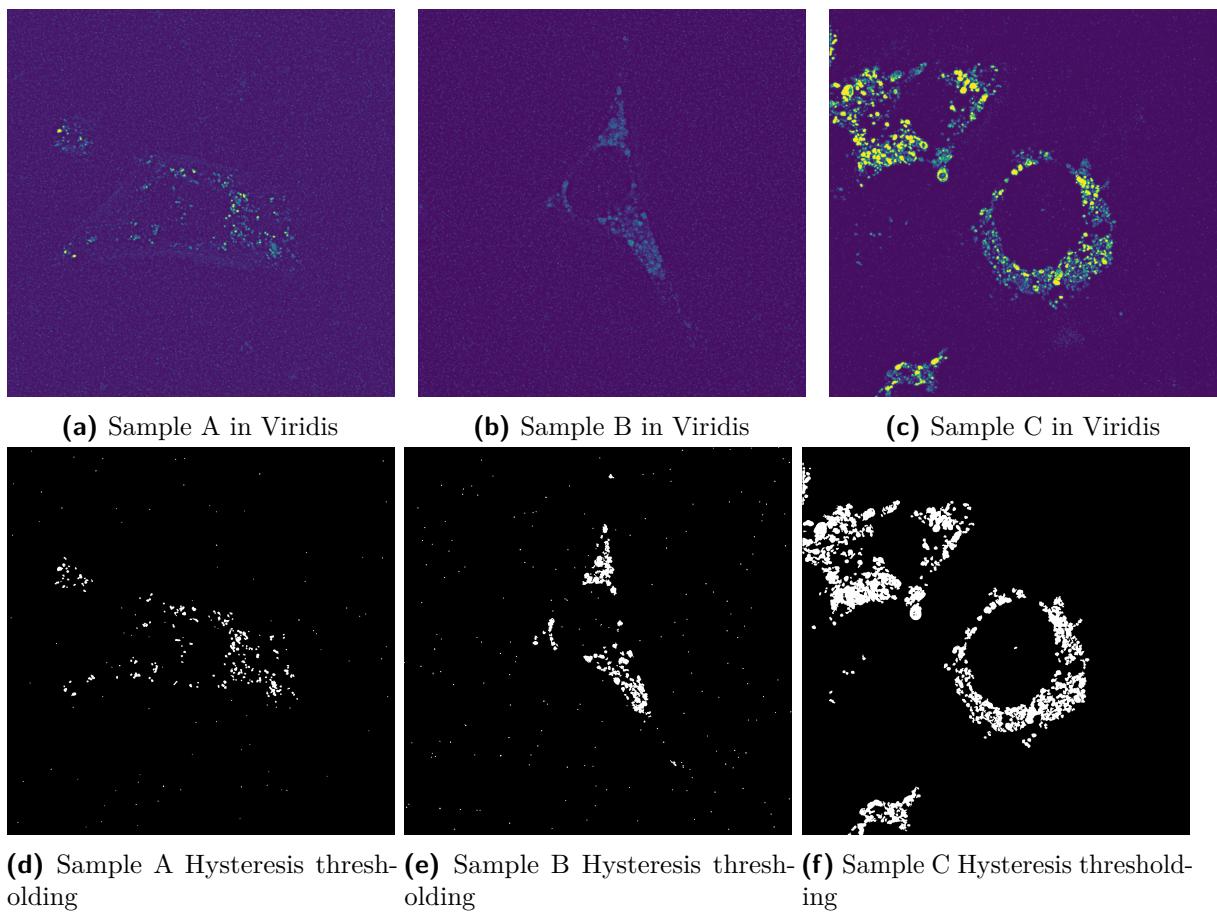


Figure 4.13: Showcase of the autophagosome evaluation samples as MIP images. The first row shows the pre-processed images, prior to any thresholding, in the viridis colourmap. The second row shows the same samples after manually tuned Hysteresis thresholding has been applied.

Global thresholding methods

For the autophagosome structures the top three global thresholding methods are Moments, IsoData and Otsu based on the mean rankings in Table 4.5.

| Method | Sample A | Sample B | Sample C | Mean Rank |
|--------------|----------|----------|----------|-----------|
| Huang2 | 1 | 2 | 2 | 1.6 |
| IsoData | 4 | 3 | 3 | 3.3 |
| Li | 1 | 1 | 1 | 1 |
| MaxEntropy | 1 | 3 | 2 | 2 |
| Moments | 3 | 3 | 5 | 3.6 |
| Otsu | 4 | 3 | 3 | 3.3 |
| RenyiEntropy | 1 | 3 | 5 | 3 |
| Triangle | 1 | 2 | 2 | 1.6 |
| Yen | 1 | 3 | 5 | 3 |

Table 4.5: Rankings for each of the global thresholding methods for the autophagosome sample images

Moments The showcase of the Moments threshold outcomes is shown in Figure 4.14 with the Moments threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. There are also binarized images for each sample of only the Moments threshold outcomes which matches the red colour channel in the red-green images. From this figure, it is observed that the Moments threshold and Hysteresis threshold outcomes for both samples A and B are very similar save for Moments having a few more foreground structures. Considering the greater low-intensity noise present in sample A, compared to the other samples, and the low average structural intensities for sample B this difference of a few structures is negligible since the Hysteresis thresholding baseline compared against is an imperfect ground truth due to the bias involved in it. The biggest deviation is for sample C where Moments separates structures that are joined in the Hysteresis threshold outcome but this joining is likely false, a flaw of Hysteresis thresholding, based on the intensity image of sample C in Figure 4.13. For this reason, it is deemed that Moments thresholding outperforms Hysteresis thresholding for sample C.

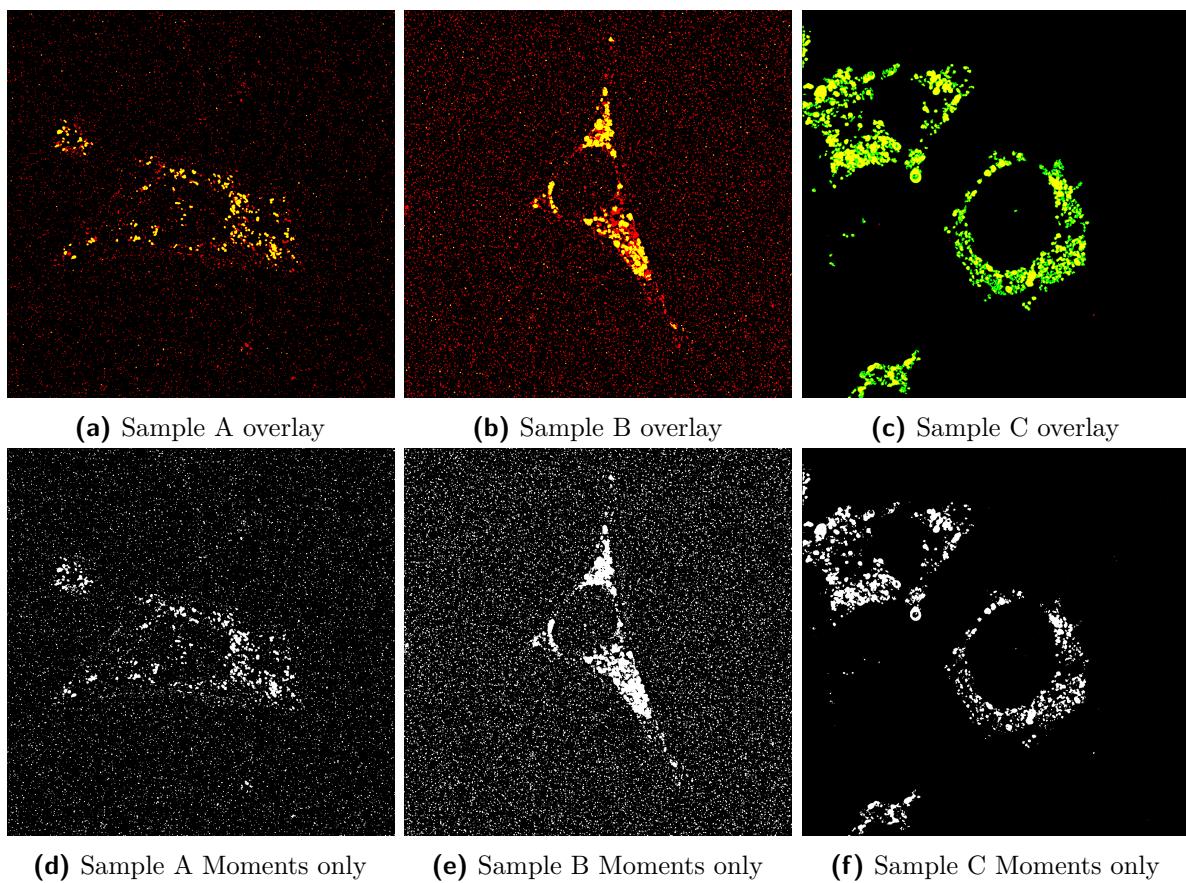


Figure 4.14: Showcase of the Moments thresholds outcomes with the top row being red-green overlay images with the Moments outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the Moments thresholding outcomes for improved clarity.

IsoData The showcase of the IsoData threshold outcomes is shown in Figure 4.15 with the IsoData threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. There are also binarized images for each sample of only the IsoData threshold outcomes which matches the red colour channel in the red-green images. The IsoData outcomes for all of the autophagosome samples are quite similar to the Hysteresis thresholding outcomes although each expresses a minor degree of under-thresholding. This is observed to be a highly effective autosome structure thresholding method as the extra structures captured by IsoData, that are not present in the Hysteresis thresholding outcomes, are for structures with a mean intensity closer to the Hysteresis thresholding labelled foreground structures than the background.

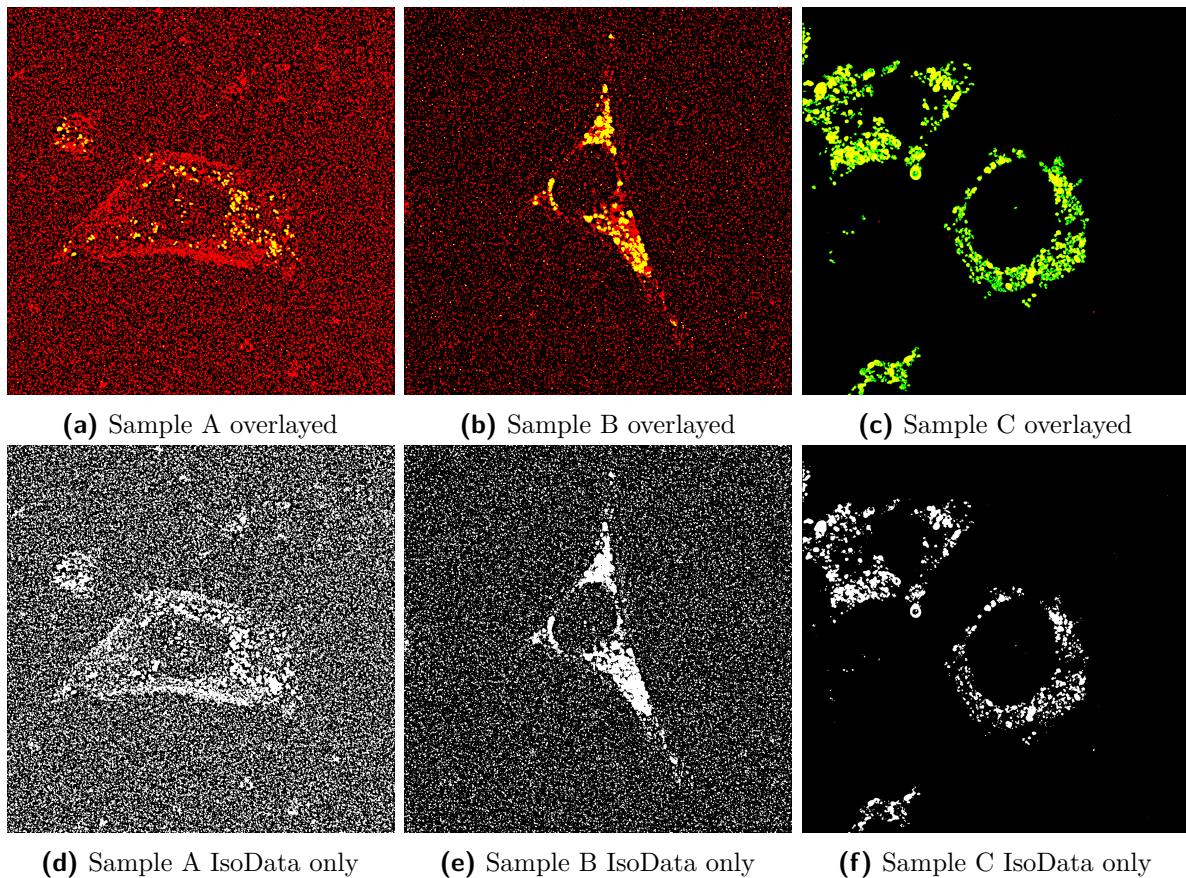


Figure 4.15: Showcase of the IsoData thresholds outcomes with the top row being red-green overlay images with the IsoData outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the IsoData thresholding outcomes for improved clarity.

Otsu The showcase of the Otsu threshold outcomes is shown in Figure 4.16 with the Otsu threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. There are also binarized images for each sample of only the Otsu threshold outcomes which matches the red colour channel in the red-green images. For sample A it is observed that the thresholding outcomes for both Otsu and Hysteresis

thresholding are similar with an arguably negligible number of extra structures present in the Otsu threshold outcome but not at a scale that could imply under-thresholding. Given the greater presence of noise in sample A than in the other samples this could imply that Otsu has some noise resilience but this might be dependent on the presence of high average intensities for foreground structures. Sample B presents structures with greater volume than in the Hysteresis threshold outcomes resulting in the increased joining of structures. While the lower average structural intensities in sample B reduces the definition of structure edges this is still quite problematic and shows that Otsu might be too lenient for lower contrast structures. Sample C presents a few extra structures for the Otsu thresholding but with some of these foreground labelled structures being quite dim, compared relative to the other structure intensities, there may be a minor degree of noise or artefacts being incorrectly labelled as foreground.

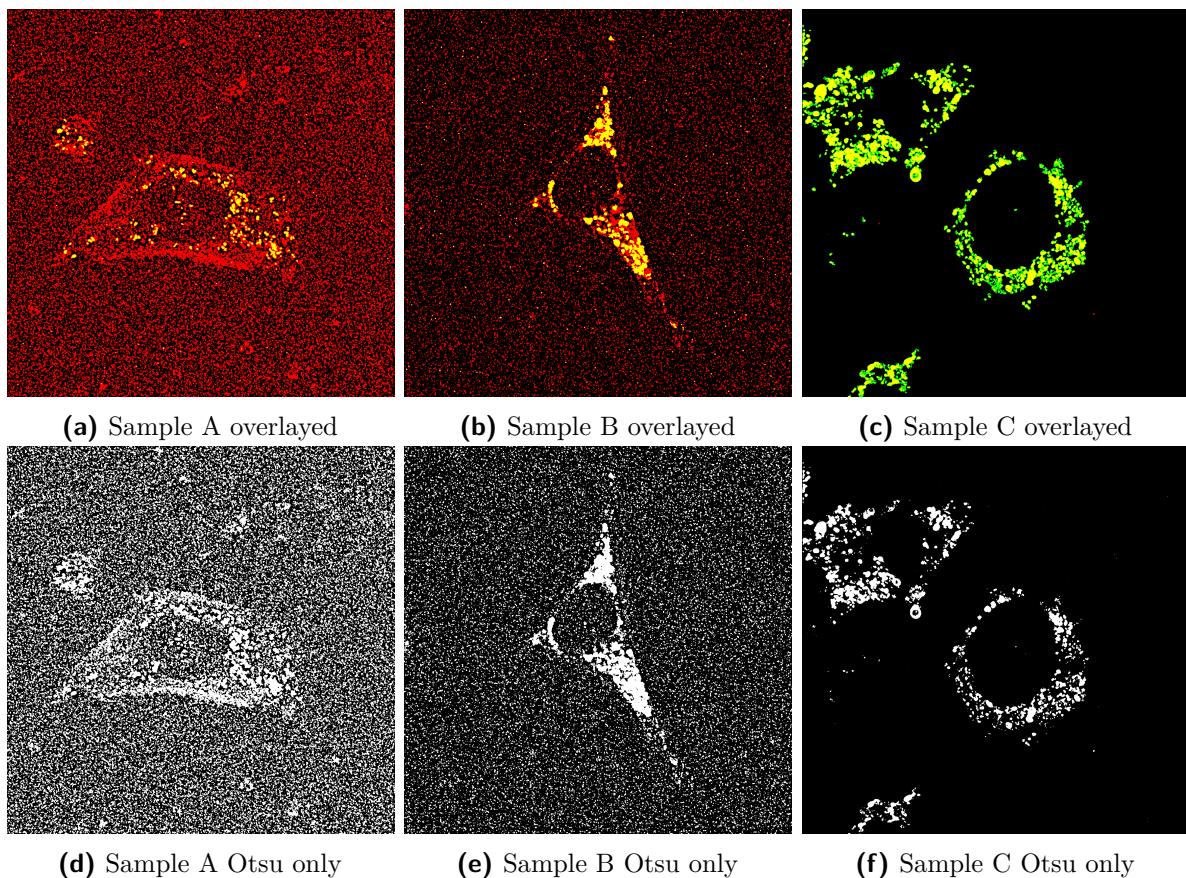


Figure 4.16: Showcase of the Otsu thresholds outcomes with the top row being red-green overlay images with the Otsu outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the Otsu thresholding outcomes for improved clarity.

Local thresholding methods

As there are only 4 local thresholding methods evaluated against only the top 2 performing methods will be explored further with these being the MidGrey and Bernsen thresold methods. An observation that can be made is that all of these thresholding outcomes are poor with all methods providing unusable results for sample A. This is likely due to the noise present in sample A being unfavourable to all of the local thresholding methods explored with performance varying for samples B and C.

| Method | Sample A | Sample B | Sample C | Mean Rank |
|---------|----------|----------|----------|-----------|
| Bernsen | 1 | 3 | 2 | 2 |
| Mean | 1 | 2 | 2 | 1.6 |
| Median | 1 | 2 | 2 | 1.6 |
| MidGrey | 1 | 4 | 3 | 2.6 |

Table 4.6: Rankings for each of the local thresholding methods for the autosome sample images

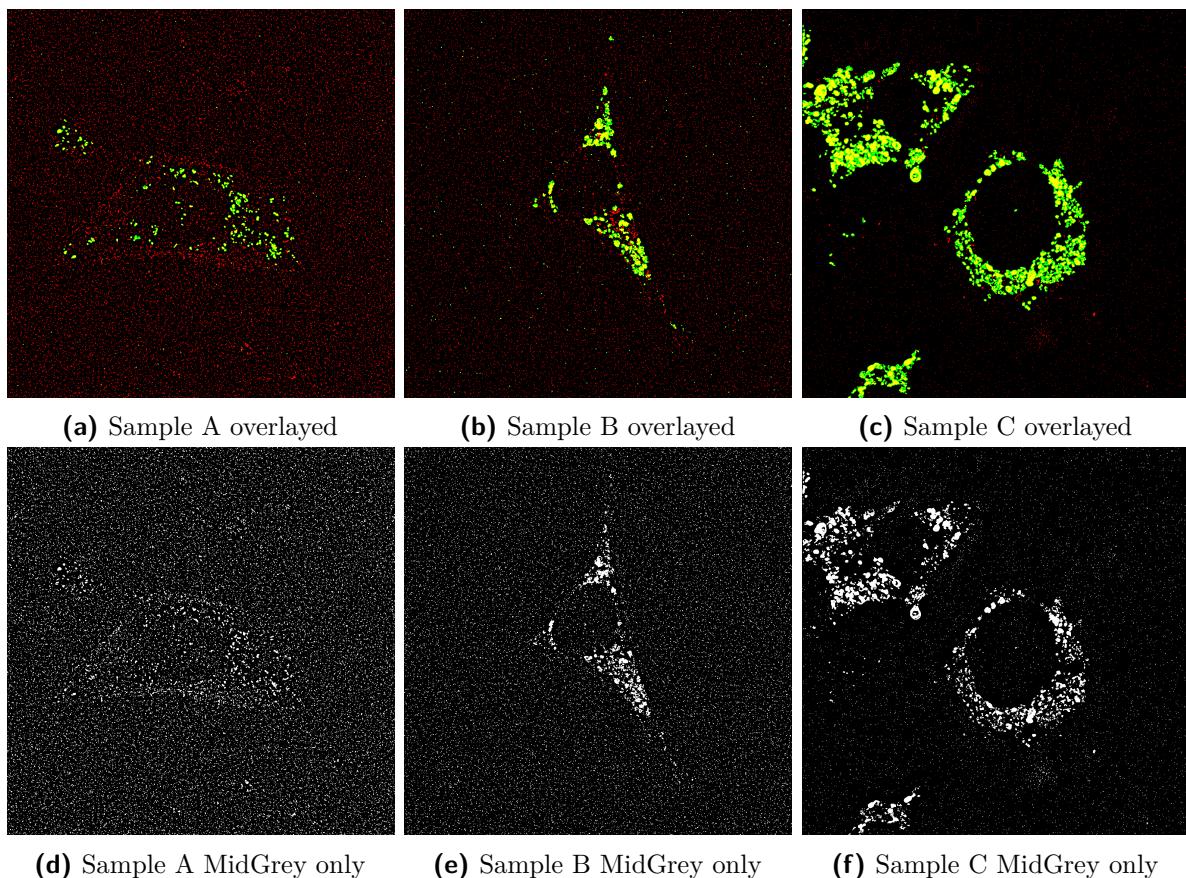


Figure 4.17: Showcase of the MidGrey thresholds outcomes with the top row being red-green overlay images with the MidGrey outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the MidGrey thresholding outcomes for improved clarity.

MidGrey The showcase of the MidGrey threshold outcomes is shown in Figure 4.17 with the MidGrey threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. There are also binarized images for each sample of only the MidGrey threshold outcomes which matches the red colour channel in the red-green images. Sample B for MidGrey is a relatively good outcome that is similar to the Hysteresis thresholding outcome, with a minor quantity of foreground structures which are subjectively viable depending on the evaluator's bias, and without any extra structure joining that is not present in the Hysteresis thresholding outcome. For sample C there are more foreground structures compared to the Hysteresis thresholding outcome that do not all map to structures with intensities closer to the foreground than the background thus there is some measure, although minor, of under-thresholding occurring.

Bernsen The showcase of the Bernsen threshold outcomes is shown in Figure 4.18 with the Bernsen threshold outcomes compared against the Hysteresis threshold outcomes in a red-green image for each sample. For sample B, while the primary foreground structures are the same as for the Hysteresis threshold outcome there is both an increased quantity of foreground structures and of structure joining. This makes the sample B outcome significantly less viable but minimal noise captured in the foreground renders this a weak but not wholly unusable result depending on the analysis. For sample C, there is an increase in the quantity of foreground structures, beyond what could be subjectively considered likely foreground, but there is an increase in structure separation. While not a great method for autosome samples it could be applied for certain samples under specific image conditions but there are likely better methods to use.

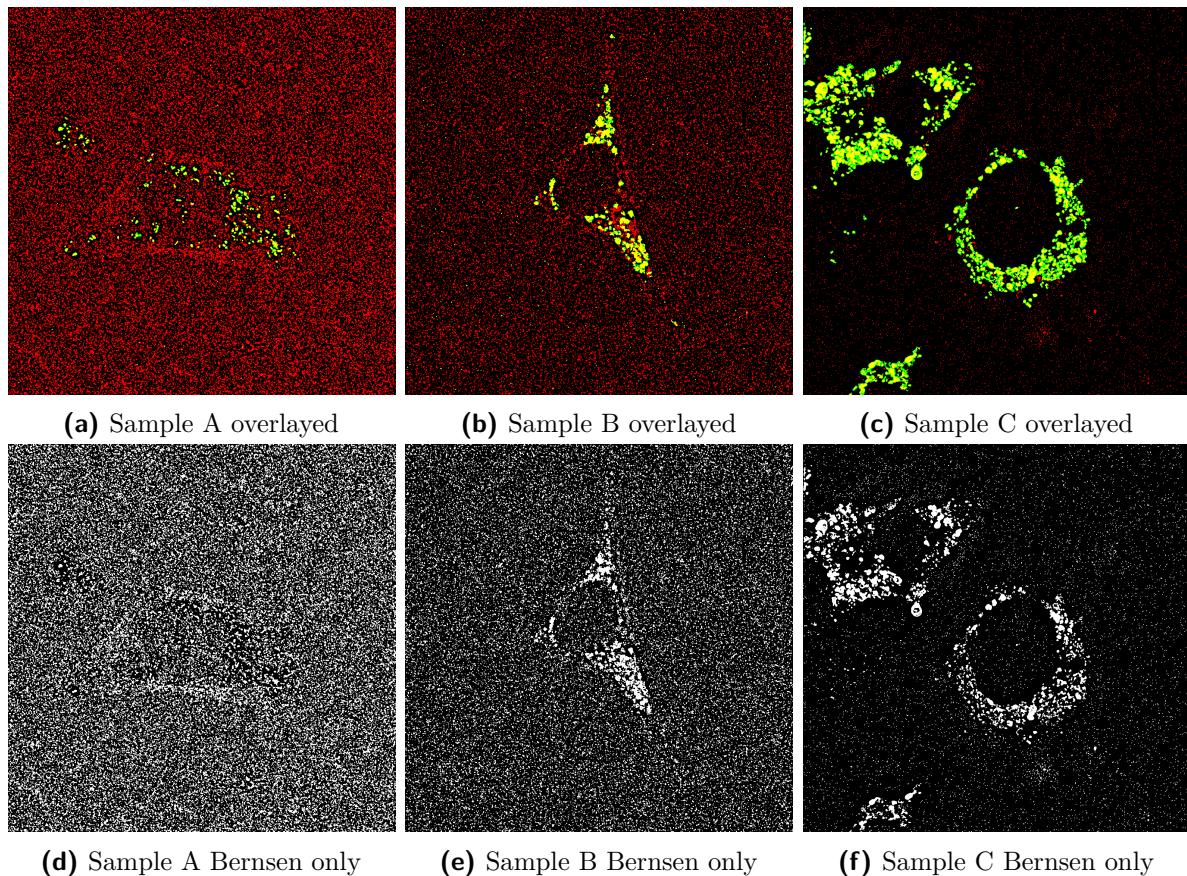


Figure 4.18: Showcase of the Bernsen thresholds outcomes with the top row being red-green overlay images with the Bernsen outcomes in red, the Hysteresis thresholding outcomes in green, and overlaps of both outcomes in yellow. The bottom row showcases only the Bernsen thresholding outcomes for improved clarity.

Chapter 5

Conclusion

5.1.

Bibliography

- [1] L. Sherwood and K. Sutton, *Human Physiology: From Cells to Systems*. Andover: CENGAGE Learning, 2015.
- [2] L. Doblado, C. Lueck, C. Rey, A. K. Samhan-Arias, I. Prieto, A. Stacchiotti, and M. Monsalve, “Mitophagy in Human Diseases,” *Int J Mol Sci*, vol. 22, no. 8, Apr 2021.
- [3] K. R. Castleman, “Chapter 3 - image digitization,” in *Microscope Image Processing*, Q. Wu, F. A. Merchant, and K. R. Castleman, Eds. Burlington: Academic Press, 2008, pp. 27–38. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B978012372578300009X>
- [4] Q. Wu and K. R. Castleman, “Chapter 9 - image segmentation,” in *Microscope Image Processing*, Q. Wu, F. A. Merchant, and K. R. Castleman, Eds. Burlington: Academic Press, 2008, pp. 159–194. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B978012372578300009X>
- [5] Y. Al-Kofahi, A. Zaltsman, R. Graves, W. Marshall, and M. Rusu, “A deep learning-based algorithm for 2-d cell segmentation in microscopy images,” *BMC Bioinformatics*, vol. 19, no. 1, p. 365, Oct 2018. [Online]. Available: <https://doi.org/10.1186/s12859-018-2375-z>
- [6] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, “scikit-image: image processing in Python,” *PeerJ*, vol. 2, p. e453, 6 2014. [Online]. Available: <https://doi.org/10.7717/peerj.453>
- [7] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.

- [8] S. de Wet, A. Mangali, R. Batt, J. Kriel, N. Vahrmeijer, D. Niehaus, R. Theart, and B. Loos, “The highs and lows of memantine—an autophagy and mitophagy inducing agent that protects mitochondria,” *Cells*, vol. 12, no. 13, 2023. [Online]. Available: <https://www.mdpi.com/2073-4409/12/13/1726>
- [9] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, and et al., “Fiji: An open-source platform for biological-image analysis,” *Nature Methods*, vol. 9, no. 7, pp. 676–682, 2012. [Online]. Available: <https://www.nature.com/articles/nmeth.2019>
- [10] A. Chaudhry, R. Shi, and D. S. Luciani, “A pipeline for multidimensional confocal analysis of mitochondrial morphology, function, and dynamics in pancreatic β -cells,” *American Journal of Physiology-Endocrinology and Metabolism*, vol. 318, no. 2, pp. E87–E101, 2020, pMID: 31846372. [Online]. Available: <https://doi.org/10.1152/ajpendo.00457.2019>
- [11] S. Castro-Obregon, “The discovery of lysosomes and autophagy,” *Nature Education*, vol. 3, no. 9, p. 49, 2010.
- [12] R. J. Youle and A. M. van der Bliek, “Mitochondrial fission, fusion, and stress,” *Science (New York, N.Y.)*, vol. 337, no. 6098, pp. 1062–1065, 2012. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4762028/>
- [13] R. J. Youle and D. P. Narendra, “Mechanisms of mitophagy,” *Nature reviews. Molecular cell biology*, vol. 12, no. 1, pp. 9–14, 2011. [Online]. Available: <https://www.nature.com/articles/nrm3028>
- [14] *Molecular biology of the cell*, 5th ed. New York: Garland Science, 2008.
- [15] M. J. Sanderson, I. Smith, I. Parker, and M. D. Bootman, “Fluorescence microscopy,” *Cold Spring Harbor Protocols*, vol. 2014, no. 10, p. pdb.top071795, 2014. [Online]. Available: <http://cshprotocols.cshlp.org/content/2014/10/pdb.top071795.full>
- [16] D.-X. Cai, Y. Quan, P.-J. He, H.-B. Tan, and Y.-Q. Xu, “Dynamic perfusion culture of human outgrowth endothelial progenitor cells on demineralized bone matrix in vitro,” *Medical Science Monitor*, vol. 22, pp. 4037–4045, 10 2016.
- [17] H. Brown-Harding, “Which fluorophores to use for your microscopy experiment.” [Online]. Available: <https://expert.cheekyscientist.com/which-fluorophores-to-use-for-your-microscopy-experiment/>
- [18] G. Haimovich, “‘the red color dilemma: how to choose the right fluorescent protein’,greenfluorescentblog.” [On-

- line]. Available: <https://greenfluorescentblog.wordpress.com/2012/05/17/the-red-color-dilemma-how-to-choose-the-right-fluorescent-protein/>
- [19] D. M. Chudakov, M. V. Matz, S. Lukyanov, and K. A. Lukyanov, “Fluorescent proteins and their applications in imaging living cells and tissues,” *Physiological Reviews*, vol. 90, no. 3, pp. 1103–1163, 2010, pMID: 20664080. [Online]. Available: <https://doi.org/10.1152/physrev.00038.2009>
- [20] S. Bolte and F. P. Cordelières, “A guided tour into subcellular colocalization analysis in light microscopy,” *Journal of Microscopy*, vol. 224, no. 3, pp. 213–232, 2006. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2818.2006.01706.x>
- [21] R. Milo and R. Phillips, *Cell Biology by the Numbers*, 12 2015.
- [22] L. Song, E. J. Hennink, I. T. Young, and H. J. Tanke, “Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy,” *Biophys. J.*, vol. 68, no. 6, pp. 2588–2600, Jun. 1995.
- [23] T. Bernas, B. P. Rajwa, E. K. Asem, and J. P. Robinson, “Loss of image quality in photobleaching during microscopic imaging of fluorescent probes bound to chromatin,” *Journal of Biomedical Optics*, vol. 10, no. 6, p. 064015, 2005. [Online]. Available: <https://doi.org/10.1117/1.2136313>
- [24] G. H. Patterson, “Fluorescence microscopy below the diffraction limit,” *Seminars in Cell & Developmental Biology*, vol. 20, no. 8, pp. 886–893, 2009, imaging in Cell and Developmental Biology Planar Cell Polarity. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1084952109001633>
- [25] P. Dedecker, J. Hofkens, and J. ichi Hotta, “Diffraction-unlimited optical microscopy,” *Materials Today*, vol. 11, pp. 12–21, 2008. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1369702109700033>
- [26] A. D. Elliott, “Confocal microscopy: Principles and modern practices,” *Current Protocols in Cytometry*, vol. 92, no. 1, p. e68, 2020. [Online]. Available: <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpcy.68>
- [27] C. M. St. Croix, S. H. Shand, and S. C. Watkins, “Confocal microscopy: comparisons, applications, and problems,” *BioTechniques*, vol. 39, no. 6S, pp. S2–S5, 2005. [Online]. Available: <https://doi.org/10.2144/000112089>
- [28] J. Jonkman, C. Brown, and R. Cole, “Quantitative confocal microscopy: Beyond a pretty picture,” *Methods in cell biology*, vol. 123C, pp. 113–134, 06 2014.

- [29] K. W. Dunn, M. M. Kamocka, and J. H. McDonald, “A practical guide to evaluating colocalization in biological microscopy,” *American Journal of Physiology-Cell Physiology*, vol. 300, no. 4, pp. C723–C742, Apr 2011.
- [30] R. P. Theart, B. Loos, and T. R. Niesler, “Regression adjusted colocalisation colour mapping (racc): A novel biological visual analysis method for qualitative colocalisation analysis of 3d fluorescence micrographs,” *PLOS ONE*, vol. 14, no. 11, pp. 1–21, 11 2019. [Online]. Available: <https://doi.org/10.1371/journal.pone.0225141>
- [31] W. Meiniel, J. C. Olivo-Marin, and E. D. Angelini, “Denoising of microscopy images: A review of the state-of-the-art, and a new sparsity-based method,” *IEEE Transactions on Image Processing*, vol. 27, pp. 3842–3856, 8 2018.
- [32] Y. Zhang, Y. Zhu, E. Nichols, Q. Wang, S. Zhang, C. Smith, and S. Howard, “A poisson-gaussian denoising dataset with real fluorescence microscopy images,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 702–11 710.
- [33] H. Kirshner, D. Sage, and M. Unser, “3D PSF models for fluorescence microscopy in ImageJ,” in *Proceedings of the Twelfth International Conference on Methods and Applications of Fluorescence Spectroscopy, Imaging and Probes (MAF’11)*, Strasbourg, French Republic, September 11-14, 2011, p. 154.
- [34] P. E. Robinson and Y. Roodt, “Blind deconvolution of gaussian blurred images containing additive white gaussian noise,” in *2013 IEEE International Conference on Industrial Technology (ICIT)*, 2013, pp. 1092–1097.
- [35] D. Sage, L. Donati, F. Soulez, D. Fortun, G. Schmit, A. Seitz, R. Guiet, C. Vonesch, and M. Unser, “Deconvolutionlab2: An open-source software for deconvolution microscopy,” *Methods*, vol. 115, pp. 28–41, 2017, image Processing for Biologists. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1046202316305096>
- [36] P. Bankhead, *Analyzing fluorescence microscopy images with ImageJ*, 06 2014.
- [37] J.-S. Lee, “Digital image smoothing and the sigma filter,” *Computer Vision, Graphics, and Image Processing*, vol. 24, no. 2, pp. 255–269, 1983, plugin: <https://imagej.net/plugins/sigma-filter>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0734189X83900476>
- [38] S. Sternberg, “Biomedical image processing,” *Computer*, vol. 16, no. 01, pp. 22–34, jan 1983, plugin: <https://imagej.net/plugins/rolling-ball-background-subtraction>.

- [39] K. Zuiderveld, *Contrast Limited Adaptive Histogram Equalization*. USA: Academic Press Professional, Inc., 1994, pp. 474–485, plugin: <https://imagej.net/plugins/clae>.
- [40] N. Otsu, “A threshold selection method from gray level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, 1979.
- [41] T. Qingzong, “Adaptivethreshold — imagej plugin.” [Online]. Available: <https://sites.google.com/site/qingzongtseng/adaptivethreshold>
- [42] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.
- [43] T. R. Singh, S. Roy, O. I. Singh, T. Sinam, and K. M. Singh, “A new local adaptive thresholding technique in binarization,” *ArXiv*, vol. abs/1201.5227, 2012.
- [44] J. Sauvola, T. Seppänen, S. Haapakoski, and M. Pietikäinen, “Adaptive document binarization.” vol. 33, 09 1997, pp. 147–152 vol.1.
- [45] P. sung Liao, T. sheng Chen, and P. choo Chung, “A fast algorithm for multilevel thresholding,” *Journal of Information Science and Engineering*, vol. 17, pp. 713–727, 2001.
- [46] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.
- [47] G. W. Zack, W. E. Rogers, and S. A. Latt, “Automatic measurement of sister chromatid exchange frequency.” *Journal of Histochemistry & Cytochemistry*, vol. 25, no. 7, pp. 741–753, 1977, pMID: 70454. [Online]. Available: <https://doi.org/10.1177/25.7.70454>
- [48] W. H. Richardson, “Bayesian-based iterative method of image restoration*,” *J. Opt. Soc. Am.*, vol. 62, no. 1, pp. 55–59, Jan 1972. [Online]. Available: <http://opg.optica.org/abstract.cfm?URI=josa-62-1-55>
- [49] L. B. Lucy, “An iterative technique for the rectification of observed distributions,” *Astronomical Journal*, vol. 79, p. 745, Jun. 1974. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/1974AJ.....79..745L>
- [50] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, “Finding a ”kneedle” in a haystack: Detecting knee points in system behavior,” in *2011 31st International Conference on Distributed Computing Systems Workshops*, 2011, pp. 166–171.
- [51] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, “Richardson-lucy algorithm with total variation regularization for 3d confocal microscope deconvolution,” *Microscopy Research*

and Technique, vol. 69, no. 4, pp. 260–266, 2006. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jemt.20294>

Appendix A

Threshold outcomes for method selection

The outcomes for the global and local thresholding methods being shortlisted with the sample images shown below in Figure A.1.

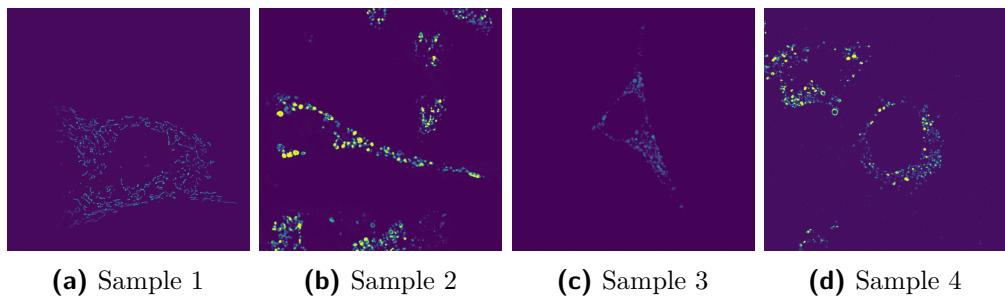


Figure A.1: Central slices of the pre-processed shortlisting samples prior to any thresholding

A.1. Global threshold outcomes of shortlist samples

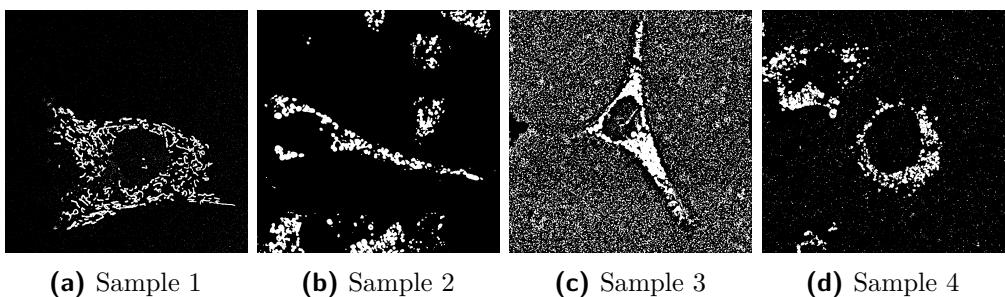


Figure A.2: Huang threshold outcomes for the shortlist sample images

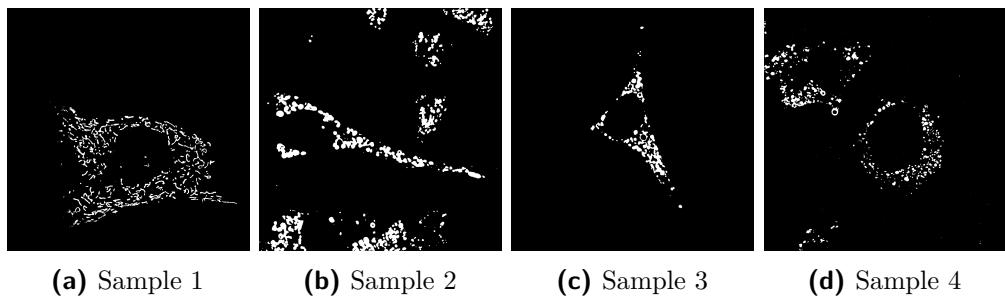


Figure A.3: Huang2 threshold outcomes for the shortlist sample images

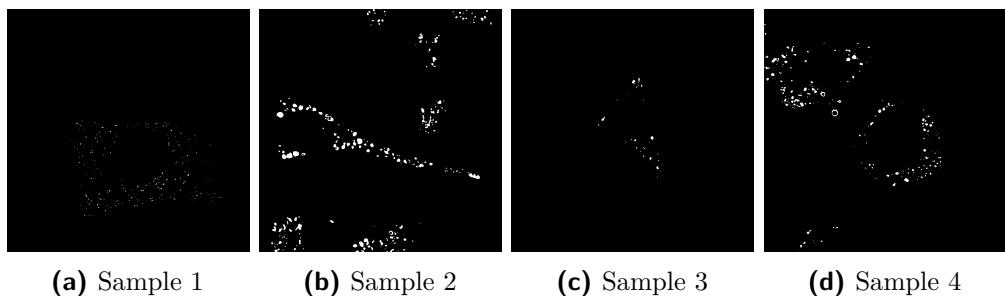


Figure A.4: Intermodes threshold outcomes for the shortlist sample images

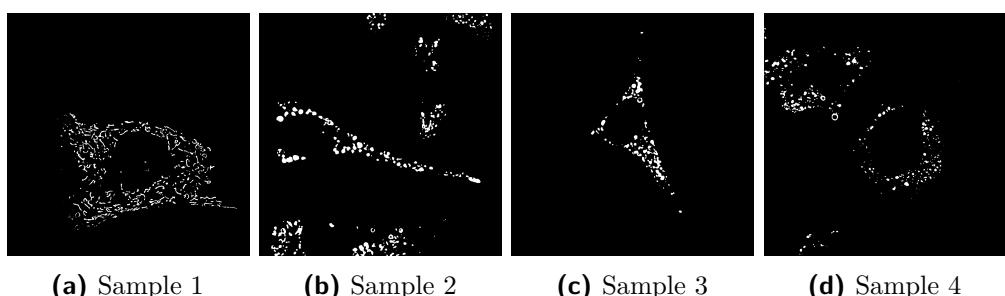


Figure A.5: IsoData threshold outcomes for the shortlist sample images

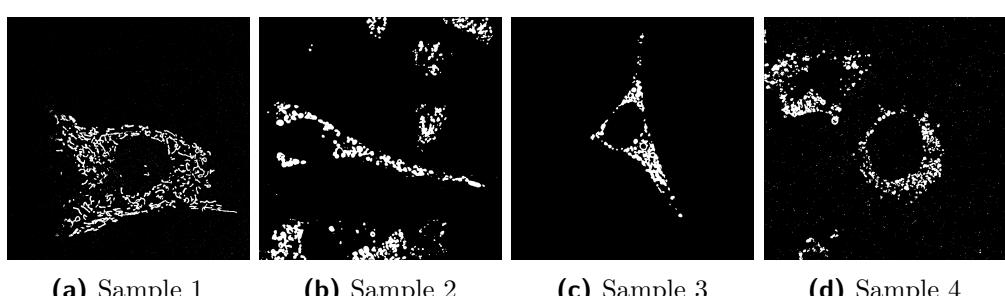


Figure A.6: Li threshold outcomes for the shortlist sample images

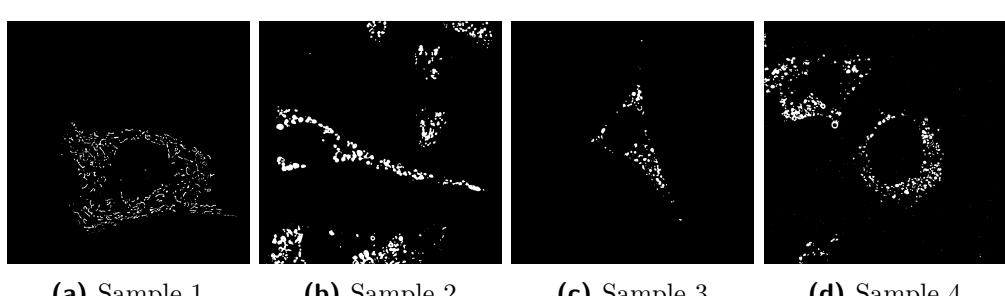


Figure A.7: MaxEntropy threshold outcomes for the shortlist sample images

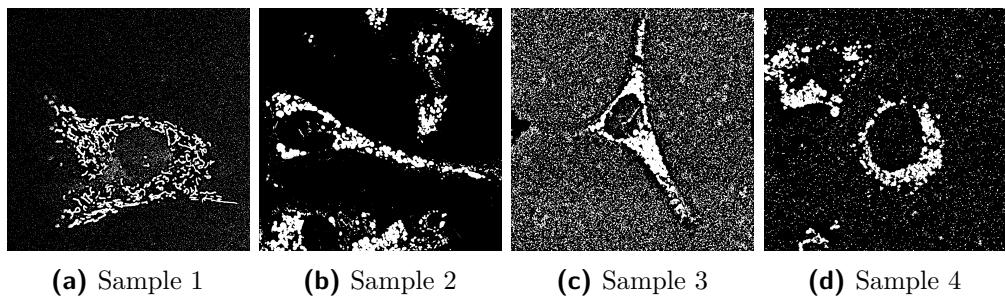


Figure A.8: Mean threshold outcomes for the shortlist sample images

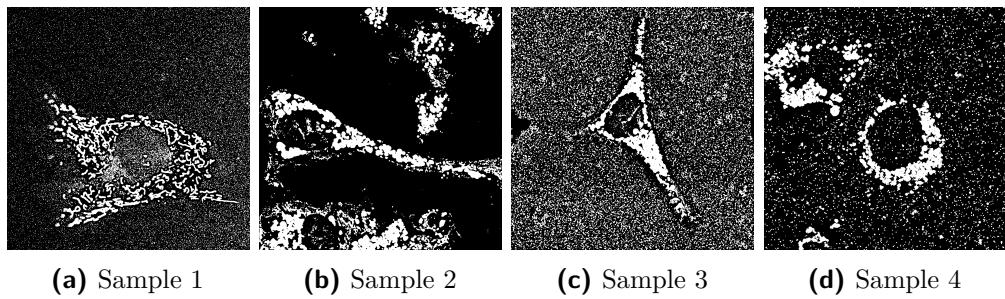


Figure A.9: MinError(I) threshold outcomes for the shortlist sample images

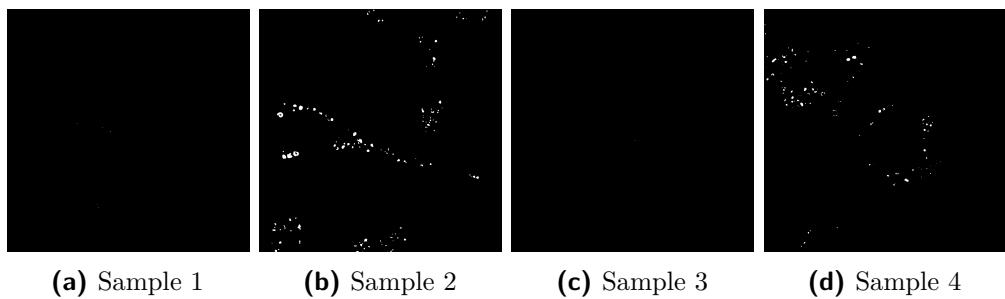


Figure A.10: Minimum threshold outcomes for the shortlist sample images

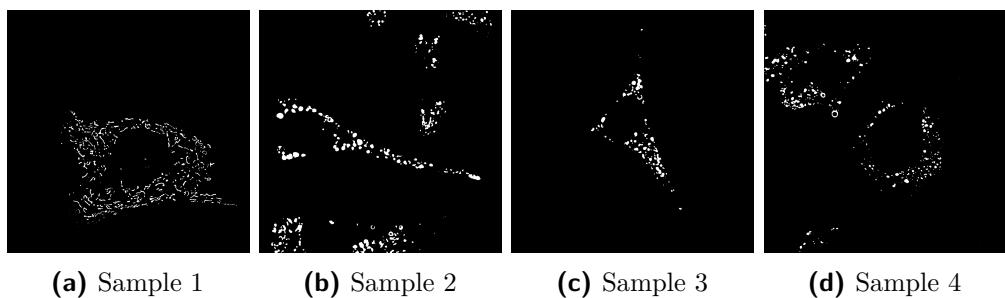


Figure A.11: Moments threshold outcomes for the shortlist sample images

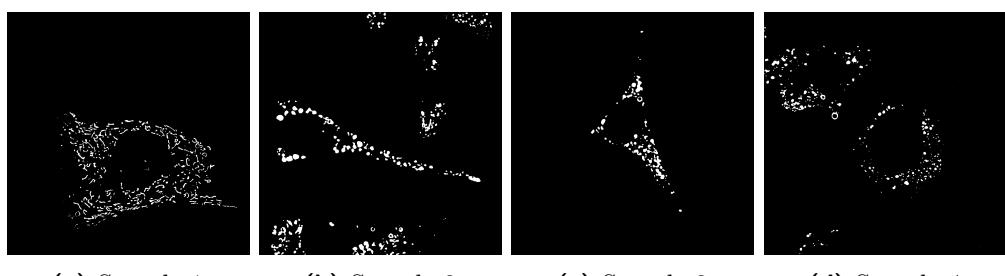


Figure A.12: Otsu threshold outcomes for the shortlist sample images

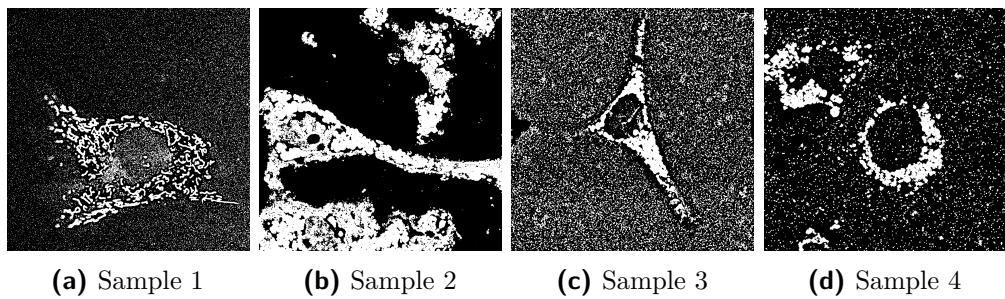


Figure A.13: Percentile threshold outcomes for the shortlist sample images

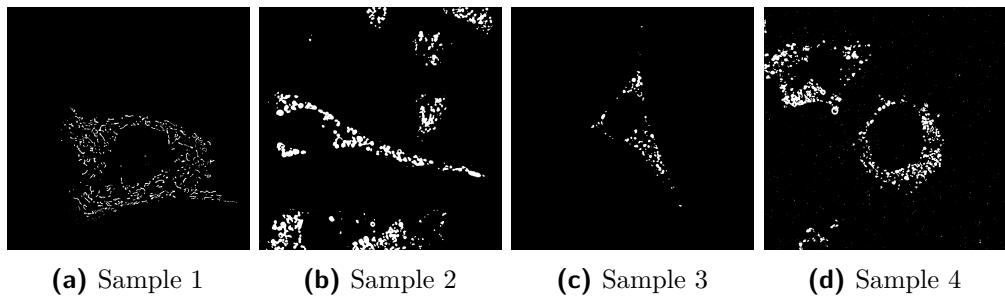


Figure A.14: RenyiEntropy threshold outcomes for the shortlist sample images

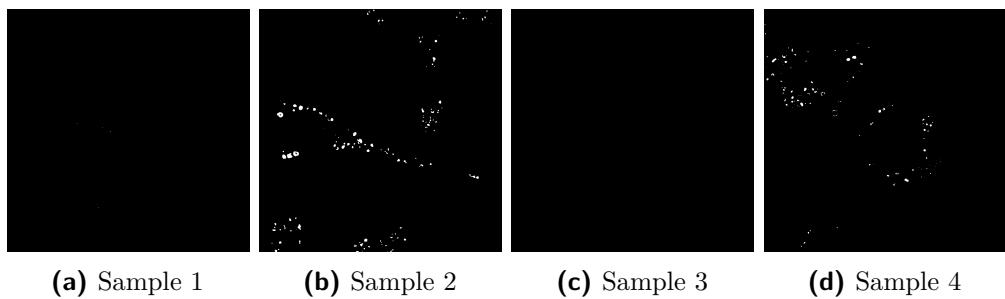


Figure A.15: Shanbhag threshold outcomes for the shortlist sample images

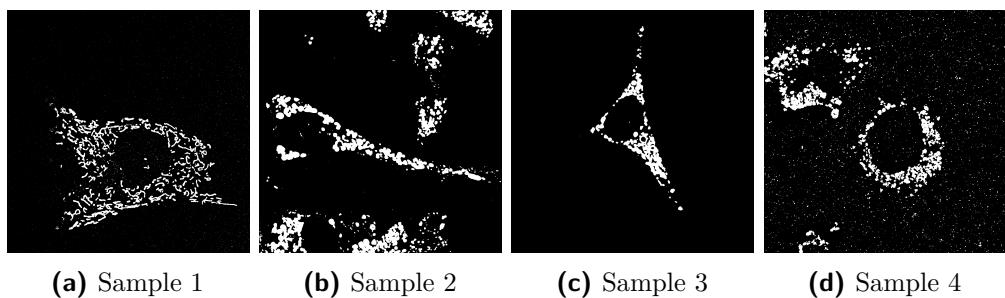


Figure A.16: Triangle threshold outcomes for the shortlist sample images

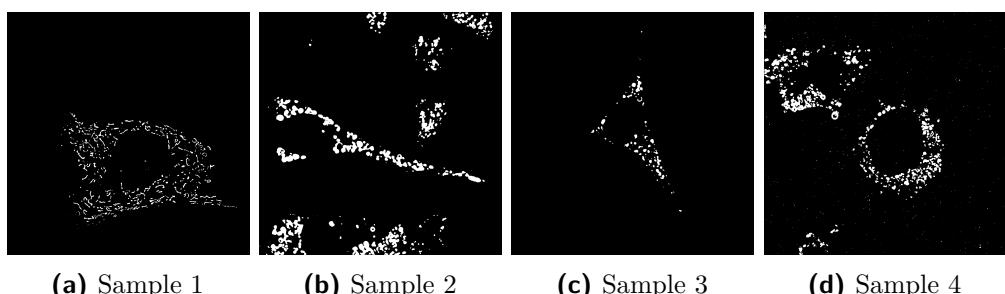


Figure A.17: Yen threshold outcomes for the shortlist sample images

A.2. Local threshold outcomes of shortlist samples

The outcomes of the local threshold methods will also depict the impact of varied window size of 15, 60, and 100 while certain methods also depict the impact secondary parameter tuning has on the outcome using a RGB range for the resultant image. For the RGB images, colour combinations are yellow (both red and green), magenta (both red and blue), cyan (both blue and green), and white (all three colours). When colour combinations depict certain regions it implies that said region is threshold as foreground by the thresholding method for both parameter variations matching the base colours. Table A.1 shows the mappings between the secondary tuning parameter values and the representative colour channel for the RGB images of the given local thresholding methods.

| Threshold method | Red channel | Green channel | Blue channel |
|------------------|-------------|---------------|--------------|
| Bernsen | 10 | 15 | 20 |
| Mean | 0 | -5 | 5 |
| Median | 0 | -5 | 5 |
| MidGrey | 0 | -5 | 5 |
| Niblack | 0.1 | 0.2 | 0.3 |
| Phansalkar | 0.25 | 0.15 | 0.35 |
| Sauvola | 0.5 | 0.4 | 0.6 |

Table A.1: Mappings between colour channel and secondary parameter values for the local thresholding methods shown.

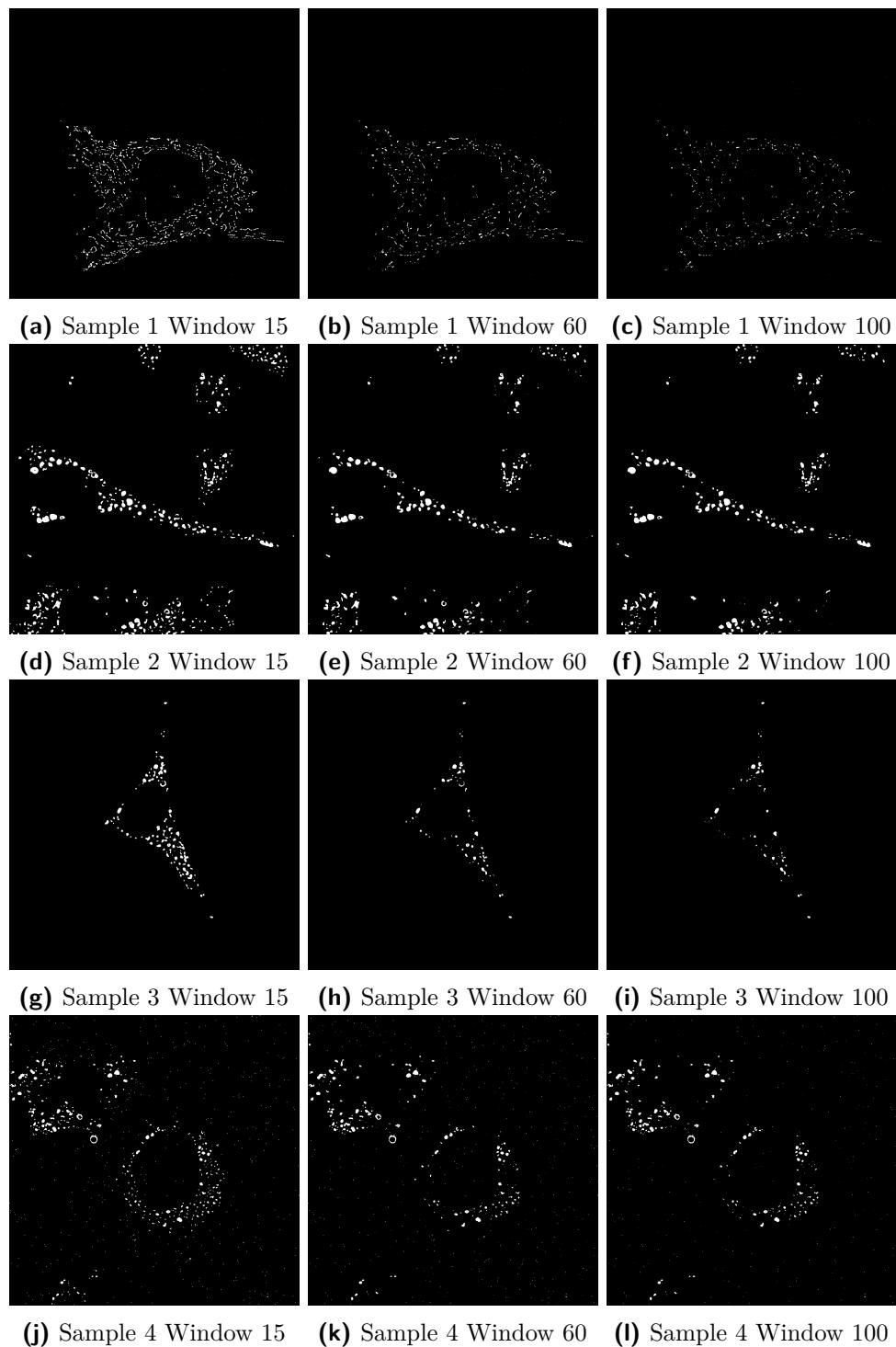


Figure A.18: Bernsen threshold binarizations of the shortlist samples at different window sizes.

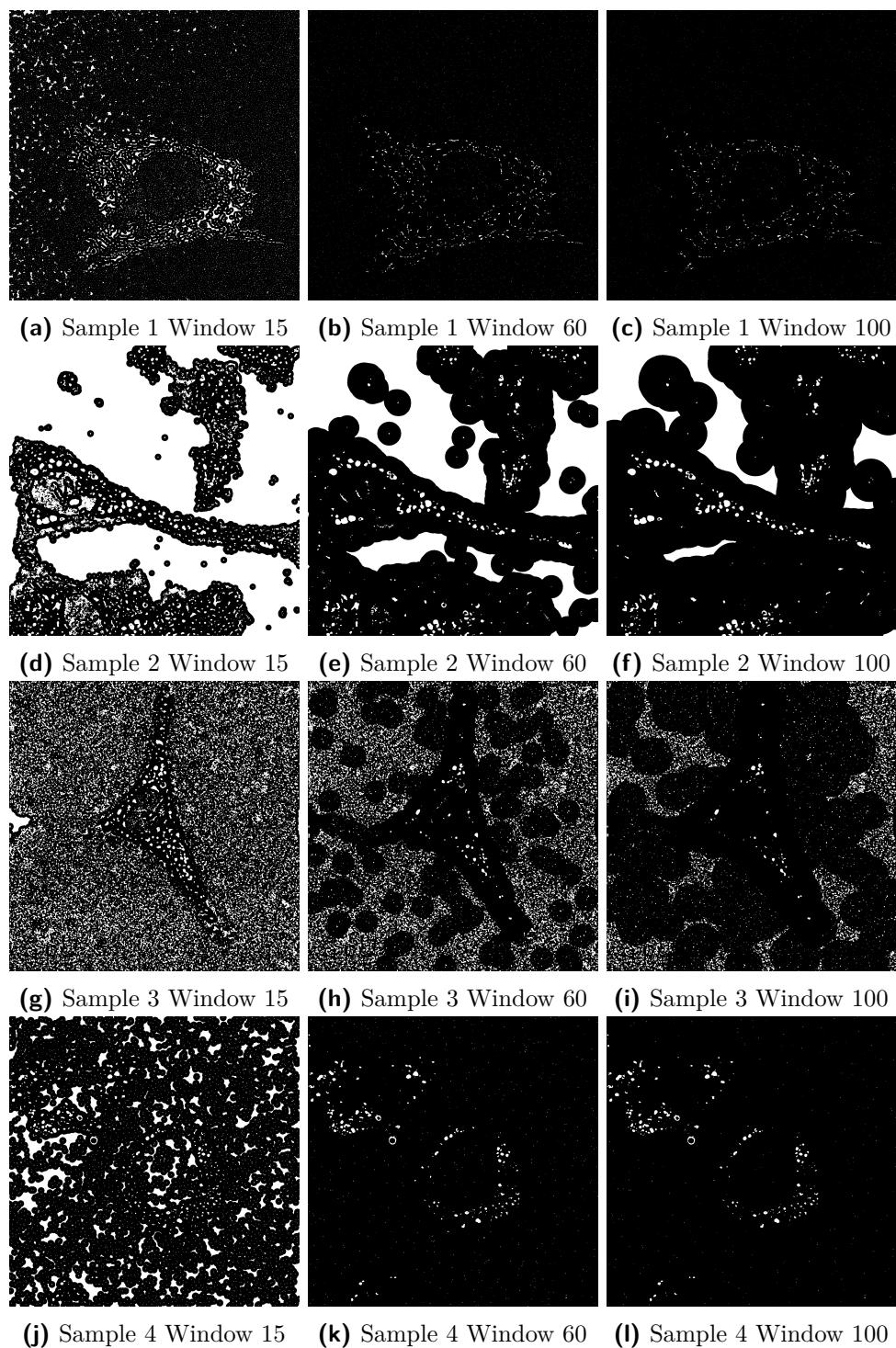


Figure A.19: Contrast binarizations of the shortlist samples at different window sizes.

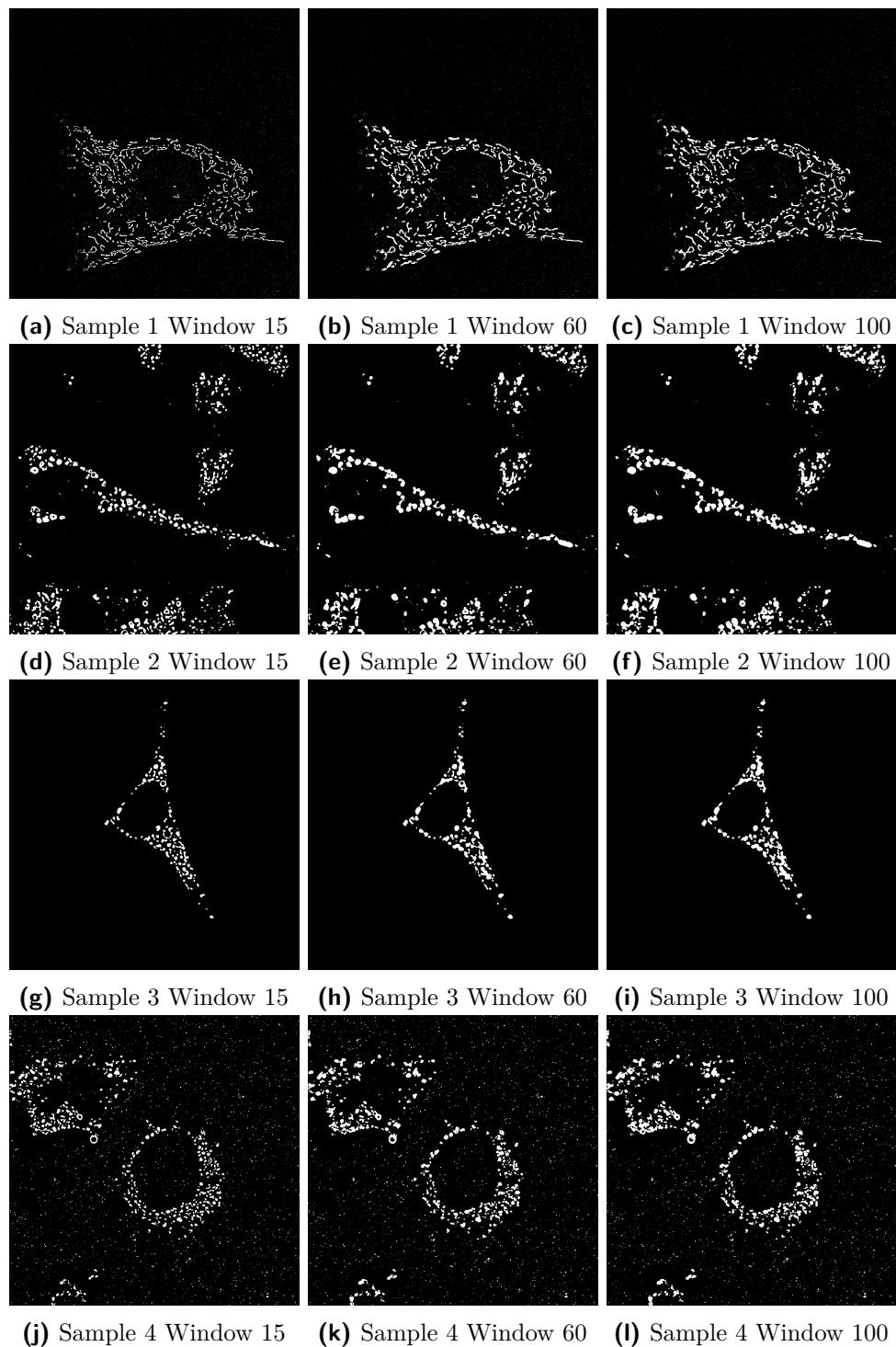


Figure A.20: Mean binarizations of the shortlist samples at different window sizes.

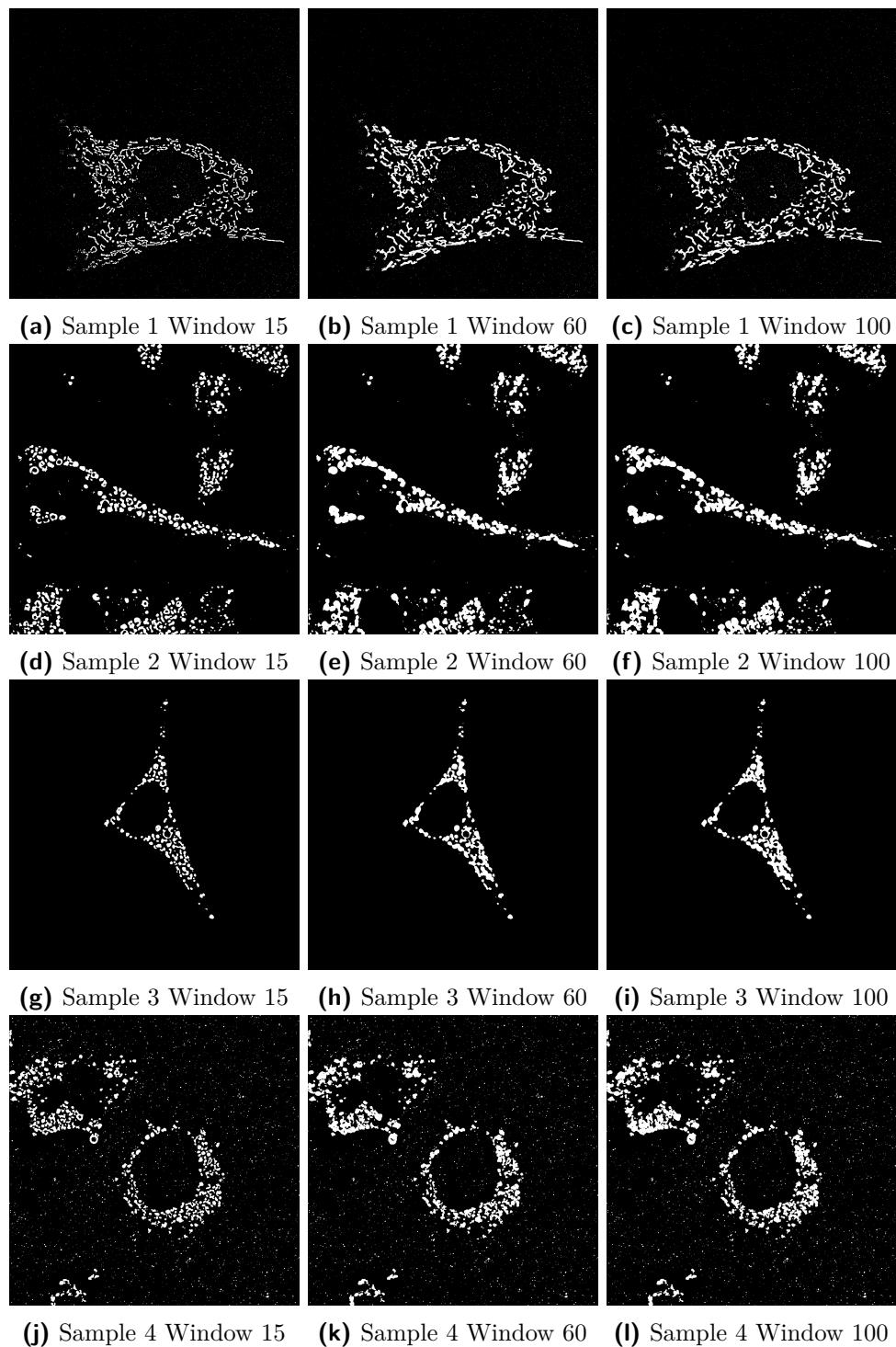


Figure A.21: Median binarizations of the shortlist samples at different window sizes.

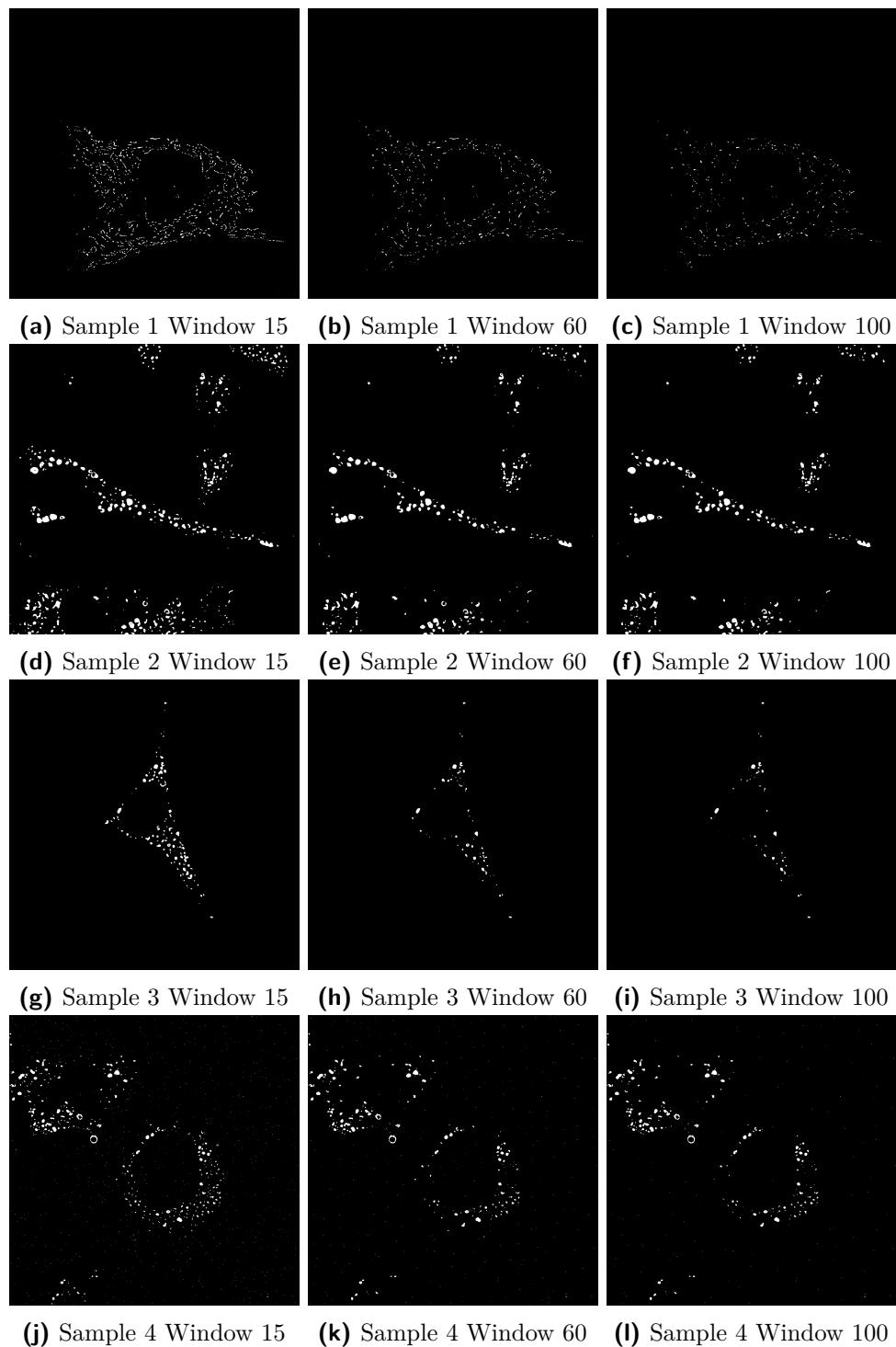


Figure A.22: MidGrey binarizations of the shortlist samples at different window sizes.

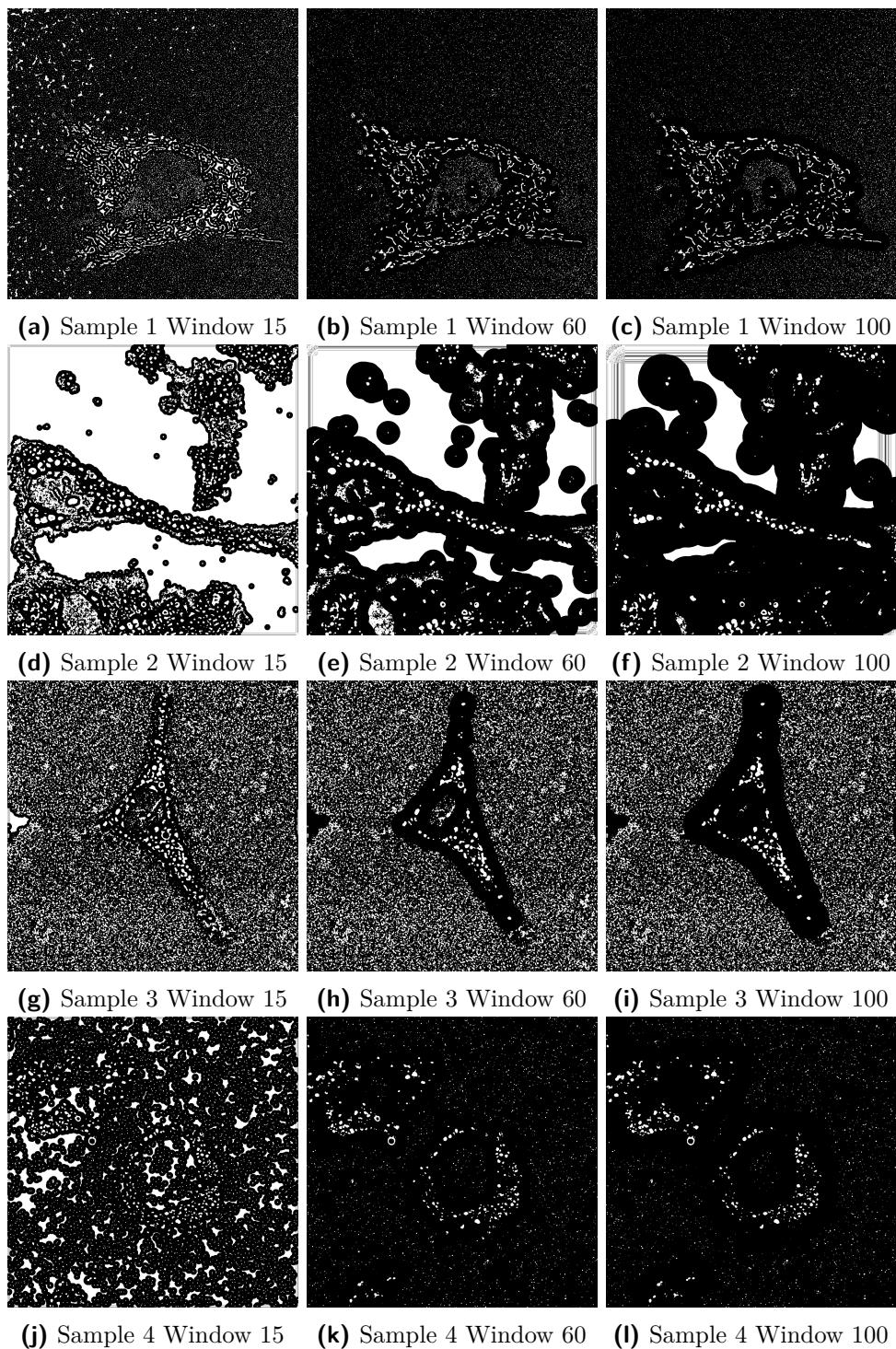


Figure A.23: Otsu binarizations of the shortlist samples at different window sizes.

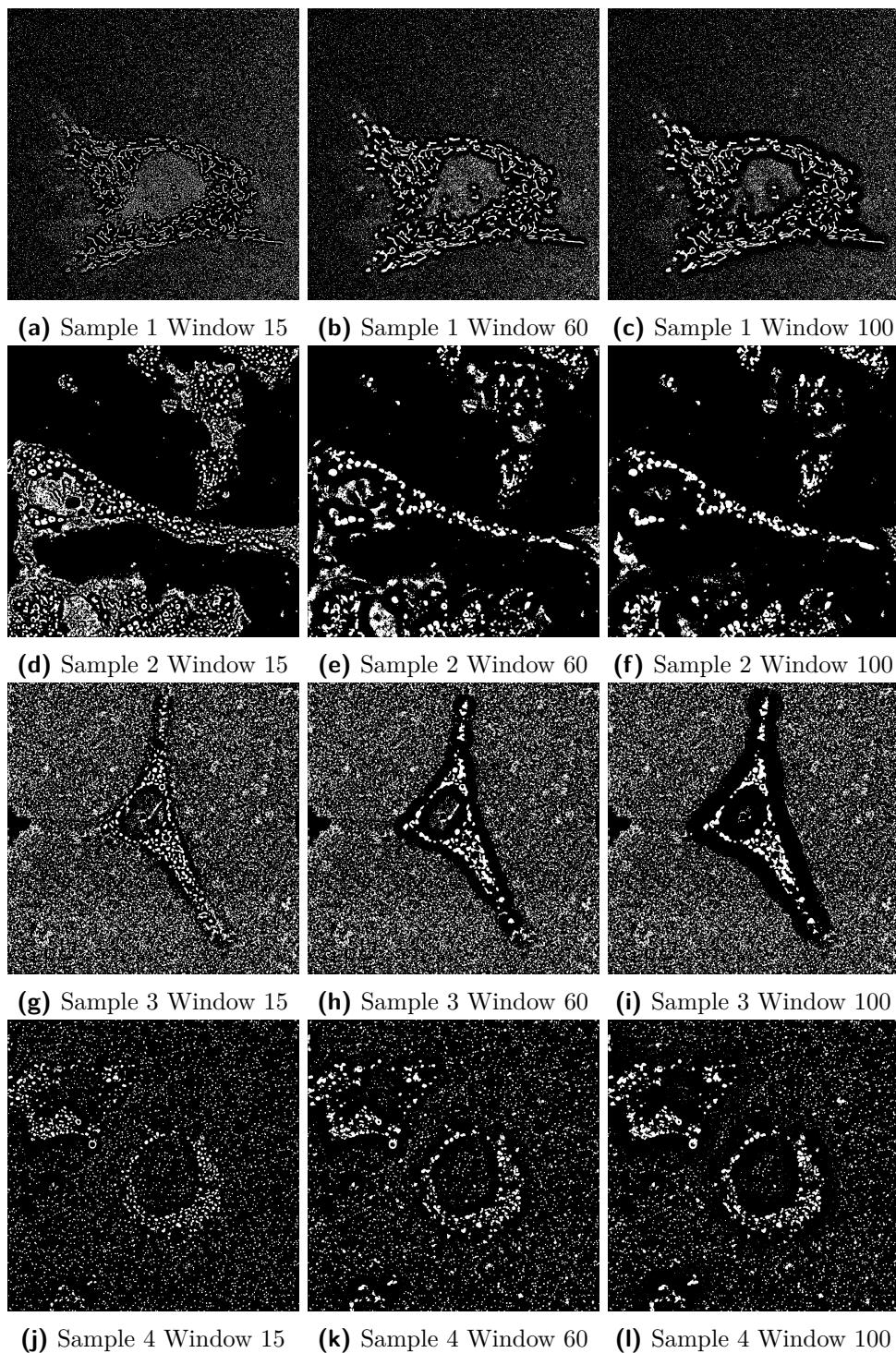


Figure A.24: Niblack binarizations of the shortlist samples at different window sizes.

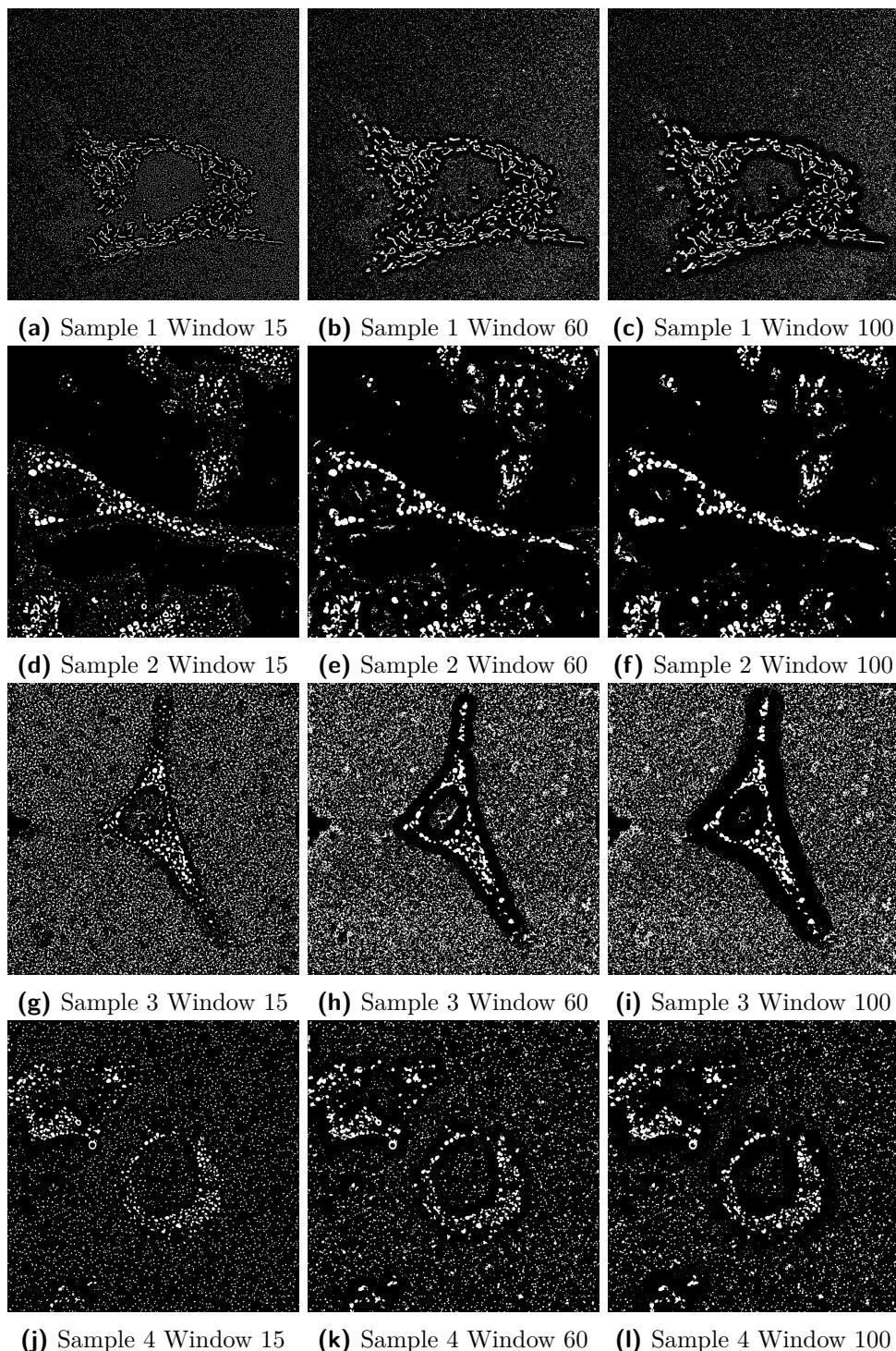


Figure A.25: Phansalkar binarizations of the shortlist samples at different window sizes.

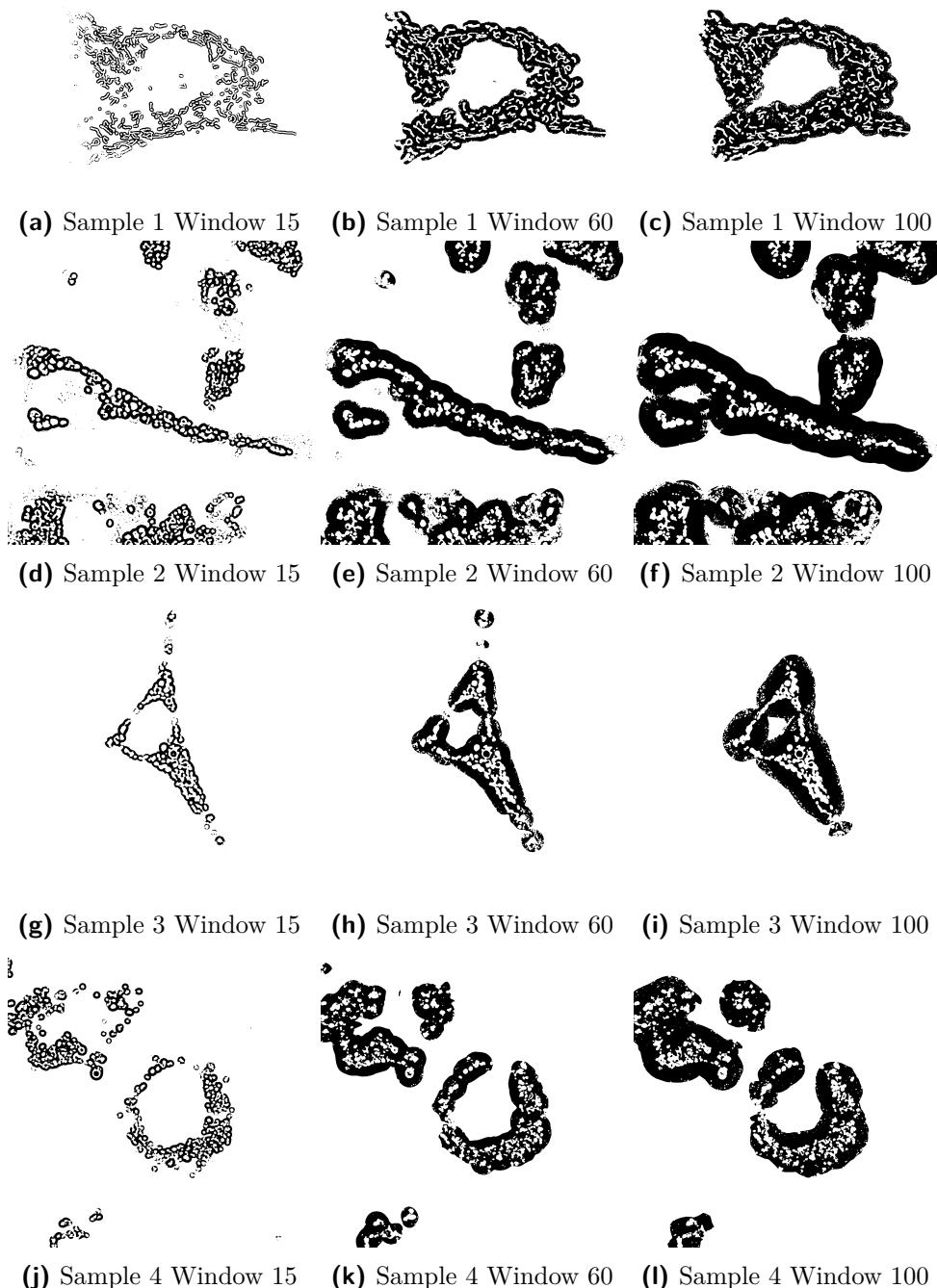


Figure A.26: Sauvola binarizations of the shortlist samples at different window sizes.

Appendix B

Visual analysis thresholding outcome images

Image panels for each thresholding method with each image in the panel stipulating the imaged organelle (lysosome, autophagosome, mitochondria) and the original sample (A-C). All binarized images produced through the thresholding methods are depicted as depth MIP images. Depth MIP images are images that have been flattened from 3D to 2D via MIP but the pixel colours denote the which slice along the 3D image depth it originates from. This is to provide a simple visualisation of the entire specimen while preserving some of the depth information. The sample images before any thresholding are showcased in Figure B.1, as MIP images with contrast enhancement applied. The contrast enhancement is purely for visualisation and was not employed for any thresholding. The Hysteresis baseline binarized images are shown in Figure B.2 as a depth MIP.

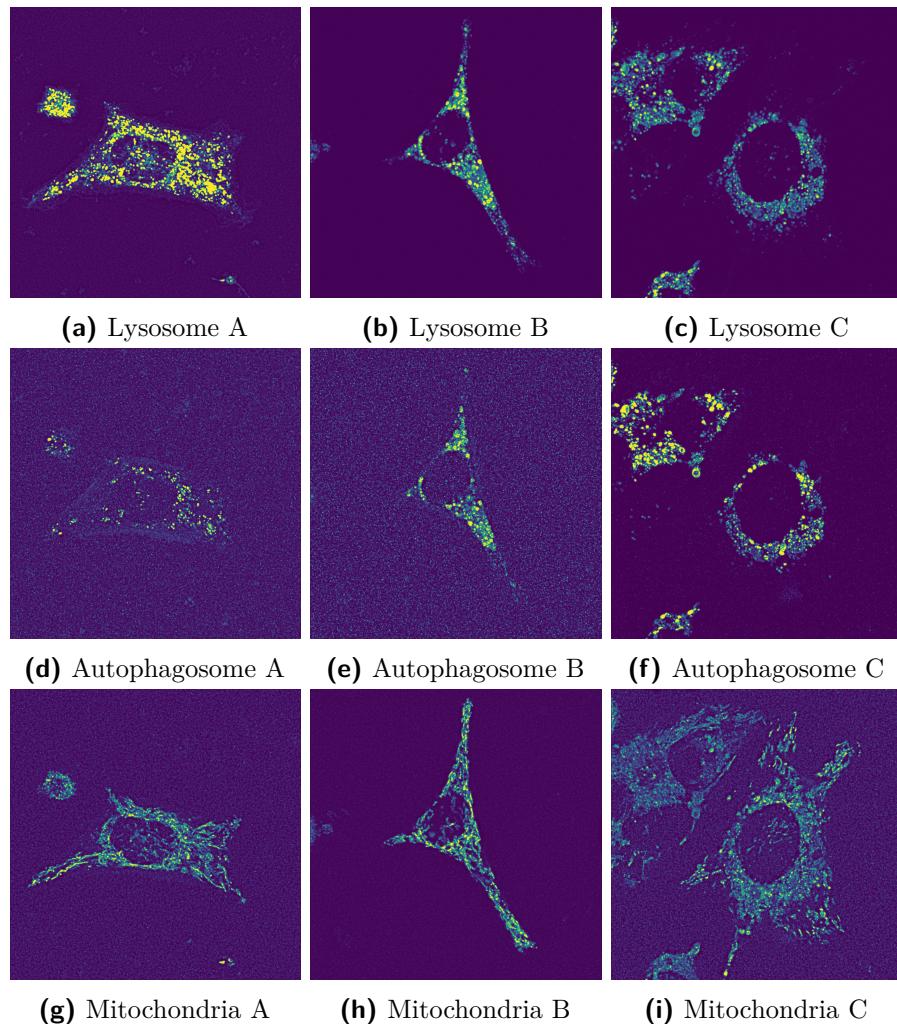


Figure B.1: Panel of the sample images, prior to any thresholding, used in the visual analysis with contrast enhancement applied to better visualise any noise. The panel rows correspond with the imaged organelle types and the columns the different sample images.

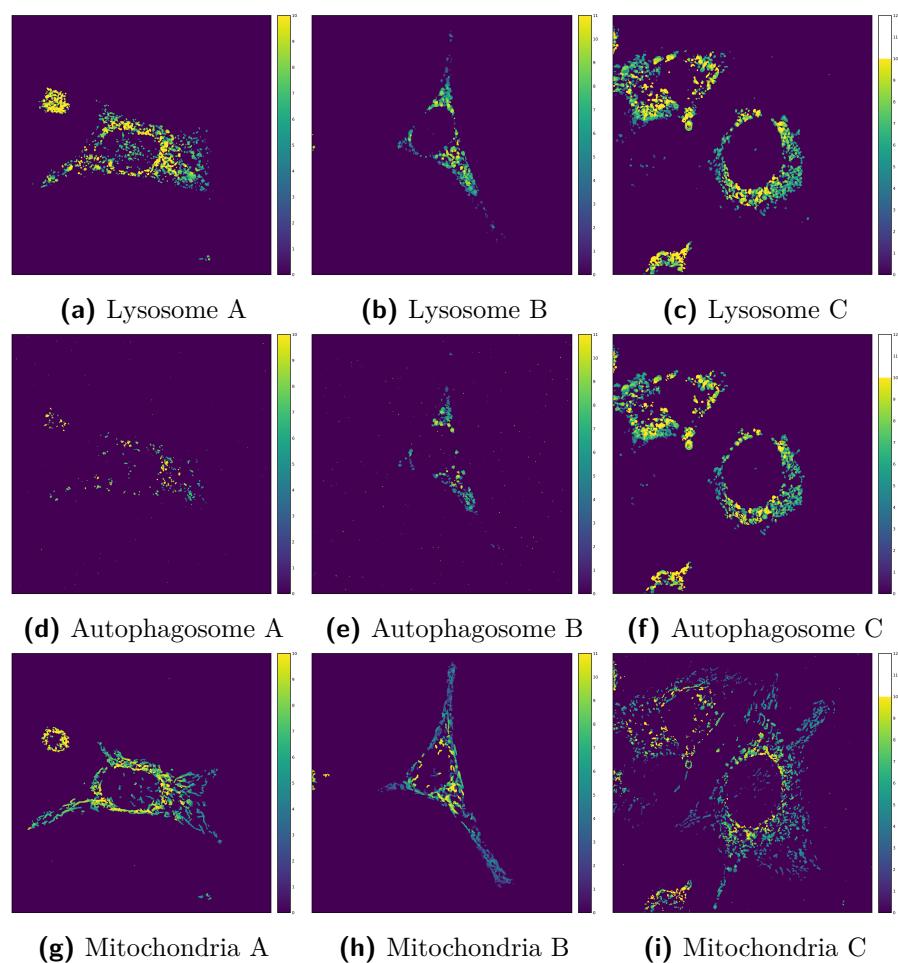


Figure B.2: Panel of the Hysteresis thresholding baseline images with organelle types separated by row.

B.1. Global threshold outcomes

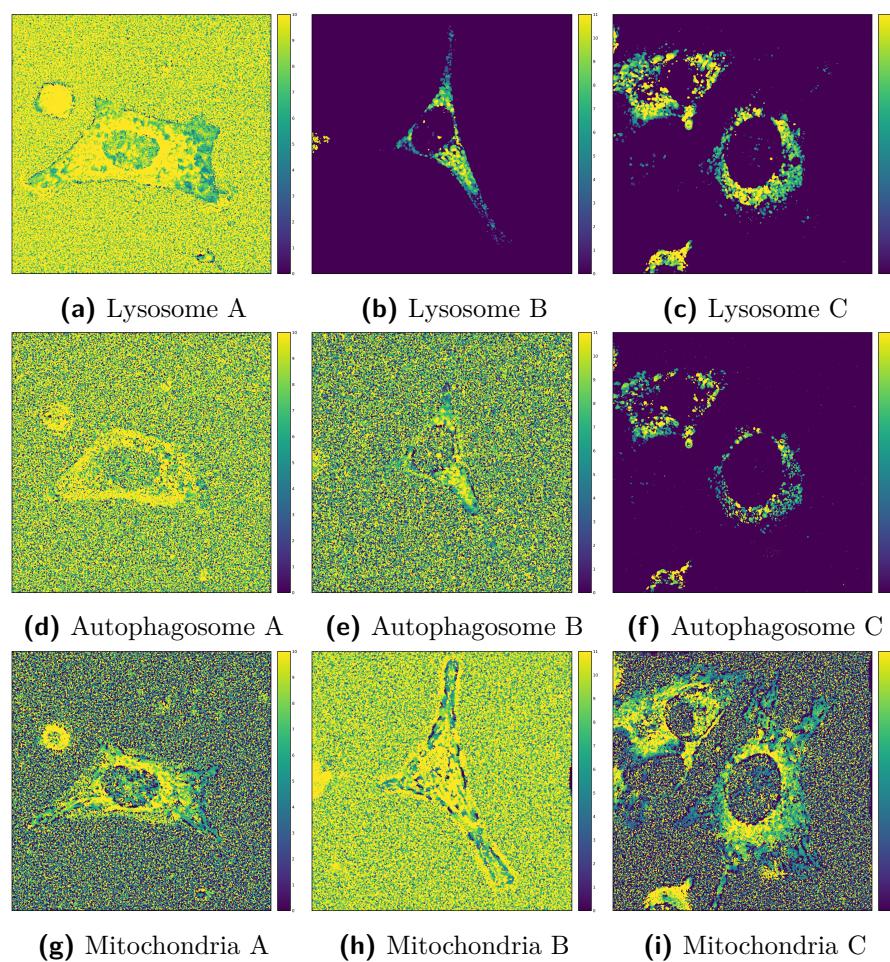


Figure B.3: Panel of Huang2 threshold outcomes

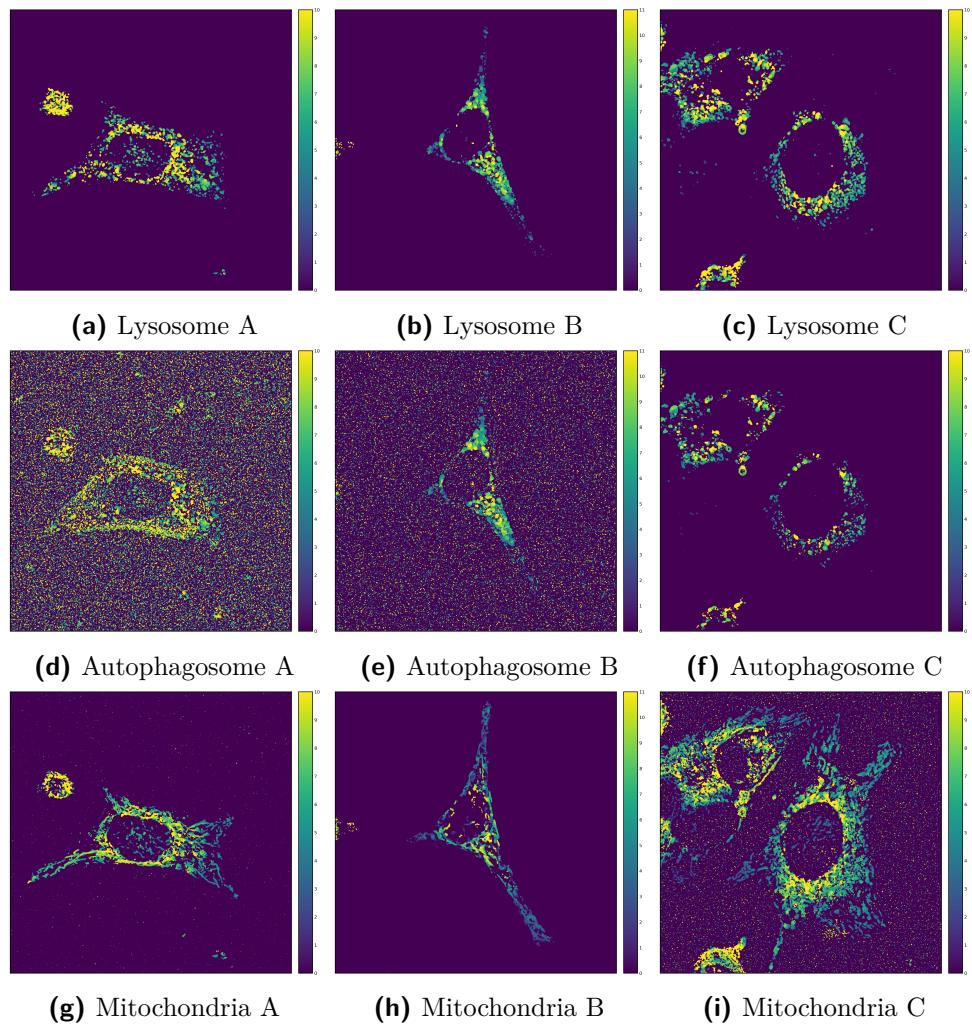


Figure B.4: Panel of IsoData threshold outcomes

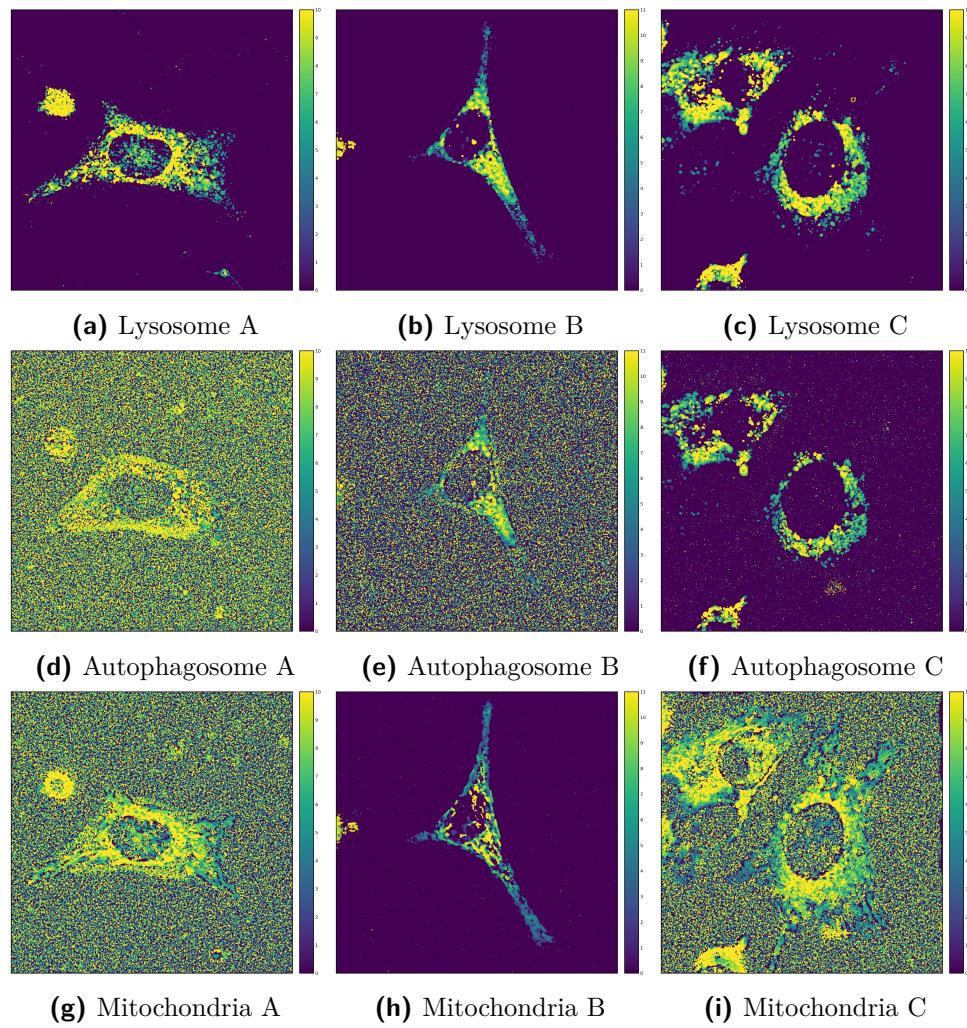


Figure B.5: Panel of Li threshold outcomes

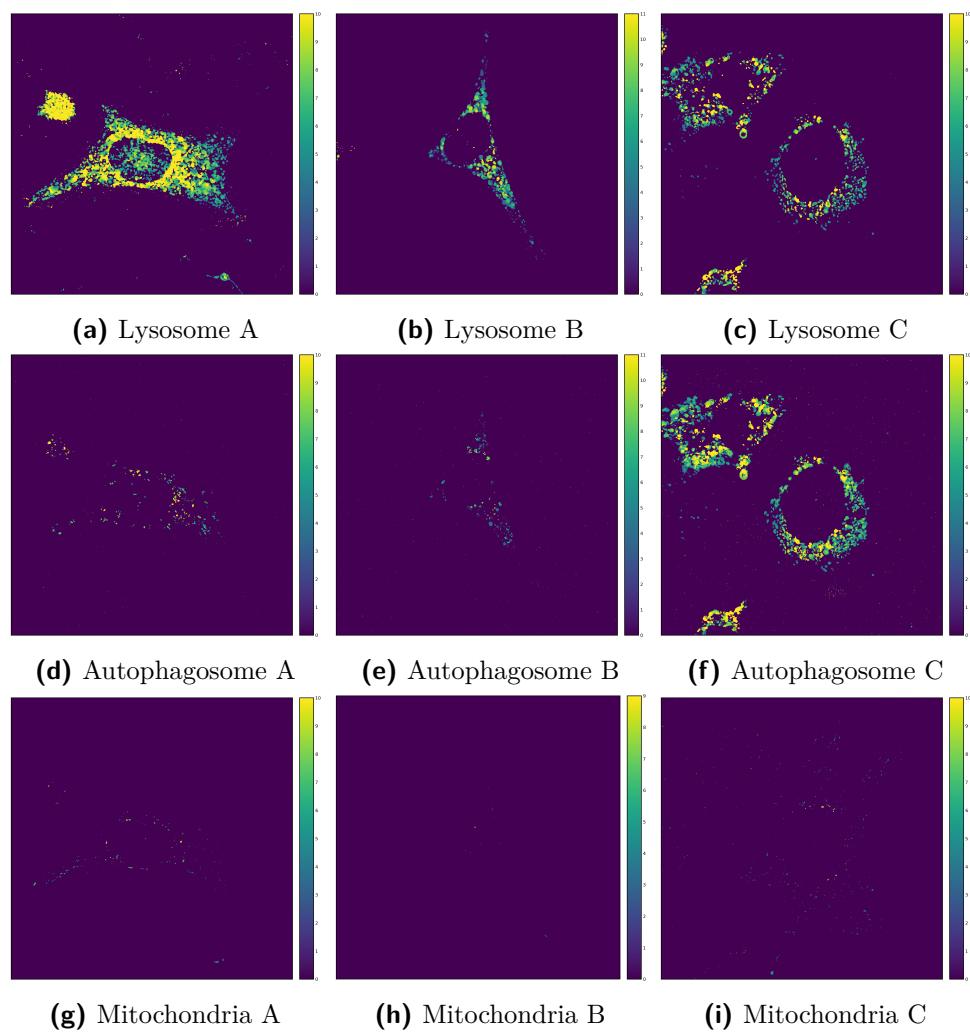


Figure B.6: Panel of MaxEntropy threshold outcomes

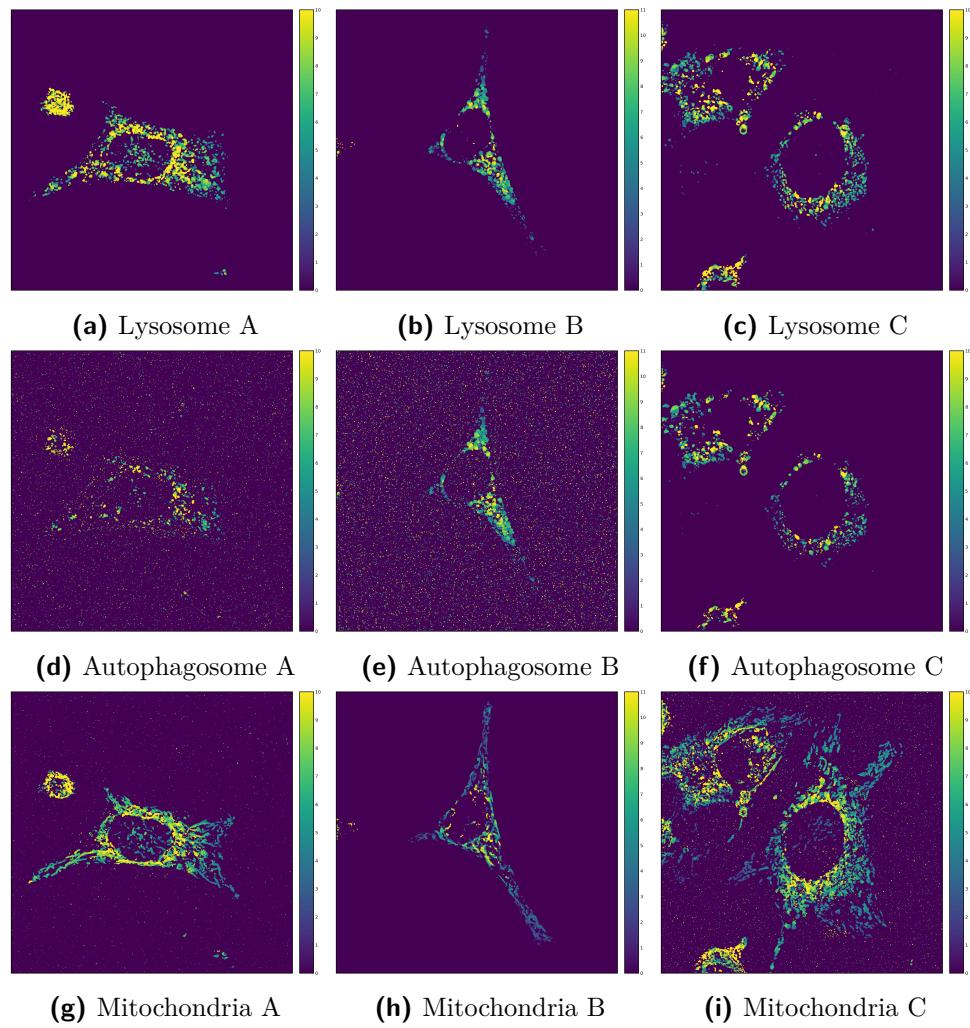


Figure B.7: Panel of Moments threshold outcomes

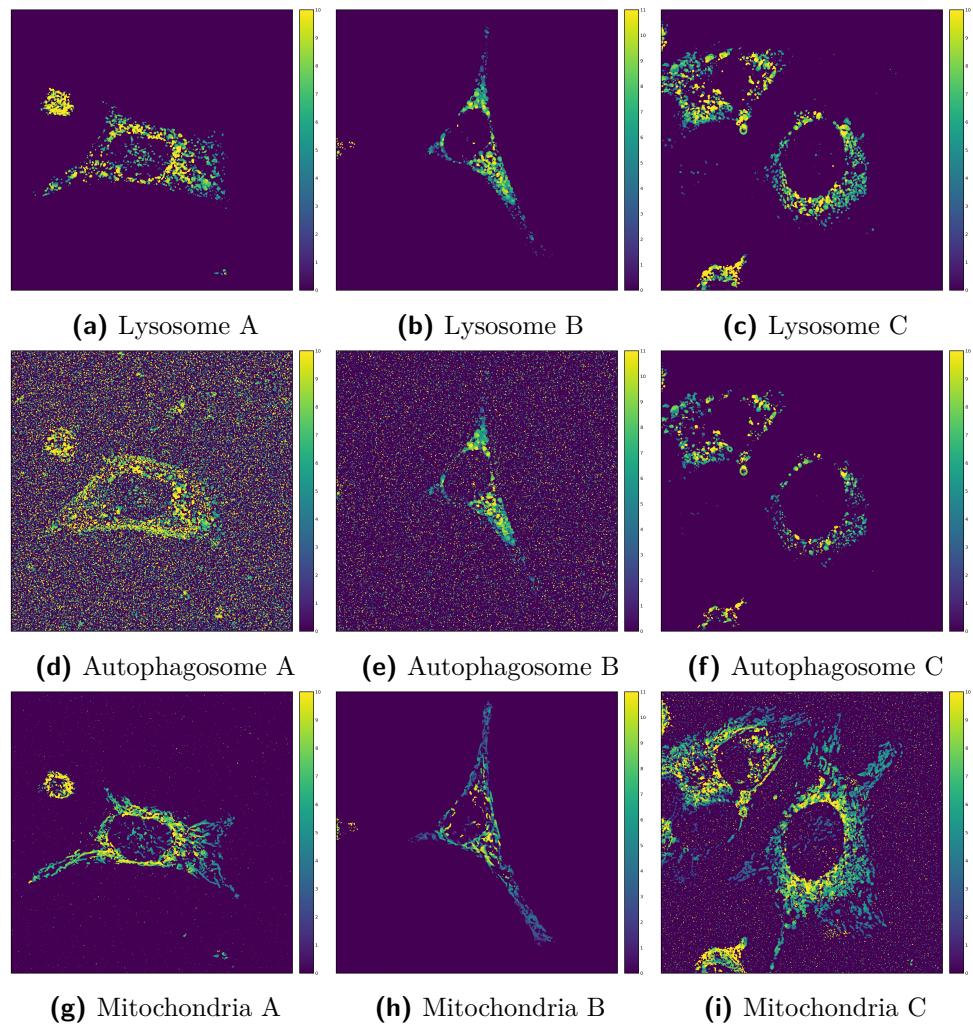


Figure B.8: Panel of Otsu threshold outcomes

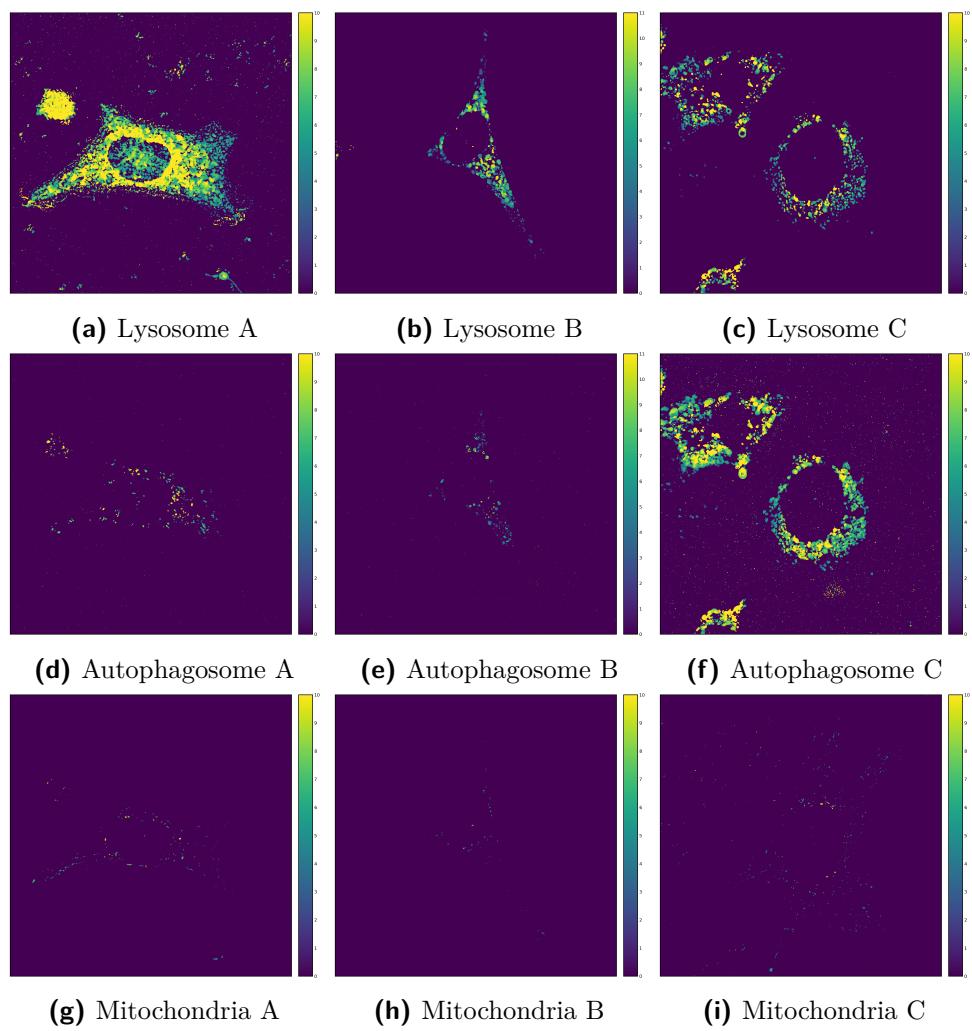


Figure B.9: Panel of RenyiEntropy threshold outcomes

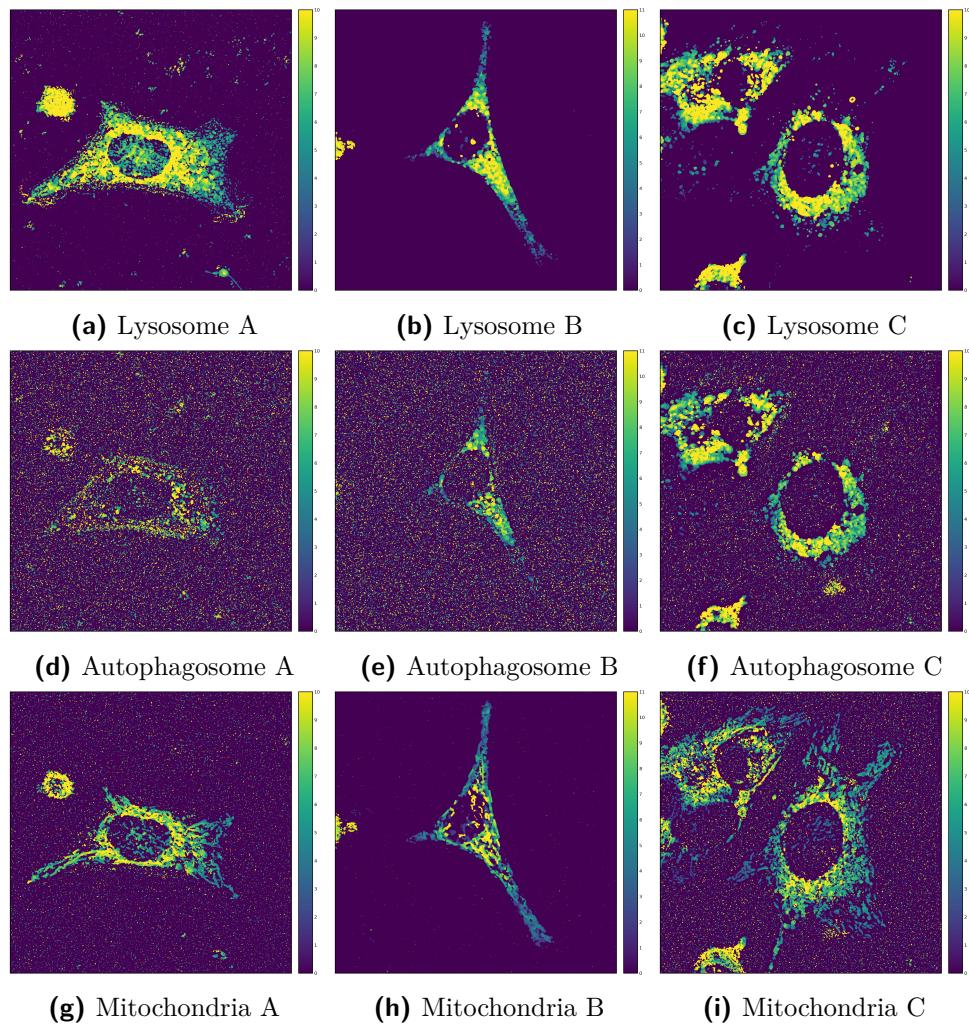


Figure B.10: Panel of Triangle threshold outcomes

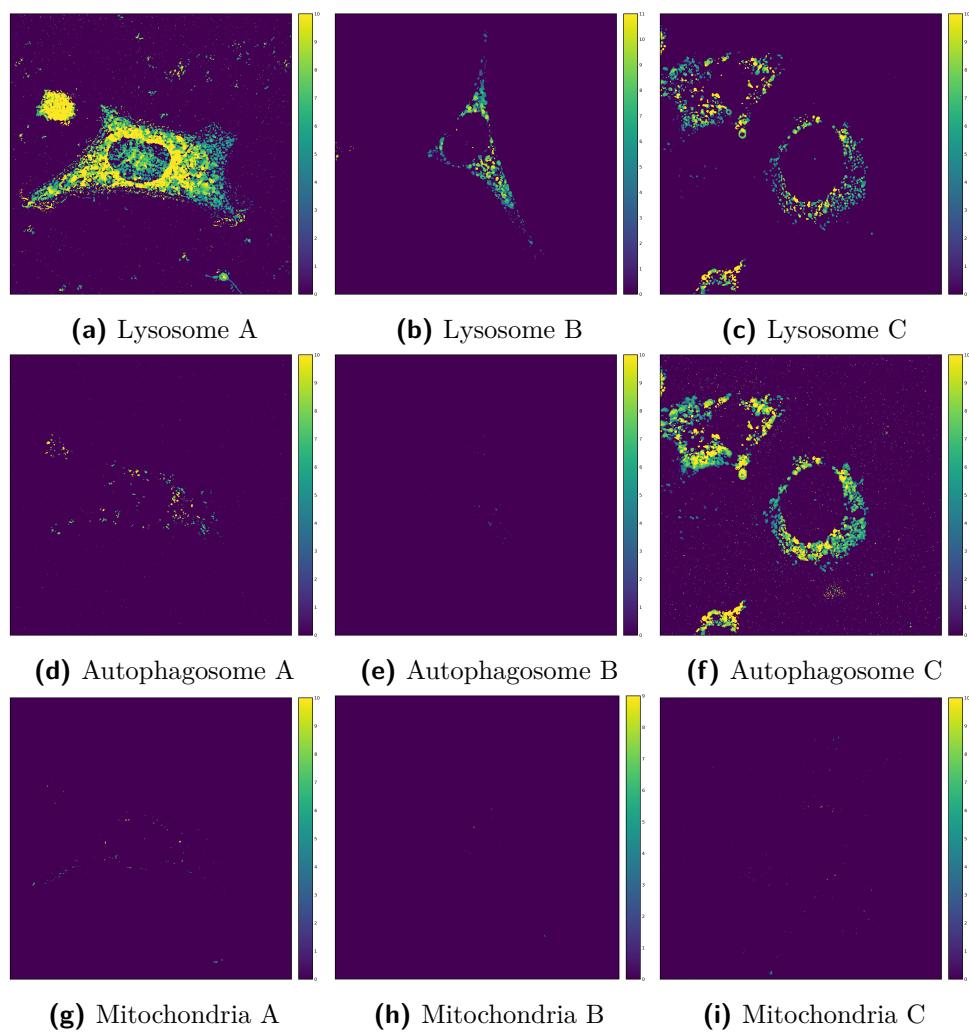


Figure B.11: Panel of Yen threshold outcomes

B.2. Local threshold outcomes

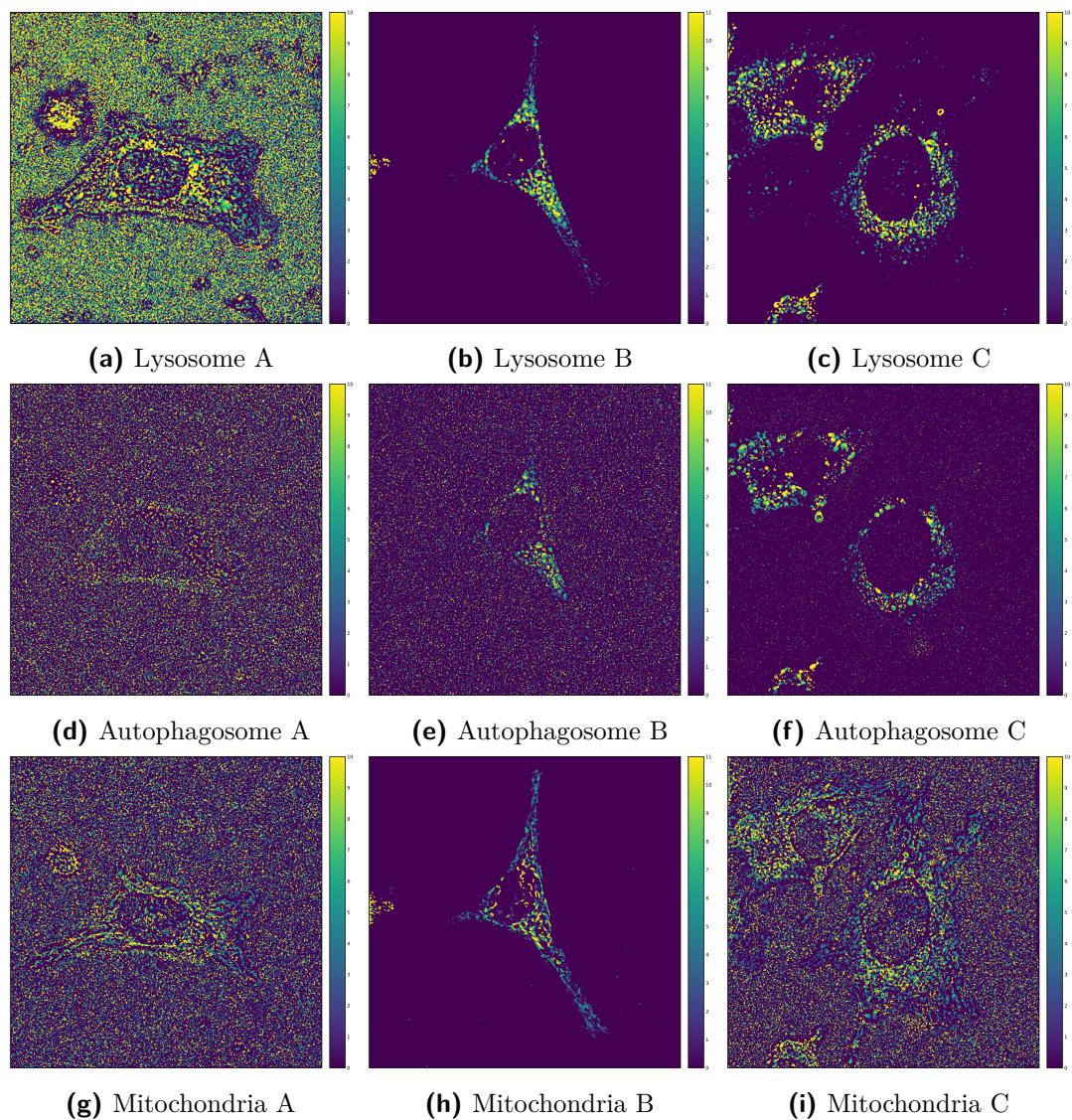


Figure B.12: Panel of Bernsen threshold outcomes

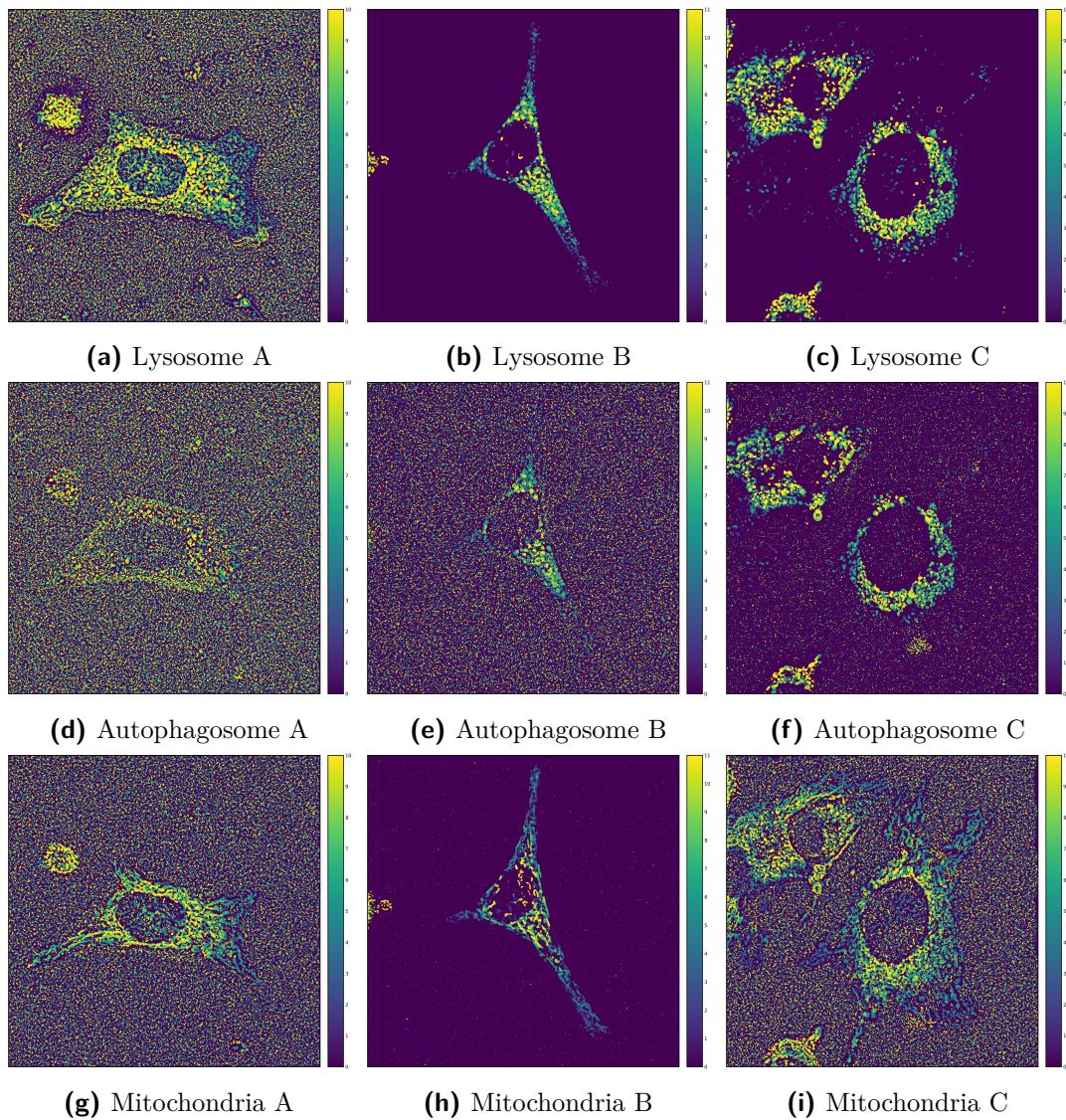


Figure B.13: Panel of Mean threshold outcomes

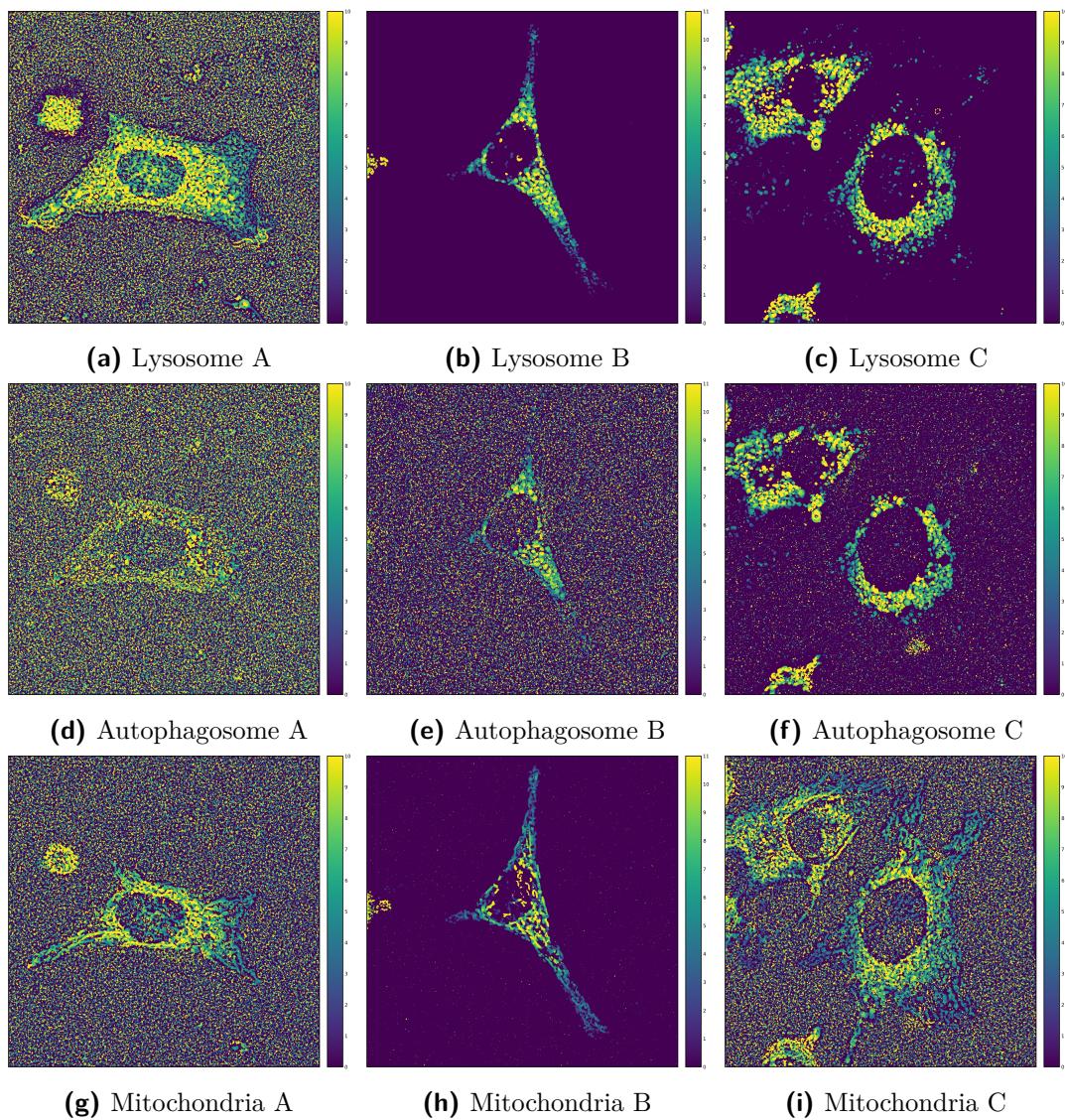


Figure B.14: Panel of Median threshold outcomes

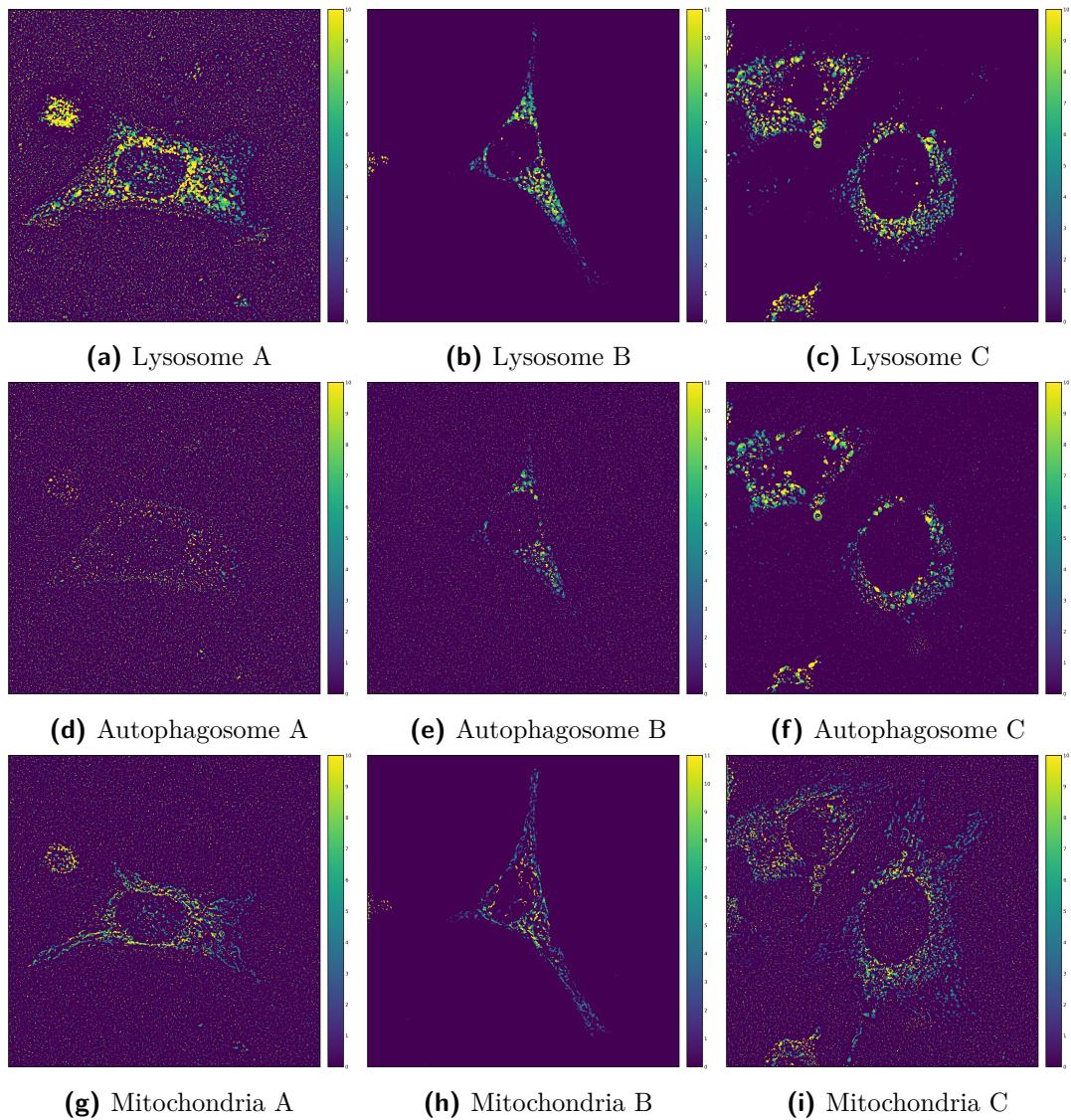


Figure B.15: Panel of MidGrey threshold outcomes

B.3. AHT outcomes

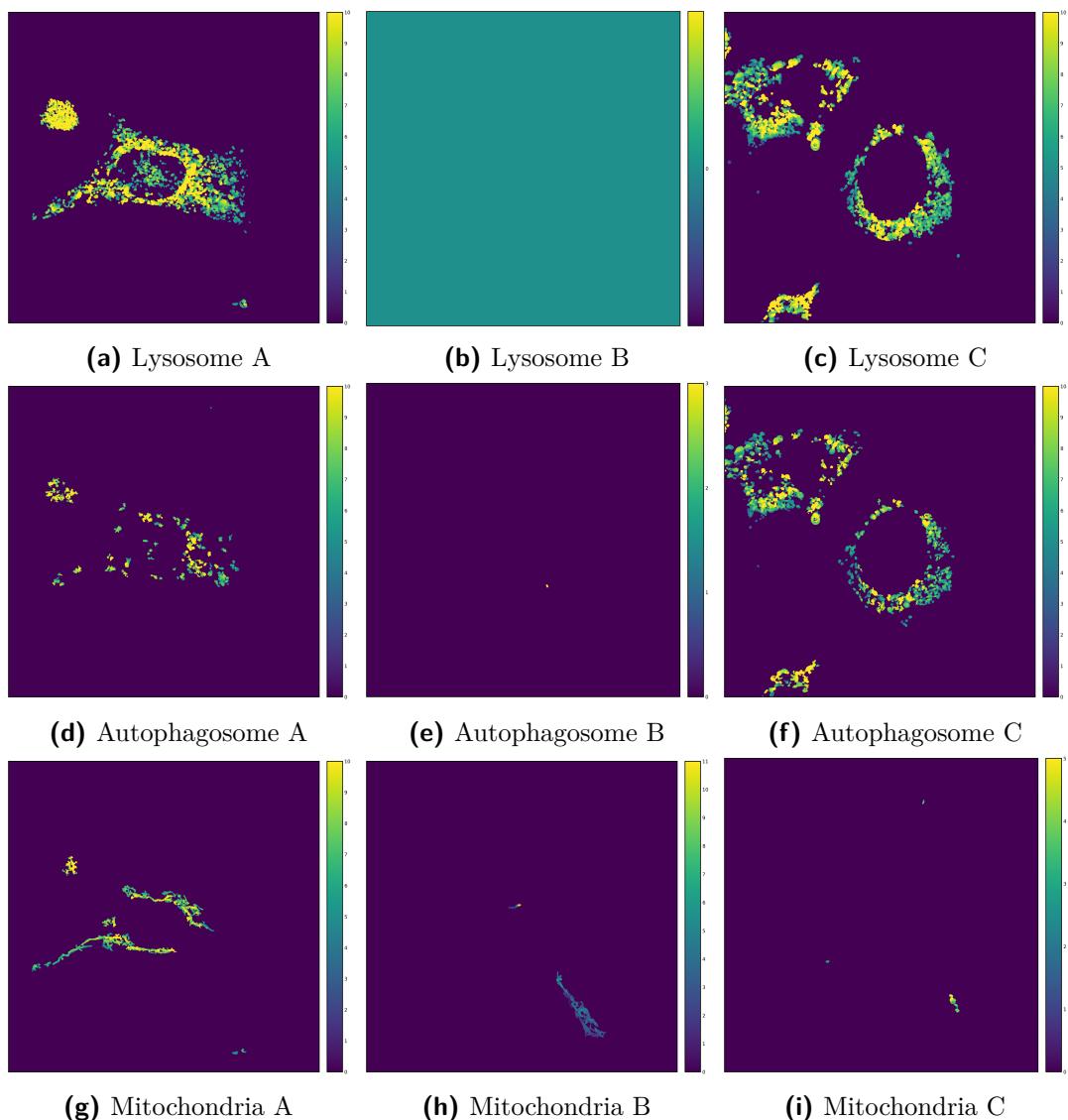


Figure B.16: MIP image panel of AHT outcomes with neither IHH or window biases applied

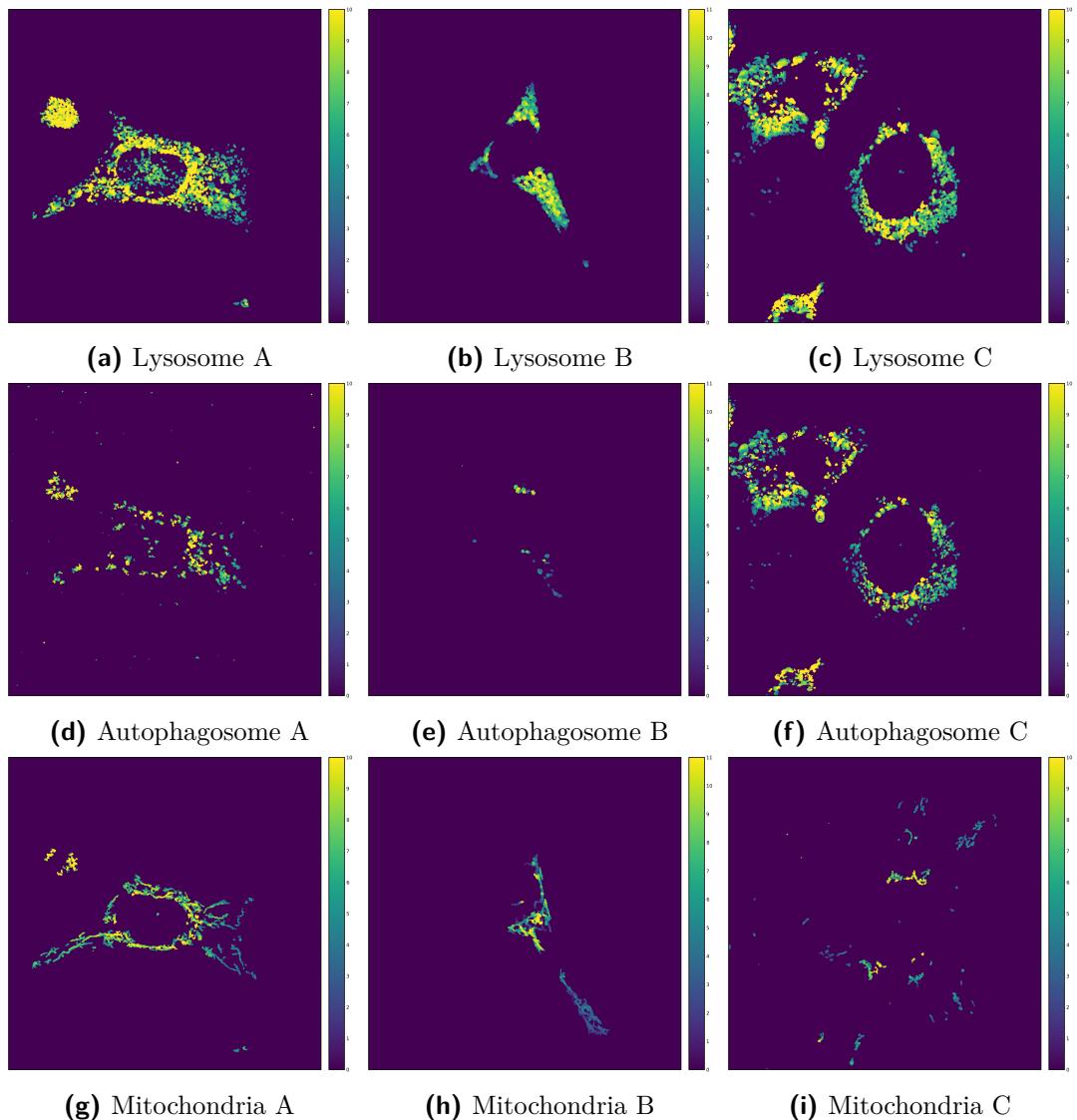


Figure B.17: MIP image panel of AHT outcomes without IHH bias but with Window Width bias applied

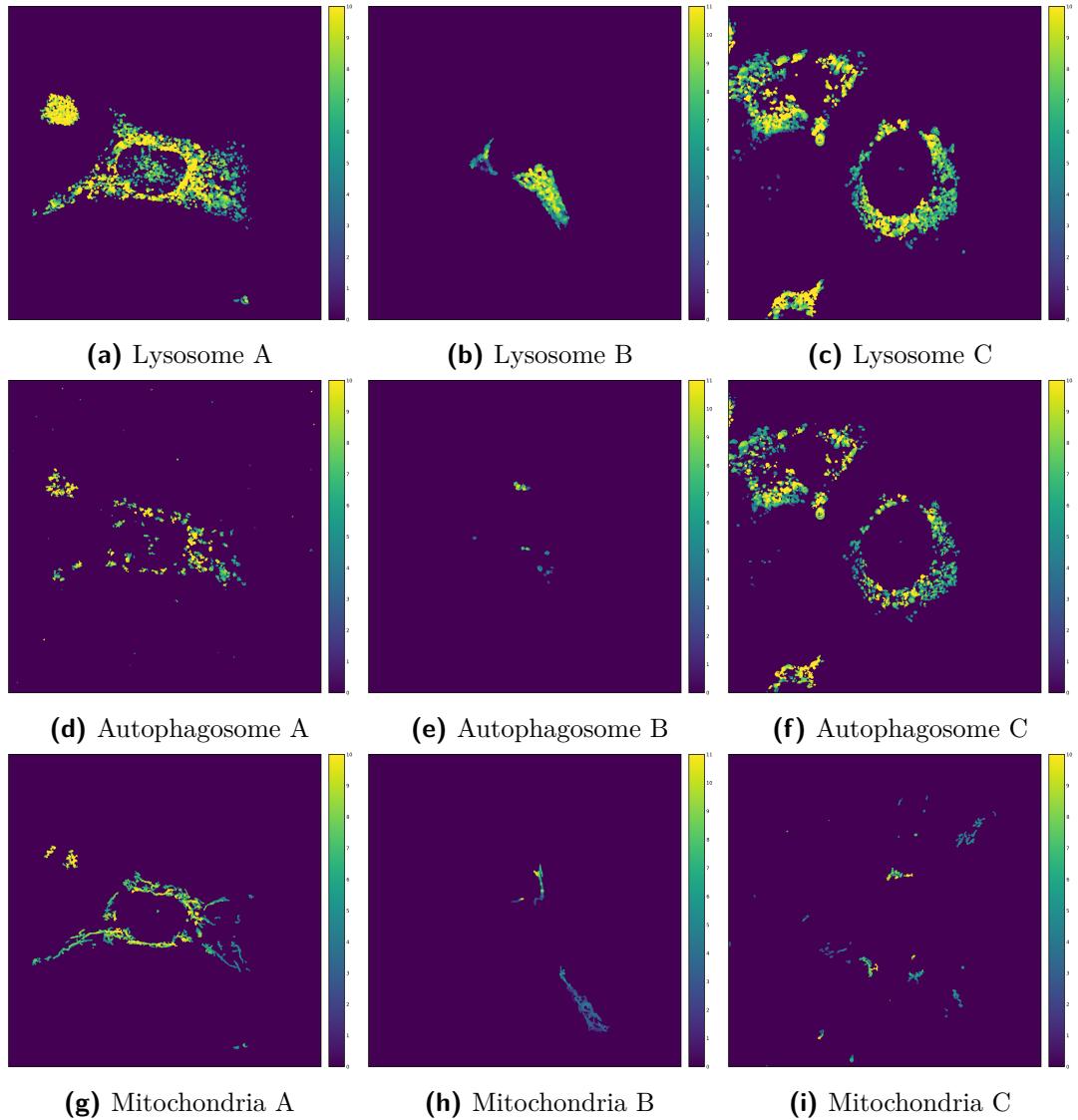


Figure B.18: MIP image panel of AHT outcomes without IHH bias but with Window Mass bias applied

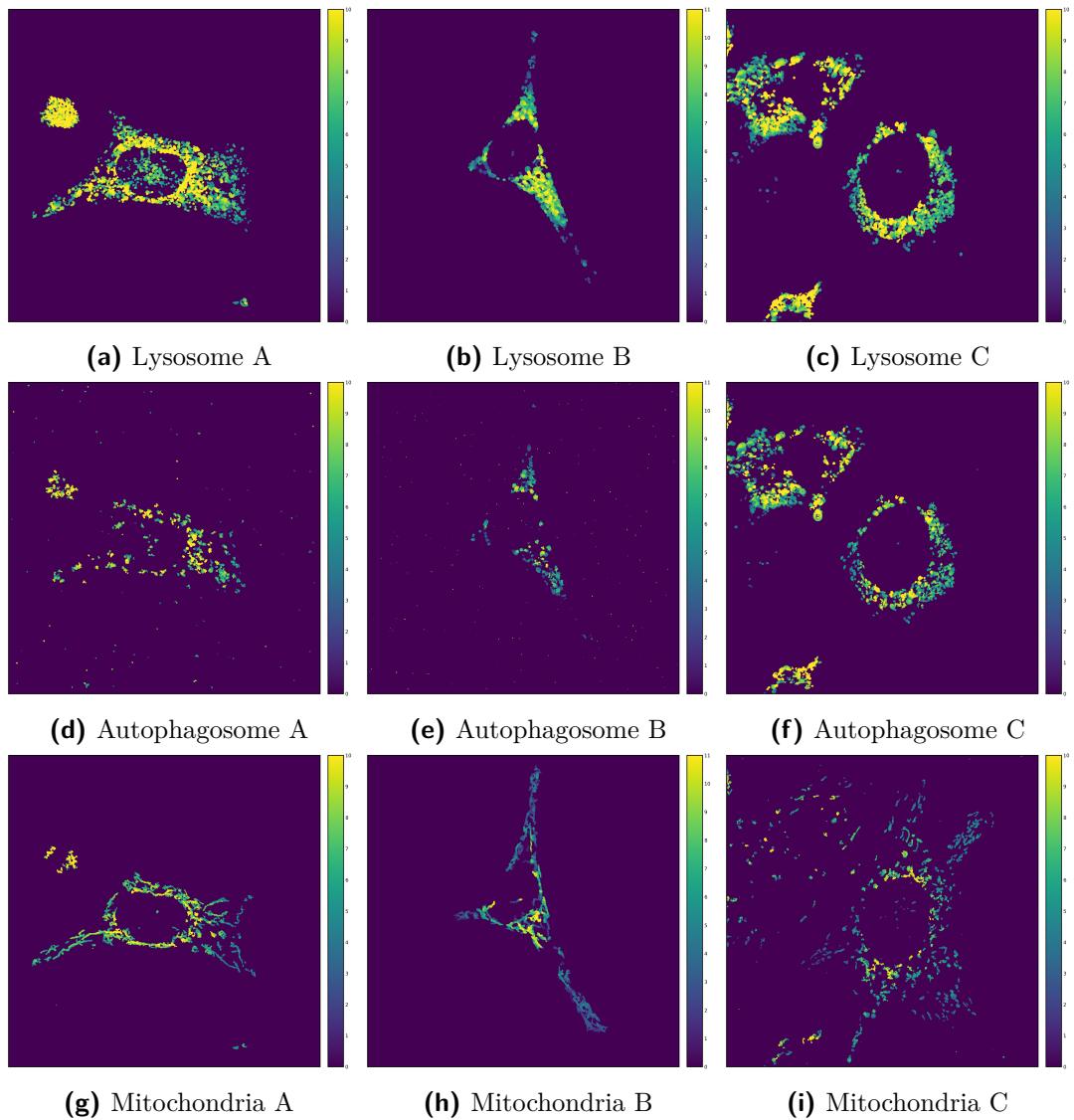


Figure B.19: MIP image panel of AHT outcomes with the IHH bias applied but no window bias applied

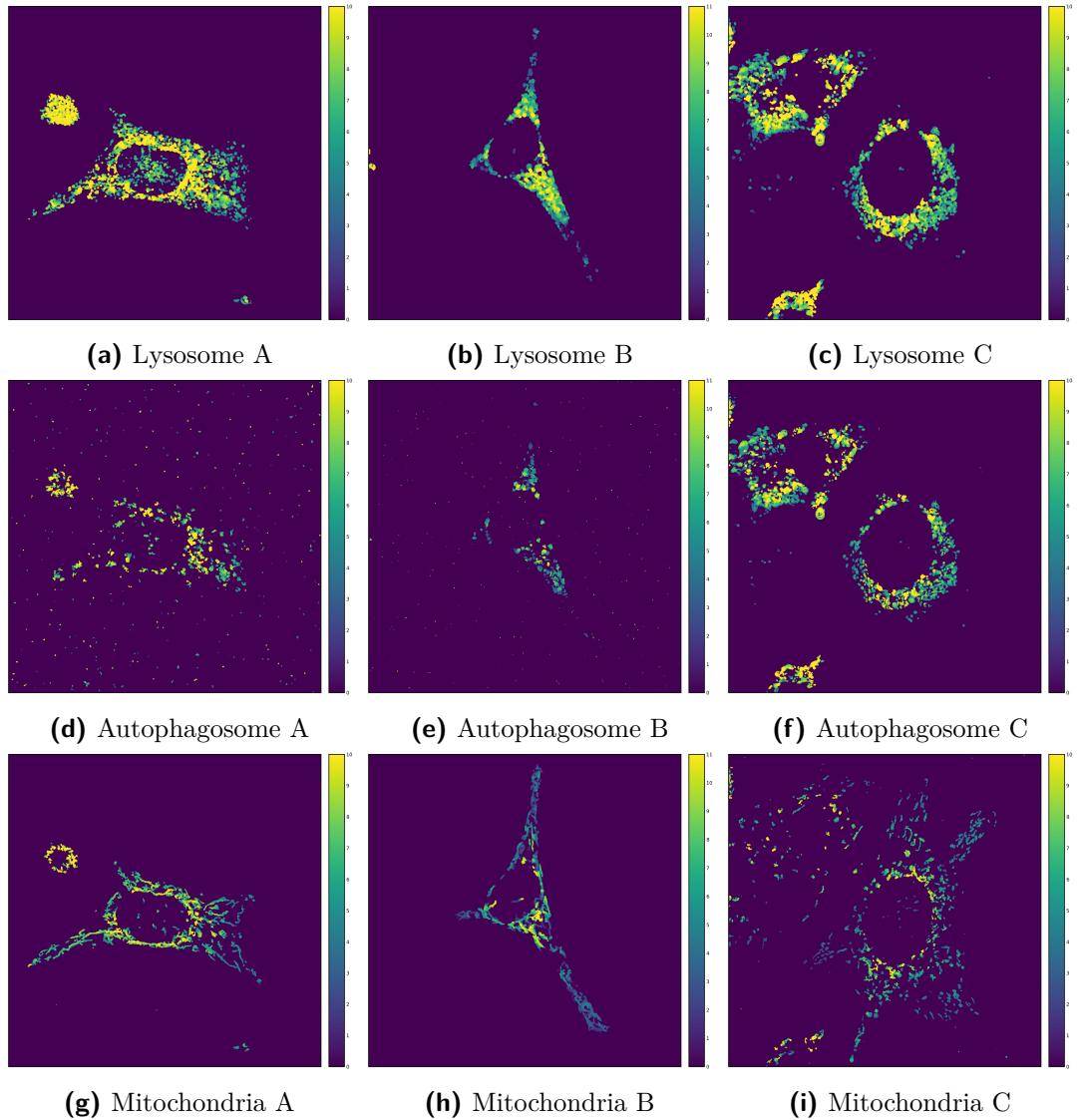


Figure B.20: MIP image panel of AHT outcomes with both the IHH and Window Width biases applied

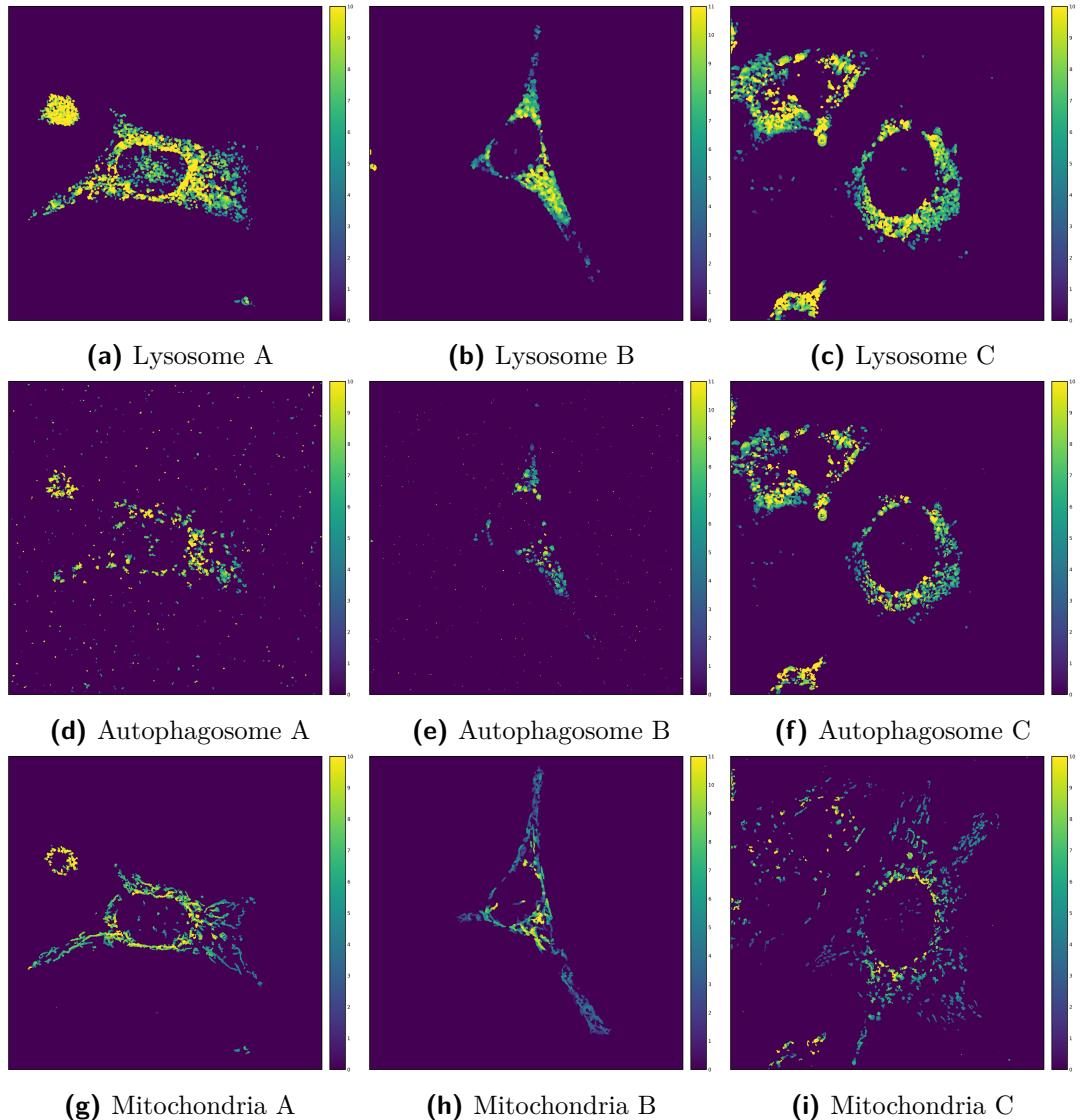


Figure B.21: MIP image panel of AHT outcomes with both the IHH and Window Mass biases applied