Adaptive optics retinal imaging in patients with retinal disease



Mital Shah

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2020

Acknowledgements

I am very grateful to my funding body, the Engineering and Physical Sciences Research Council, for providing me with the opportunity to pursue this DPhil. I would like to thank my supervisors, Susan Downes, Hannah Smithson and Laura Young, for their continued support and guidance during this DPhil. A special thank you goes to Ana Namburete from the Institute of Biomedical Engineering for her guidance, without which Chapter 5 would not have been possible. I would like to thank all members of the Eye Research Group Oxford, with special thanks to Caroline Justice, for their invaluable clinical support that made this research possible. I would like to thank all members of the Oxford Perception Lab for their advice, help and support over the last three years. Finally, I would like to thank my family, to whom this thesis is dedicated, for their ongoing support, patience and kindness.

Abstract

High-resolution *in vivo* retinal imaging with an adaptive optics scanning laser ophthalmoscope (AOSLO) allows visualisation of degenerative changes at the cellular level. The improvement in resolution with this type of imaging when compared to conventional retinal imaging technologies has made the use of cell-specific imaging biomarkers possible. With the development of new therapies for retinal disease, cell-specific biomarkers of disease are becoming increasingly important for monitoring treatment outcomes. This thesis aims to investigate the AOSLO as a clinical tool and to develop fully automated image processing and analysis pipelines for AOSLO images from patients with retinal disease.

Participants with Stargardt disease were investigated using multimodal conventional clinical imaging and high-resolution *in vivo* retinal imaging. Hyper-reflective spots, that were termed discrete hyper-reflective foci, were identified in AOSLO images in areas of retinal degeneration. We hypothesize that they may be cellular in nature and could represent cellular debris related to ongoing retinal degeneration. To optimise AOSLO image quality in patients with retinal disease, recommendations for future participant selection criteria and AOSLO image acquisition protocols are described.

Reference frames are used to measure and correct intraframe eye motion in order to produce averaged AOSLO images with a high signal to noise ratio. A fully automated algorithm for reference frame selection was developed and validated using synthetic data simulating normal retinae. Although the results did not show a significant difference between selected reference frames, averaged images generated using the developed algorithm were significantly better when compared to the default method used for reference frame selection with the Oxford AOSLO. We hypothesise that the improvement in averaged images may be due to clustering of spatially overlapping frames that occurs in the developed algorithm, which may lead to an improvement in the subsequent step of eye motion correction.

A deep learning approach for automated cone identification in AOSLO images of the photoreceptor mosaic was developed and validated. There was good agreement between the developed convolutional neural network (CNN), trained with synthetic and real data, and the current gold standard of manual marking. The developed CNN also marginally outperformed the results of current state-of-the-art automated cone detection algorithms.

Although high-resolution *in vivo* retinal imaging provides substantial advantages over conventional clinical imaging, a number of technical, image processing and image analysis related challenges need to be overcome to promote its widespread clinical application.

Contents

List of Figures **ix**

List of Tables **xii**

1 Introduction **1**

1.1 Overview . . . 1
1.2 The retina . . . 2
1.2.1 Photoreceptors . . . 2
1.2.2 Retinal pigment epithelium . . . 3
1.3 Conventional retinal imaging . . . 4
1.3.1 Fundus photography . . . 5
1.3.2 Confocal scanning laser ophthalmoscope . . . 5
1.3.3 Fundus autofluorescence . . . 7
1.3.4 Optical coherence tomography . . . 8
1.3.5 Challenges of conventional clinical imaging . . . 9
1.4 Adaptive optics retinal imaging . . . 11
1.4.1 Adaptive optics scanning laser ophthalmoscope (AOSLO) . . . 12
1.4.2 Oxford AOSLO . . . 15
1.4.3 AOSLO image processing . . . 15
1.5 Stargardt disease . . . 17
1.6 Challenges of AOSLO imaging . . . 18
1.6.1 Signal to noise ratio . . . 19
1.6.2 Ocular media opacities . . . 20
1.6.3 Deformable mirror stroke . . . 20
1.6.4 Retinal anatomy . . . 21
1.6.5 Small field of view . . . 21
1.6.6 Imaging range . . . 21
1.7 Challenges of AOSLO image processing . . . 21
1.8 Challenges of AOSLO image analysis . . . 23
1.9 Research aims and thesis outline . . . 24
1.9.1 Chapter 2 . . . 24
1.9.2 Chapter 3 . . . 24

1.9.3 Chapter 4 . . . 25
1.9.4 Chapter 5 . . . 25
1.9.5 Chapter 6 . . . 25
1.9.6 Chapter 7 . . . 25

2 The participant cohort 26
2.1 Contributions . . . 26
2.2 Background . . . 27
2.3 Methods . . . 28
2.3.1 Participants . . . 28
2.3.2 Clinical assessment . . . 29
2.3.3 Clinical data analysis . . . 31
2.3.4 Statistical analyses . . . 34
2.4 Results . . . 34
2.4.1 Demographics . . . 34
2.4.2 Retinal phenotype of participants with Stargardt disease . . . 37
2.4.3 Preferred retinal locus . . . 47
2.5 Discussion . . . 49
2.5.1 Hyper-reflective foci . . . 49
2.5.2 Leading edge of disease . . . 52
2.5.3 Clinical retinal imaging . . . 53
2.5.4 Macular microperimetry . . . 53
2.5.5 Limitations . . . 54

3 Participant selection and image acquisition with the AOSLO 55
3.1 Contributions . . . 55
3.2 Background . . . 56
3.3 Participant selection and image acquisition . . . 56
3.3.1 Fixation stability . . . 56
3.3.2 Fundus-guided retinal imaging . . . 57
3.3.3 Refractive error . . . 58
3.3.4 Ocular media opacity . . . 59
3.3.5 Assessment of retinal structure . . . 59
3.3.6 Ocular surface . . . 60
3.3.7 Maintaining active individual participation . . . 60
3.3.8 Avoiding dark adaptation . . . 61
3.3.9 Variations in local retinal topography . . . 62
3.4 Recommendations for future participant recruitment . . . 62
3.4.1 Participant selection criteria . . . 62
3.4.2 AOSLO image acquisition protocol . . . 63

4 Automated reference frame selection **65**
4.1 Contributions . . . 65
4.2 Background . . . 66
4.3 Methods . . . 67
4.3.1 Total intensity . . . 69
4.3.2 Interframe grouping . . . 71
4.3.3 Hierarchical clustering . . . 75
4.3.4 Image quality metrics . . . 76
4.3.5 Synthetic images . . . 80
4.3.6 Synthetic datasets . . . 82
4.3.7 Estimating residual aberration . . . 83
4.3.8 Estimating signal to noise ratio . . . 86
4.3.9 Estimating eye motion . . . 86
4.3.10 Reference frame selection . . . 93
4.4 Results . . . 95
4.4.1 Algorithm evaluation . . . 95
4.5 Discussion . . . 99

5 Automated cone identification using deep learning **104**
5.1 Contributions . . . 105
5.2 Background . . . 105
5.3 Methods . . . 109
5.3.1 Evaluation protocol . . . 109
5.3.2 Data . . . 113
5.4 Experiment 1 . . . 117
5.4.1 Input images . . . 117
5.4.2 FCRN and U-Net architectures . . . 117
5.4.3 Results . . . 121
5.5 Experiment 2 . . . 124
5.5.1 Results . . . 125
5.6 Experiment 3 . . . 130
5.6.1 Input images . . . 131
5.6.2 New U-Net architecture . . . 132
5.6.3 Results . . . 136
5.7 Experiment 4 . . . 140
5.7.1 Input images . . . 142
5.7.2 U-Net transfer learning . . . 143
5.7.3 Results . . . 144
5.7.4 Comparison with state-of-the-art . . . 150
5.8 Discussion . . . 152

6 AOSLO imaging in participants with Stargardt disease 156
6.1 Contributions . . . 157
6.2 Background . . . 157
6.3 Methods . . . 159
6.3.1 AOSLO imaging . . . 159
6.3.2 AOSLO image scaling and processing . . . 164
6.4 Results . . . 165
6.4.1 AOSLO images in Stargardt disease . . . 165
6.5 Automated reference frame selection in participants with Stargardt disease . . . 173
6.6 Automated cone photoreceptor identification in participants with Stargardt disease . . . 174
6.7 Discussion . . . 176
6.7.1 Discrete hyper-reflective foci . . . 177
6.7.2 ABCA4-associated optical gap phenotype . . . 178
6.7.3 Retinal pigment epithelium mosaic . . . 179
6.7.4 Automated reference frame selection in participants with Stargardt disease . . . 181
6.7.5 Automated cone photoreceptor identification in participants with Stargardt disease . . . 181
6.8 Conclusions . . . 182

7 Summary and conclusions 184
7.1 Chapter 2: The participant cohort . . . 185
7.2 Chapter 3: Participant selection and image acquisition with the AOSLO . . . 185
7.3 Chapter 4: Automated reference frame selection . . . 187
7.4 Chapter 5: Automated cone photoreceptor identification using deep learning . . . 187
7.5 Chapter 6: AOSLO imaging in participants with Stargardt disease . 189
7.6 Conclusions . . . 189

References 191

List of Figures

1.1 Spatial distribution of photoreceptors in the retina 3
1.2 Retinal surface anatomy 3
1.3 Photoreceptor anatomy 4
1.4 Anatomy of the retinal layers 5
1.5 Anatomy of the retinal pigment epithelium 6
1.6 Conventional clinical retinal imaging 7
1.7 Raster scanning 8
1.8 Cross section of the human eye 10
1.9 Adaptive optics imaging 11
1.10 Adaptive optics imaging versus post-mortem tissue histology 12
1.11 Adaptive optics retinal imaging system schematic 13
1.12 AOSLO imaging geometries 14
1.13 AOSLO image processing pipeline used with the Oxford AOSLO . . 16
1.14 Conventional clinical retinal images in Stargardt disease 18
1.15 Role of ABCA4 in Stargardt disease 19
1.16 Improvements in the AOSLO image signal to noise ratio with averaging 20
1.17 Eye movement related image distortion 22

2.1 Normal fundus autofluorescence and optical coherence tomography images 33
2.2 Retinal phenotype in participants with Stargardt disease (S01 - S04) 39
2.3 Retinal phenotype in participants with Stargardt disease (S05 - S08) 40
2.4 Retinal phenotype in participants with Stargardt disease (S09 - S12) 41
2.5 Retinal phenotype in participants with Stargardt disease (S13 - S16) 42
2.6 Comparison between near-infrared and short-wavelength fundus autofluorescence 44
2.7 Early spectral-domain optical coherence tomography features of outer retinal degeneration 45
2.8 Hyper-reflective foci on spectral-domain optical coherence tomography 46
2.9 Comparison of hyper-reflective foci with retinal flecks on spectral-domain optical coherence tomography 51

4.1 Automated reference frame selection algorithm workflow 68

4.2 Total intensity . . . 70
4.3 Histogram of total frame intensity . . . 71
4.4 Interframe registration . . . 74
4.5 Hierarchical agglomerative clustering . . . 77
4.6 Synthetic frames used for evaluating image quality metrics . . . 84
4.7 Relationship between image quality metrics and Strehl ratio . . . 85
4.8 Estimating intraframe eye motion . . . 87
4.9 Predicting normal simulated eye movements . . . 89
4.10 Synthetic frames with induced shear used for evaluating intraframe eye motion detection . . . 90
4.11 Relationship between r_{DFT} and induced shear . . . 91
4.12 Logistic regression model results investigating r_{DFT} as a predictor of induced image shear . . . 93
4.13 Reference frame selection module workflow . . . 96
4.14 Evaluation of reference frame selection . . . 97
4.15 Performance of the automated algorithms . . . 98
4.16 Evaluation of averaged images . . . 99

5.1 Artificial neuron . . . 107
5.2 Identifying cones from the convolutional neural network prediction . 110
5.3 Identifying true and false positive predicted cones . . . 111
5.4 Dealing with edge cones . . . 112
5.5 Jaccard similarity coefficient . . . 113
5.6 Confocal AOSLO images from Dataset 4 . . . 115
5.7 Confocal AOSLO image from Dataset 5 . . . 116
5.8 Training labels . . . 117
5.9 Experiment 1 input images . . . 118
5.10 Fully Convolutional Regression Network model architecture . . . 122
5.11 U-Net model architecture . . . 123
5.12 Fully Convolutional Regression Network training losses . . . 124
5.13 Fully Convolutional Regression Network accuracy . . . 125
5.14 U-Net training losses . . . 126
5.15 U-Net accuracy - loss function . . . 127
5.16 U-Net accuracy - kernel size . . . 128
5.17 U-Net accuracy stratified by input image signal to noise ratio . . . 129
5.18 Average cone size from synthetic input images (Dataset 3) . . . 130
5.19 Synthetic input images used for training and evaluating the new U-Net model . . . 133
5.20 Real input images used for transfer learning and evaluation . . . 134
5.21 New U-Net model architecture . . . 135

5.22 New U-Net model training losses . . . 137
5.23 Performance of the new U-Net model after training with Dataset 3a and Dataset 3b . . . 138
5.24 New U-Net model accuracy after training with Datasets 3a and 3b . 139
5.25 New U-Net model training losses during parameter fine tuning . . . 140
5.26 Performance of the new U-Net model after parameter fine-tuning with Dataset 3c . . . 141
5.27 New U-Net model accuracy after training with Dataset 3c . . . 142
5.28 Transfer learning new U-Net model architecture . . . 145
5.29 New U-Net model training losses during transfer learning with Dataset 5a . . . 146
5.30 New U-Net model results after transfer learning with Dataset 5a . . 146
5.31 New U-Net model accuracy after transfer learning with Dataset 5a . 147
5.32 New U-Net model training losses during transfer learning with Dataset 5b . . . 148
5.33 New U-Net model results after transfer learning with Dataset 5b . . 148
5.34 New U-Net model accuracy after transfer learning with Dataset 5b . 149
5.35 Performance of the new U-Net model with current state-of-the-art cone detection methods . . . 153

6.1 AOSLO imaging locations in participants S02, S04 and S06 . . . 160
6.2 AOSLO imaging locations in participants S07, S08 and S10 . . . 161
6.3 AOSLO imaging locations in participants S12, S13 and S15 . . . 162
6.4 AOSLO imaging locations in participants S01, S03, S09, S11 and S16 163
6.5 AOSLO imaging locations in participants S05 and S14 . . . 164
6.6 Normal photoreceptor mosaic . . . 167
6.7 Degenerated photoreceptor mosaic . . . 168
6.8 Size and distribution of discrete hyper-reflective foci . . . 169
6.9 Discrete hyper-reflective foci . . . 170
6.10 Retinal pigment epithelium mosaic . . . 172
6.11 Automated reference frame selection in participants with Stargardt disease . . . 175
6.12 Automated cone photoreceptor identification in participants with Stargardt disease . . . 176

List of Tables

2.1 Ophthalmic clinical assessment 29
2.2 Demographics and baseline clinical characteristics of participants with Stargardt disease 35
2.3 Demographics and baseline clinical characteristics of control participants 36
2.4 Genetic testing results of participants with Stargardt disease 38
2.5 Retinal phenotype in participants with Stargardt disease 43
2.6 Retinal structure and function at the preferred retinal locus 48

4.1 Multiple linear regression to predict Strehl ratio 86
4.2 Logistic regression model results investigating r_{DFT} as a predictor of induced image shear 94

5.1 U-Net results for each kernel size tested 130
5.2 New U-Net architecture results after training with Dataset 3a and Dataset 3b 137
5.3 New U-Net architecture results after training with Dataset 3c . . . 140
5.4 Comparing the new U-Net with the current state-of-the-art 152

6.1 Discrete hyper-reflective foci in participants with Stargardt disease. 171
6.2 Adaptive optics images of the retinal pigment epithelium 180

1 Introduction

1.1 Overview

In 2013 approximately 2 million people in the United Kingdom were living with sight loss and this number is predicted to double by 2050 [1]. Retinal disease accounts for two of the five leading causes of sight loss in the United Kingdom and over 20% of sight loss in adults is attributable to retinal disease [1]. Inherited retinal diseases are now the commonest cause of blindness in working age adults [2], and the second commonest cause of severe visual impairment or blindness in children [3].

Retinal imaging has a higher sensitivity and specificity than clinical examination alone [4] and plays an essential role in patient diagnosis and management. High-resolution *in vivo* retinal imaging with adaptive optics (AO) allows degenerative changes in the retina to be tracked at significantly finer spatial and temporal scales than is possible with conventional clinical imaging. With the development of new therapies, such as gene replacement or cell-based therapy, cell-specific biomarkers of disease are becoming increasingly important for monitoring treatment outcomes. The highly sensitive measures of retinal structure and function provided by AO retinal imaging systems have enabled visualisation of disease progression earlier than conventional clinical imaging [5] and are increasingly being used during interventional clinical trials [6]. Additionally, these measures can be used to provide

a greater understanding of underlying disease mechanisms, as biomarkers of retinal disease and may enable earlier disease detection.

The overall aims of this thesis are to investigate the adaptive optics scanning laser ophthalmoscope (AOSLO) as a clinical tool and to develop fully automated image processing and analysis pipelines for AOSLO images from patients with retinal disease. These aims will be achieved by investigating the relationship between multimodal conventional clinical and AOSLO imaging in participants with Stargardt disease (an inherited retinal disease affecting the central vision) to determine if AO retinal imaging will provide additional clinically relevant information, and developing and validating fully automated algorithms with the ultimate aim of using them on images from patients with retinal disease.

1.2 The retina

The retina lies at the back of the eye and converts light, which is focussed by the cornea and lens on to the retinal surface, into electro-chemical impulses that are transmitted to the brain through complex neural pathways. The macula (central retina) is dominated by cone photoreceptors whereas the peripheral retina predominantly contains rod photoreceptors (Figure 1.1). The macula is located temporal to the optic disc and between the two major (superior and inferior) temporal retinal vascular arcades (Figure 1.2). The fovea, which lies at the centre of the macula contains the highest density of cones with a rod free zone at its centre (Figures 1.1 and 1.2).

1.2.1 Photoreceptors

Retinal photoreceptors (cones and rods) consist of four distinct segments: the outer segment, inner segment, cell body and synaptic terminal (Figure 1.3). The photoreceptor outer segments lie adjacent to the retinal pigment epithelium (RPE) and consist of disc-like structures that contain visual pigment (Figure 1.3). Connecting cilia connect the outer segments to the inner segments, which contain the metabolic and biosynthetic machinery of the cell and are rich in mitochondria, endoplasmic

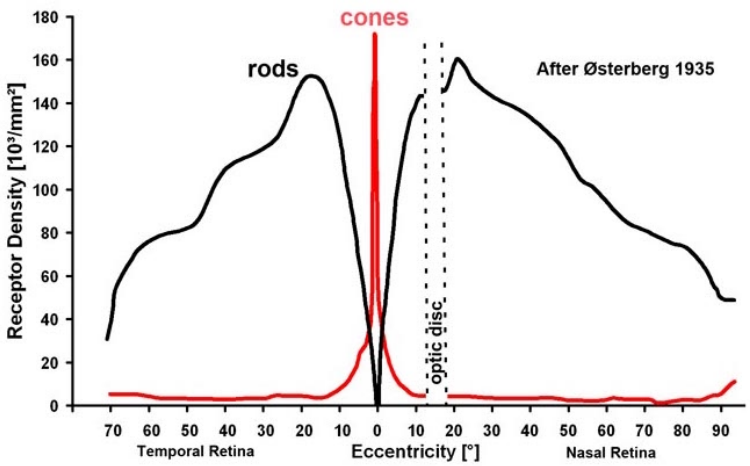


Figure 1.1: Spatial distribution and density of rod and cone photoreceptors along the horizontal meridian (adapted from [7]).

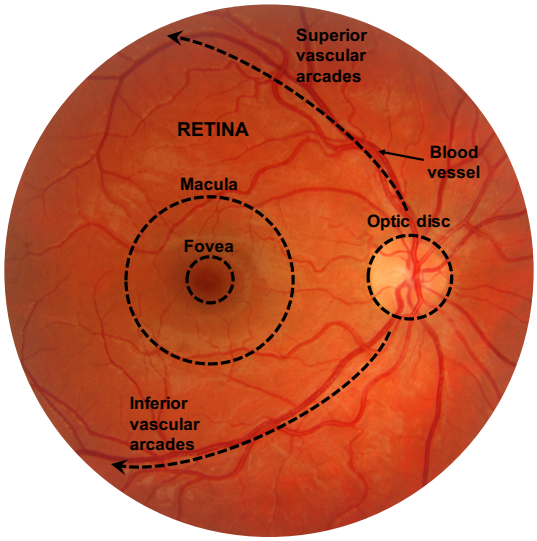


Figure 1.2: Right eye colour fundus photograph demonstrating anatomical features on the surface of a normal retina.

reticulum and other subcellular organelles [8]. The cell bodies of cones form a single row within the outer nuclear layer adjacent to the external (outer) limiting membrane, while those of rods make up the remainder of the outer nuclear layer (Figure 1.4).

1.2.2 Retinal pigment epithelium

Interactions between photoreceptors and RPE cells are essential for visual function [11]. The RPE is located between the photoreceptor outer segments and the choroid. It is a monolayer of hexagonally packed cells, connected by tight junctions,

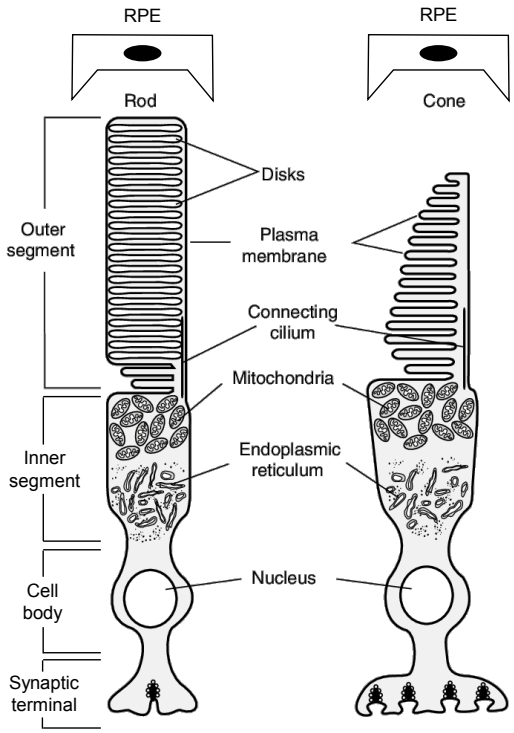


Figure 1.3: Rod and cone photoreceptor anatomy (adapted from [9]).

that contain pigment granules and organelles for digestion of the outer segment discs that are shed from photoreceptors. The apical processes of the RPE cells are directed towards the subretinal space to facilitate interactions with the photoreceptor outer segments (Figure 1.5). The RPE also plays a crucial role in the visual cycle, the process of converting a photon of light into an electrical signal in the retina, by converting all-*trans*-retinol to 11-*cis*-retinal, which is delivered back to photoreceptors [11].

1.3 Conventional retinal imaging

Conventional retinal imaging has become crucial in evaluating the retina clinically: for diagnostic purposes; for making treatment decisions and assessing treatment response; and in monitoring patients for signs of disease progression. Conventional retinal imaging technologies provide an *en face* or cross sectional view of the retina and include colour fundus photography (Figure 1.6A), fundus autofluorescence imaging (Figure 1.6B and C), and optical coherence tomography (Figure 1.6D).

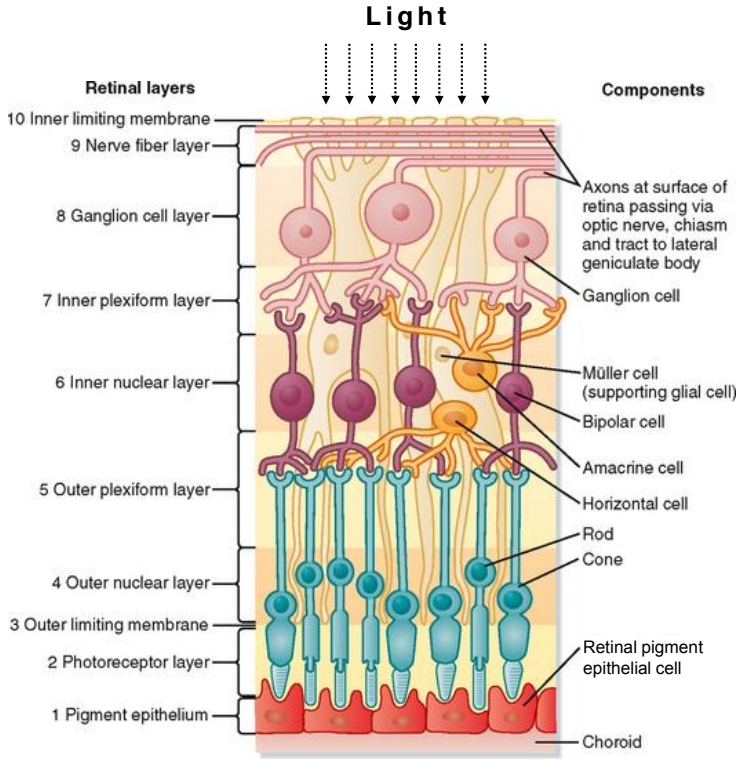


Figure 1.4: Anatomy of the retinal layers (adapted from [10]). The retina consists of 10 layers that span from the pigment epithelium, adjacent to the choroid, to the inner limiting membrane, which is closest to the front of the eye. The outer plexiform layer corresponds to the inner fibres of the photoreceptors and the synapses between the photoreceptors and cells from the inner nuclear layer. The outer nuclear layer corresponds to the photoreceptor cell bodies. Henle's nerve fibre layer, composed of the photoreceptor axons, contributes to the inner part of this layer (closest to the outer plexiform layer). The external (outer) limiting membrane corresponds to the junctional complexes between the Müller cells and photoreceptors. The photoreceptor layer contains the inner segments and photosensitive outer segments of rods and cones.

1.3.1 Fundus photography

Fundus photography (Figure 1.6A), the earliest retinal imaging technology to be developed, enables colour photographic documentation of changes in the retina and optic nerve head that are visible on clinical examination.

1.3.2 Confocal scanning laser ophthalmoscope

The confocal scanning laser ophthalmoscope is capable of imaging at very high magnification and uses monochromatic laser light to produce high contrast *en face* retinal images with an optical resolution that is superior to flood illuminated

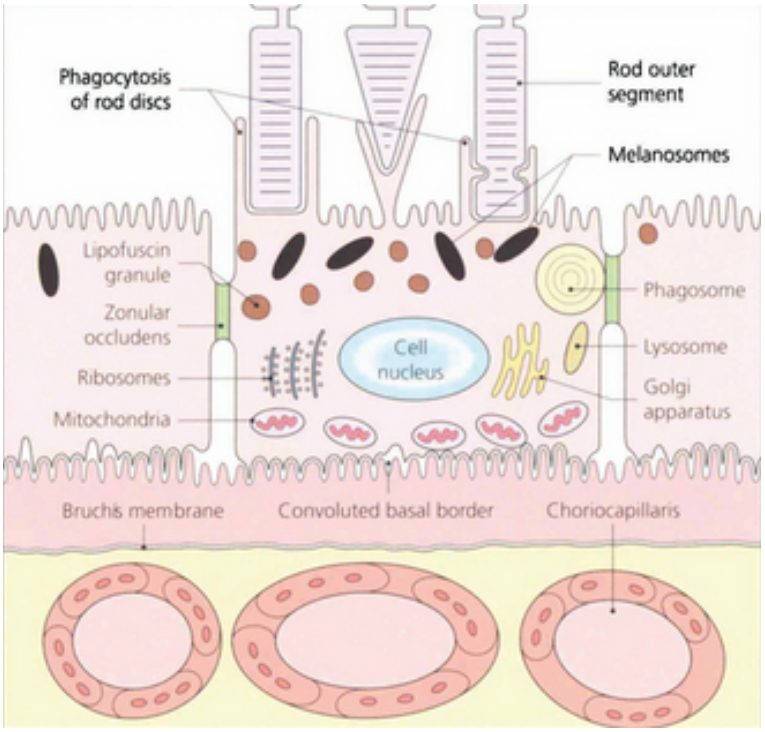


Figure 1.5: Anatomy of the retinal pigment epithelium (RPE, adapted from [12]). The RPE is located between the photoreceptor outer segments and the choroid. It consists of a single sheet of cells that are hexagonally packed, connected by tight-junctions, and contain pigment granules and organelles for the digestion of photoreceptor outer segments. Its apical membrane and processes face the subretinal space and enable interactions with photoreceptor outer segments. Its basolateral membrane is in contact with Bruch's membrane, which facilitates interaction between the RPE and the fenestrated blood vessels of the choroid (the choriocapillaris) [11].

imaging systems, such as a fundus camera. It is utilised by a number of conventional clinical retinal imaging modalities, such as fundus autofluorescence (Section 1.3.3), retinal and choroidal angiography (fluorescein and indocyanine green angiography), and ultra-widefield retinal imaging. The term "confocal" refers to the arrangement of both illumination and detection arms having a common focal point. This is achieved by placing a pinhole in front of the detector conjugate to the plane of focus on the retina, which only allows light returning from that plane to reach the detector and blocks out-of-focus light. Image formation occurs by scanning this focal point across the field of view. A set of scanning mirrors are used to rapidly scan the laser across the retina in a raster pattern (Figure 1.7), one mirror scans horizontally while the other scans vertically. Light from outside the focal point

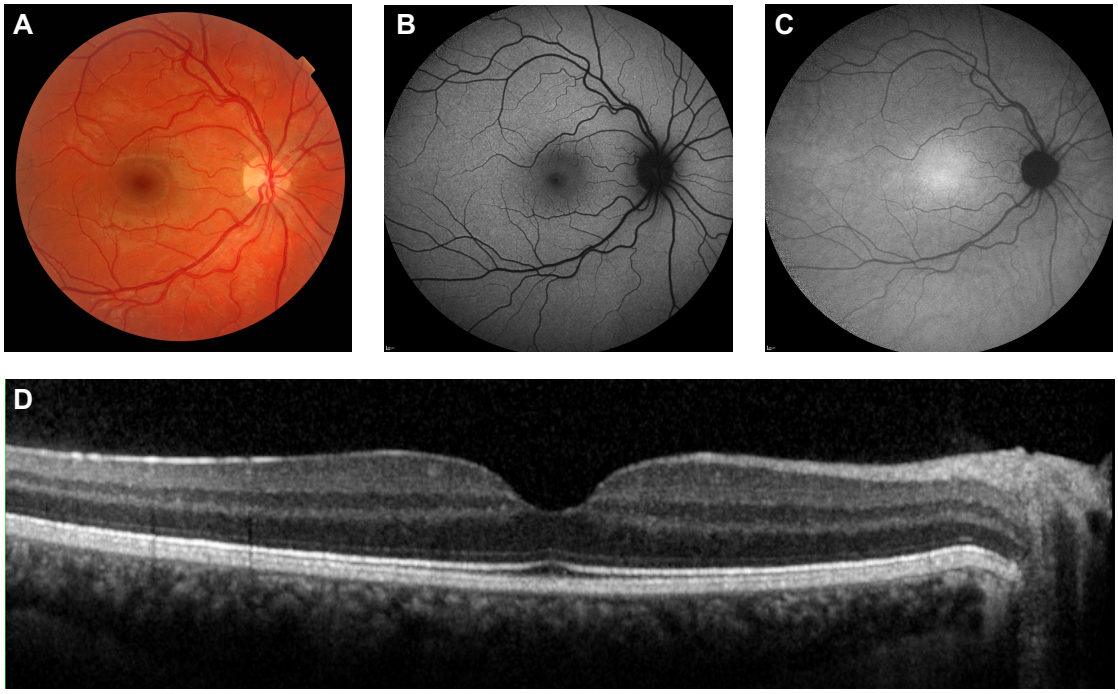


Figure 1.6: Examples of conventional clinical retinal imaging of a normal retina from a right eye. **A** - colour fundus photograph (50° field of view). **B** – short-wavelength fundus autofluorescence using a scanning laser ophthalmoscope (55° field of view), **C** - near-infrared fundus autofluorescence using a scanning laser ophthalmoscope (55° field of view), **D** - spectral-domain optical coherence tomography providing a cross section of the retina (cross section shown through the fovea).

is blocked by the pinhole, which reduces the overall signal intensity but greatly improves contrast. Despite the higher resolution afforded by the confocal scanning laser ophthalmoscope, resolution is still limited by ocular wavefront aberrations.

1.3.3 Fundus autofluorescence

Autofluorescence is the emission of light from fluorescent molecules (fluorophores) that occurs naturally in structures/cells, when they absorb light from an excitation light source. Fundus autofluorescence imaging utilises the intrinsic fluorescence of retinal pigments to provide information about the health and function of the retina. There are a number of retinal fluorophores, each producing a peak fluorescence signal at a different excitation wavelength. The signal from autofluorescence imaging with short wavelength light (488 nm excitation, 500 nm cut-off filter, Figure 1.6B) is purported to be derived primarily from lipofuscin [14–16], a by-product of

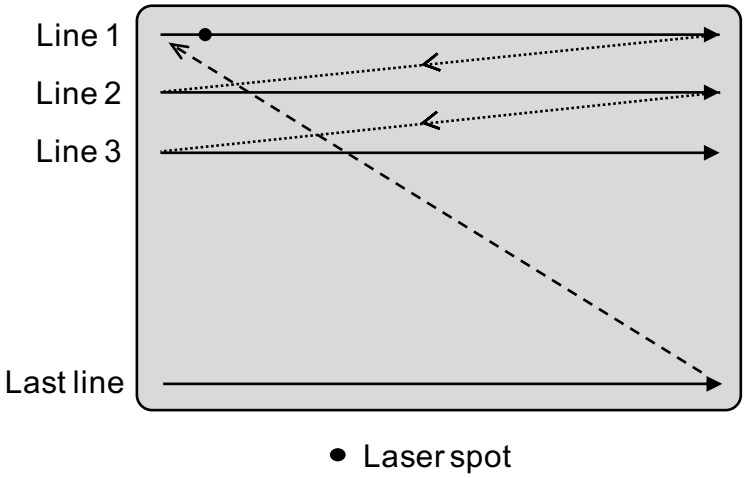


Figure 1.7: Schematic illustrating the concept of raster scanning (adapted from [13]). Note that to image a 2D patch scanning occurs both horizontally and vertically.

the visual cycle, while with near-infrared light (787 nm excitation, 830 nm cut-off filter, Figure 1.6C) it appears to originate predominantly from melanin within in the RPE and choroid [16–18].

1.3.4 Optical coherence tomography

Optical coherence tomography (OCT) is the most commonly used retinal imaging modality in ophthalmology and is the third commonest procedure or intervention carried out by the National Health Service in England [19]. OCT is analogous to ultrasound and measures the back-reflected light from tissues, such as the retina, to produce two-dimensional cross-sectional images that can be stacked to produce a 3D volume. Representations of tissue structure within OCT images are based on interference. The back-reflected light is combined with a reference beam: if the distance travelled by light in both paths is the same, this leads to positive interference; if the distance differs by half of the wavelength, this leads to destructive interference. As the interference pattern is sensitive to small changes in the distance travelled by light, it provides superior axial (depth) resolution. Since the first description of OCT in 1991 [20], a number of technological improvements have led to the development of systems with higher resolution and faster imaging speeds. The axial and transverse resolution of modern OCT systems is 4-7 µm and 15-20 µm, respectively, and the

development of newer spectral-domain OCT (SD-OCT) systems, with fewer moving parts, has led to a 50-100 times improvement in imaging speed [21]. Despite these technological improvements image quality with OCT is still limited by ocular wavefront aberrations, as with the confocal scanning laser ophthalmoscope.

1.3.5 Challenges of conventional clinical imaging

Although conventional clinical imaging technologies are widely used, there are challenges associated with their use. Image quality is sensitive to opacities of the ocular media and the resolution of retinal images is limited by optical imperfections in the eye.

1.3.5.1 Ocular media opacities

Altered transparency of the ocular media (the transparent structures of the eye) reduces the signal intensity from the retina and can affect image quality. Causes of altered ocular media transparency include:

- Corneal opacities (e.g. scarring) - the cornea is the clear external surface of the eye (Figure 1.8) and (along with the tear film) accounts for approximately 70% of the total optical power of the eye [22].
- Cataract - a cataract is clouding of the eye lens (Figure 1.8) and can cause significant scattering of back-reflected light impairing image quality.
- Vitreous opacities (e.g. age-related degeneration of the vitreous humor) - the vitreous humor is a clear gel-like substance that fills the space behind the eye lens and in front of the retina (Figure 1.8).

1.3.5.2 Ocular wavefront aberrations

The resolution of an optical system is proportional to the wavelength of light and inversely proportional to the diameter of the aperture. There is a theoretical limit to the resolution of an optical imaging system due to the physics of diffraction. An optical system capable of achieving this theoretical resolution limit is said to

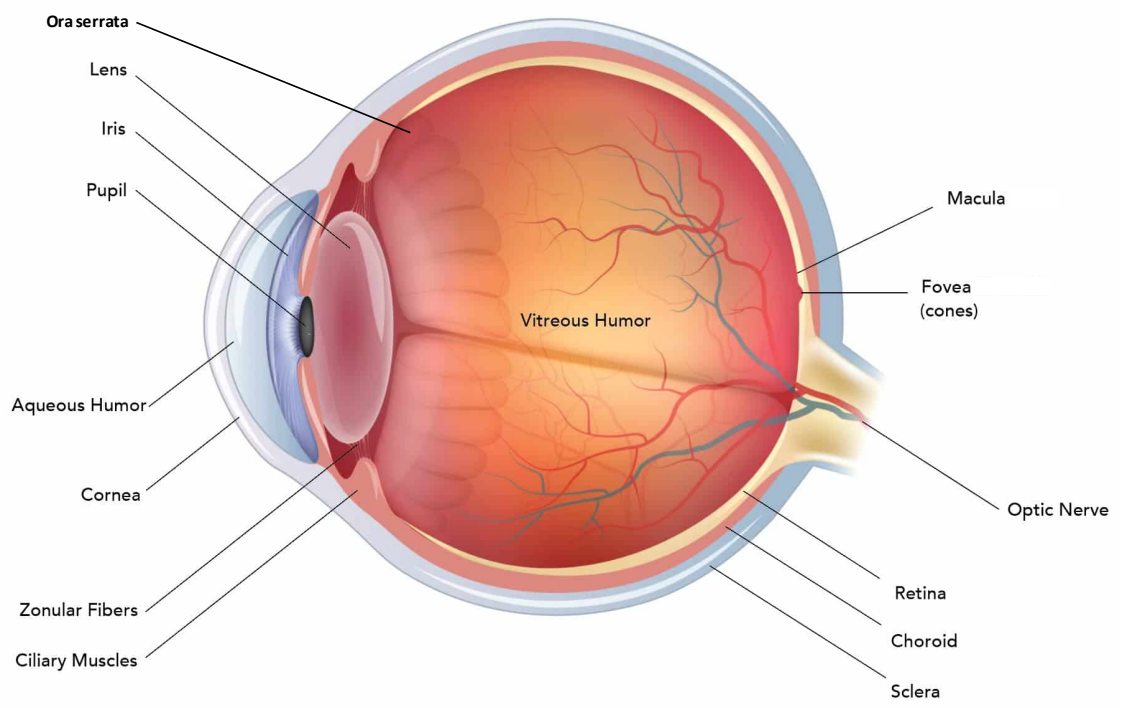
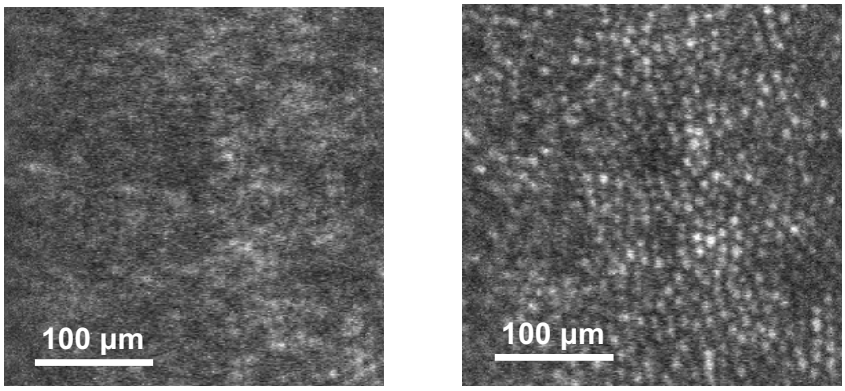


Figure 1.8: Cross section of the human eye. Schematic illustrating the ocular anatomy (adapted from [23]). Light enters the eye through the cornea, passes through the aqueous humor, pupil, lens and vitreous humor before reaching the retina.

be diffraction limited. The aperture of the eye is the pupil and the iris is the diaphragm that controls its size (Figure 1.8). A larger pupil leads to a higher resolution by reducing the diffraction limit. With near-infrared light the theoretical resolution limit of the human eye is approximately 2 µm on the retina for a maximally dilated pupil [24], however, this is not reached as the eye is not diffraction limited. This occurs due to the presence of optical imperfections in the eye, termed ocular wavefront aberrations, which result in blurring and limit both the optical quality of retinal images and vision.

Common optical aberrations affecting the eye include positive (hyperopia or long-sighted) and negative (myopia or short-sighted) defocus, as well as astigmatism. These low order aberrations account for approximately 90% of the ocular wavefront aberrations while high order aberrations account for the remaining 10% [22]. Monochromatic ocular aberrations arise from various sources, which include the tear film, cornea and lens. While the anterior surface of the cornea (and tear film) contribute approximately two-thirds of the eye's optical power and aberrations, the



(a) No adaptive optics correction **(b)** Adaptive optics correction

Figure 1.9: Adaptive optics imaging. **(a)** - A single raw frame with no adaptive optics correction applied. Only blurred, coarse structures can be seen due to the eye's optical aberrations. **(b)** - By applying adaptive optics to correct the optical aberrations in **(a)** it is possible to obtain a high-resolution image of the same region where individual cone photoreceptors (the bright spots) are clearly visible.

lens is largely responsible for the rest. The pupil controls how much light enters the eye, and a large pupil improves the diffraction-limited resolution. However, a large pupil also increases aberrations, which are generally larger at the edge of the pupil.

1.4 Adaptive optics retinal imaging

Limitations in the resolution of conventional clinical retinal imaging technologies caused by ocular wavefront aberrations has precluded the use of cell-specific retinal biomarkers with this type of imaging. By correcting these naturally occurring optical aberrations of the eye, adaptive optics (AO) retinal imaging systems can achieve optical resolutions of around 2 µm in the living human eye [24] (Figure 1.9). This cellular resolution has enabled measurements of retinal structures that were previously only possible in humans with post-mortem tissue histology (Figure 1.10).

AO retinal imaging systems typically consist of a wavefront sensor and a wavefront corrector coupled together by a control system (Figure 1.11). The wavefront sensor measures the ocular wavefront aberrations while the wavefront corrector, commonly a deformable mirror, changes the wavefront under computer control. The surface of a deformable mirror can be deformed such that it corrects the distortions to the wavefront of light passing into and out of the eye. Conventionally

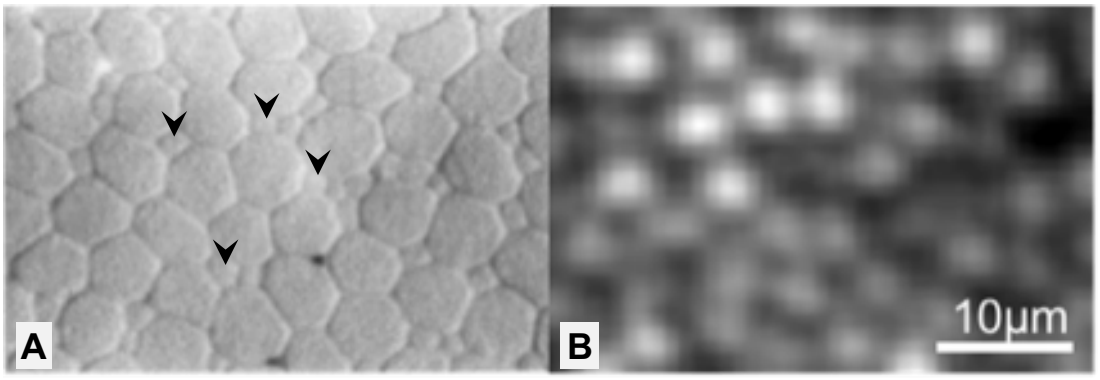


Figure 1.10: Adaptive optics imaging versus post-mortem tissue histology. **A** - optical section of the photoreceptor mosaic showing individual cones with few rods (examples highlighted with black arrow heads) at 1.4° nasal to the foveal centre from an *ex-vivo* human histological specimen of a whole-mounted retina (adapted from [25]). **B** – *in vivo* AOSLO image of the photoreceptor mosaic from the Oxford AOSLO [26] demonstrating individual cones (the bright spots) at 1.4° temporal to the foveal centre in a normal control participant.

these components operate in a closed loop system [27] in order to minimise wavefront aberrations, ideally providing diffraction limited performance. However, this is not always possible [28] and residual aberrations can affect image quality. To date, a number of optical imaging systems have been assisted by AO, such as: flood illumination systems (fundus camera) [29–31], raster or line scanning laser ophthalmoscopes [26, 32–35] and optical coherence tomography [36–38].

1.4.1 Adaptive optics scanning laser ophthalmoscope

The addition of AO to a scanning laser ophthalmoscope (AOSLO) provides high-resolution retinal images with increased contrast and the standard confocal imaging geometry of the AOSLO allows for optical sectioning. As with a scanning laser ophthalmoscope, in the AOSLO a single spot is scanned across the retina in a raster pattern and the back-scattered light from each location that reaches the detector is recorded to construct an image. When used for confocal imaging, only light returning from the plane of focus on the retina reaches the detector and light from other layers is blocked by the pinhole. When focused on the photoreceptor layer, this provides high contrast *en face* images of the highly reflective photoreceptor outer segments [41, 42].

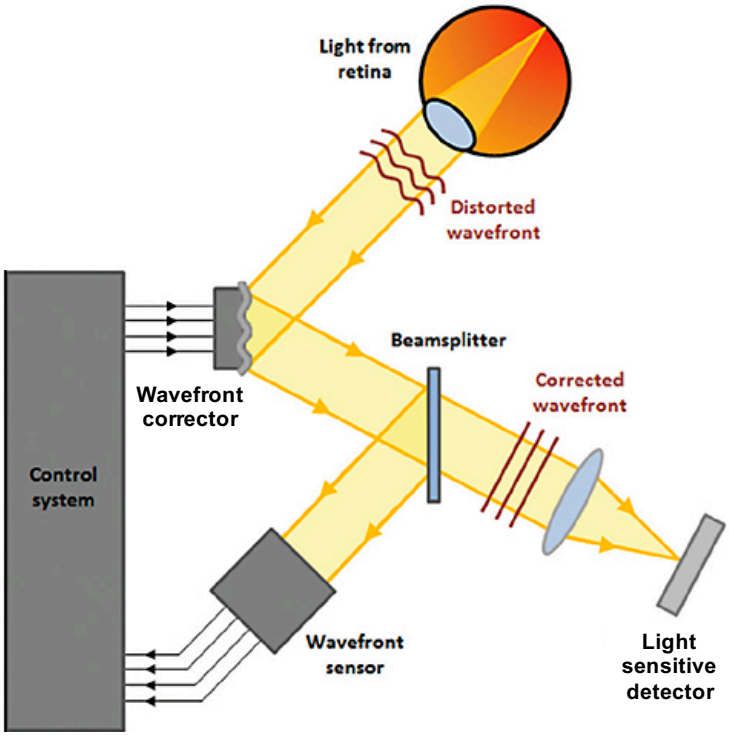


Figure 1.11: Schematic of an adaptive optics retinal imaging system (adapted from [39] and [40], Figure 10.1). A beam of light is used to illuminate the retina. Back-scattered light returning from the retina enters the optical system and is split by a beam splitter between the wavefront sensor, which measures the wavefront aberrations, and the light-sensitive detector. Based on the wavefront measurement the control system computes the actuator voltages for the wavefront corrector, causing its surface to change shape. This computation is performed in a closed-loop feedback system that attempts to minimise the wavefront aberrations.

In confocal AOSLO imaging, photoreceptors appear as bright spots because of their waveguiding properties; incoming light is propagated by total internal reflection along the photoreceptor inner segment and back-reflected light is coupled back through the photoreceptor towards the pupil centre [43]. It has been suggested that the back-reflected light originates from two primary sources, the photoreceptor inner and outer segment junction [44], and the interface between the photoreceptor outer segment and RPE [45]. Therefore the outer segment must be intact and in contact with the RPE in order to be observed as a bright region of the image [45].

A number of non-confocal imaging geometries, where multiply-scattered light is collected, have recently emerged (Figure 1.12). Split-detection imaging analyses the direction of the non-confocal signal (multiply-scattered light) by calculating the

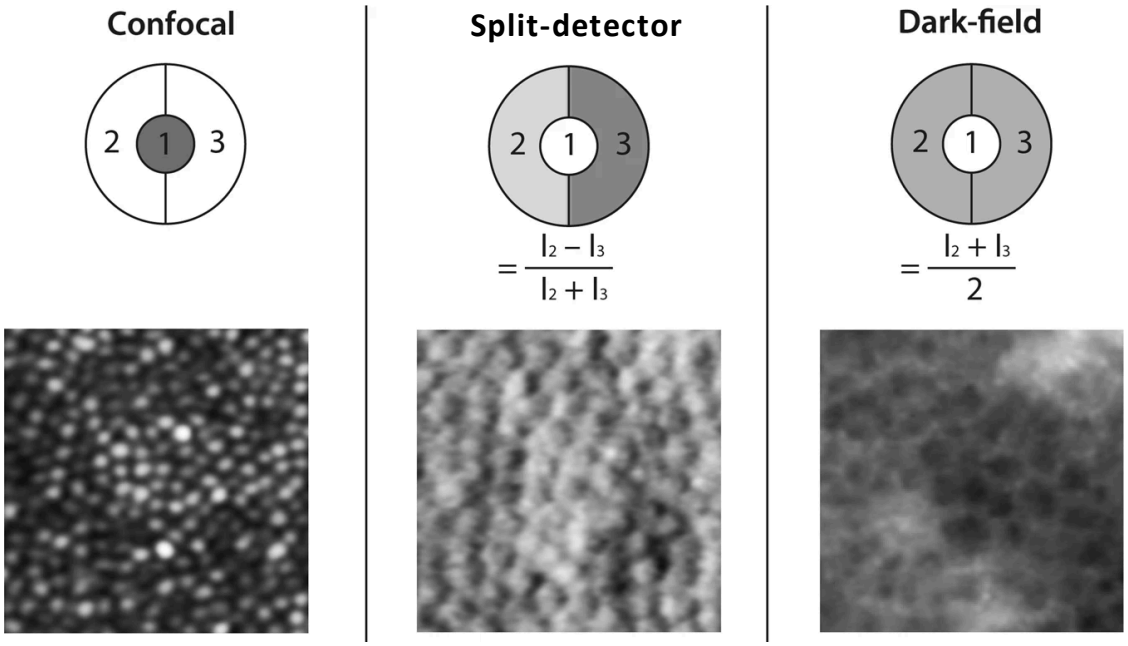


Figure 1.12: AOSLO imaging geometries (adapted from [28]). **Top row**, configuration of the annular aperture and the relative proportions of multiply-scattered light that contribute to image formation. **Bottom row**, an example of the images derived from each imaging geometry from different retinal locations in the same healthy eye. **Confocal**, only light from the plane of focus on the retina (1) is used to produce images of photoreceptor outer segments. **Split-detector**, multiply-scattered light (the difference between 2 and 3) is used to produce images of photoreceptor inner segments. **Dark-field**, uses multiply-scattered light (the sum of 2 and 3) to produce images of the hexagonal retinal pigment epithelium mosaic.

contrast between the two halves of the split and enables the separate imaging of cone inner segments [46], allowing them to be visualised even when the outer segments are missing or damaged [45, 46]. The related technique of dark-field imaging collects all of the non-confocal signal and enables direct visualisation of individual RPE cells [47] without the need for autofluorescence imaging. Contrast from split-detector and dark-field imaging appears to be generated by spatial variations in the cellular refractive index [48]. While confocal images have high resolution because the pinhole restricts the detected light to a small spatial region, non-confocal imaging geometries are unable to achieve the lateral resolution of confocal imaging [24].

1.4.2 Oxford AOSLO

The AOSLO imaging in this thesis was carried out using the Oxford AOSLO [26],

equipped with near-infrared reflectance confocal imaging capabilities (wavelength = 850 nm, full width at half maximum = 50 nm). The use of reconfigurable modules in the Oxford AOSLO, based on an Offner relay system, allows for a compact, in-plane design. A periscope is used with a beamsplitter to enable the participant to be shown a distant visual display unit containing a fixation target during experiments. By using the periscope to raise the beamsplitter above the main optical system, a wide field of view for distant fixation targets can be achieved. A fixation target is provided on a visual display unit located 1.75 m from the eye. A participant's head is stabilised using a wax dental impression on an adjustable bite-bar. Alignment of the eye to the system is performed using the wavefront-sensor image as a guide. The stroke of the deformable mirror (MultiDM, Boston Micromachines Corporation) is 3.5 µm, which enables the correction of approximately 1 dioptre of defocus or approximately 0.5 dioptres of astigmatism. Due to the limited stroke of the deformable mirror the use of refractive correction (prescription glasses or trial lenses) is required in many participants and the selection of lenses is based on retinal image quality. Images are acquired at a rate of 30 Hz with a $1.6° \times 1.0°$ field of view, and retinal locations are targeted by directing the participant's gaze using the fixation target. Pharmacologic pupil dilation and cycloplegia are not performed routinely, which allows participants to focus naturally on the visual display unit but limits the maximum resolution achievable. The selection of a long imaging wavelength reduces illumination-related pupil constriction and, to maximise natural pupil dilation, all imaging is performed in a darkened room. The system has been shown in control participants to achieve diffraction-limited performance, allowing resolution of foveal cones [26].

1.4.3 AOSLO image processing

Following data acquisition a number of image processing steps are required, which include: additional processing to reconstruct an image from the sequentially collected pixels; the detection and correction of intraframe eye motion; and alignment and subsequent averaging of individual images to improve their signal to noise ratio. Figure 1.13 provides an overview of the current image processing pipeline used with

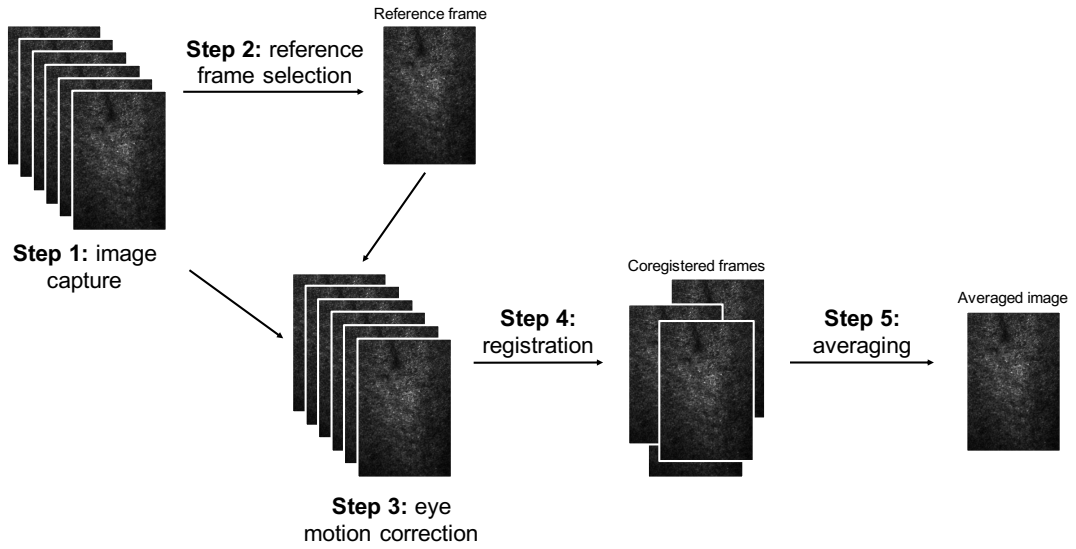


Figure 1.13: The default image processing pipeline used with the Oxford AOSLO. Following acquisition of an image sequence (Step 1), a minimally distorted reference frame is selected (Step 2) to measure and subsequently correct the eye motion present in each frame within the image sequence (Step 3). Motion-corrected frames are spatially aligned (Step 4) to create an image stack before averaging down the stack (Step 5) to produce an averaged image with an improved signal to noise ratio.

the Oxford AOSLO [26]. Image acquisition, the first step in the image processing pipeline, occurs at a rate of 30 Hz with a $1.6° \times 1.0°$ field of view using a previously described confocal AOSLO setup [26].

Unlike a fundus camera, image capture in a scanning system is not instantaneous. Therefore, eye movements that occur during image capture can lead to distortion in the image. The scanning nature of image capture with an AOSLO coupled with a participant's fixational eye movements will lead to eye movement-related image distortions. To measure and correct this intraframe eye motion requires a reference frame. As all frames in an image sequence will contain some eye movement-related image distortion, identifying the best reference frame is challenging. Therefore, a minimally distorted reference frame is selected from the acquired image sequence by using normalised cross correlation to identify the frame that is most similar to all other frames within the sequence. This is based on the assumption that fixational eye movements are random and average motion will be zero. The frame that is most similar to all others will therefore contain the least motion. Once a reference

frame has been selected the eye motion in each image is estimated using an optical flow algorithm [49] and the images are subsequently corrected for the measured distortion. Optimal performance of a motion extraction algorithm requires the retinal location of each frame to partially overlap with the reference frame. The greater the overlap between the frame being corrected and the reference frame, the better the algorithm's performance at measuring eye motion-related distortion. Averaging a co-registered (spatially aligned) and motion-corrected sequence of images down the stack is required to improve the signal to noise ratio as the signal in a single frame is restricted by light safety limits.

1.5 Stargardt disease

Stargardt disease has a reported annual incidence in the United Kingdom of between 0.110 and 0.128 per 100,000 individuals [50] and results in progressive loss of central vision. It is the commonest inherited macular dystrophy in adults and children [51] and is highly heterogeneous both clinically and genetically with over 900 disease-causing variants reported to date [52] (Figure 1.14).

The gene for Stargardt disease, *ABCA4*, encodes for an ATP-binding cassette superfamily transporter, also known as ABCA4. This transmembrane protein, ABCA4, localises to the photoreceptor outer segment disc membranes [53, 54] and is involved in the visual cycle; a chain of biochemical reactions that lead to the regeneration of visual pigments following light exposure. Under normal conditions ABCA4 is believed to actively transport N-retinylidene-phosphatidylethanolamine (Ret-PE), a by-product of the visual cycle, across the outer segment disc membrane into the cell cytoplasm. In Stargardt disease, accumulation of Ret-PE leads to a build up of diretinoid-pyridinium-phosphatidylethanolamine (A2PE). Following phagocytosis of photoreceptor outer segments by the RPE cells, although A2PE can be hydrolysed to diretinoid-pyridinium-ethanolamine (A2E), a major component of lipofuscin, it cannot be degraded further. Accumulation of A2E, which is toxic to RPE cells, results in apoptosis and consequent photoreceptor degeneration [54] (Figure 1.15).

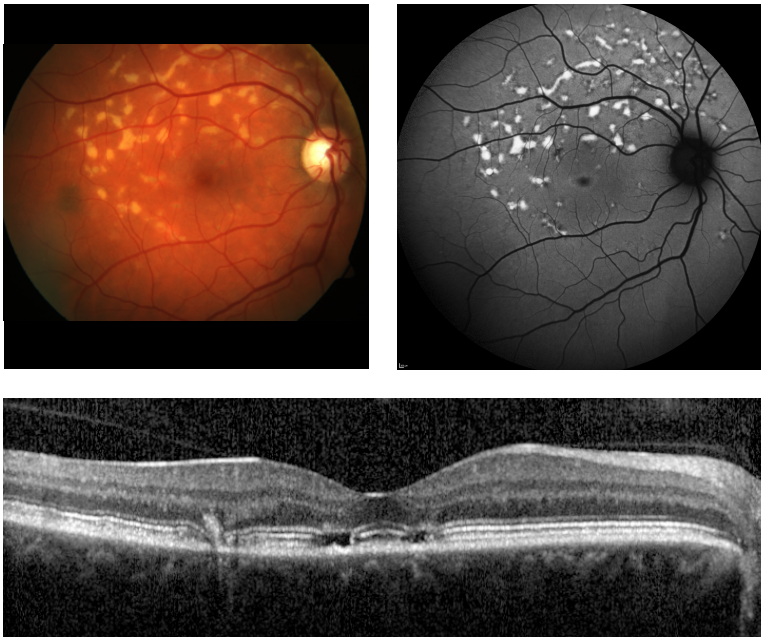


Figure 1.14: Example of conventional clinical retinal images from a patient with Stargardt disease. **Top left**, colour fundus photograph of the right eye centred on the macula showing the macroscopic yellow-white retinal flecks typically found in patients with Stargardt disease (50° field of view). **Top right**, short-wavelength fundus autofluorescence image of the same eye showing hyper-autofluorescence of the retinal flecks (55° field of view). **Bottom**, spectral-domain optical coherence tomography image showing a cross section through the fovea of the right eye. Disruption of the outer retinal layers around the foveal centre can be seen.

Although there are currently no approved treatments for Stargardt disease there are a number of ongoing clinical trials involving gene replacement therapy, cell-based therapy and pharmacotherapy [51]. Cell-specific biomarkers of disease are becoming increasingly important for monitoring outcomes of these new therapies and AO retinal imaging systems could be used to provide highly sensitive cell-based measures of retinal structure and function.

1.6 Challenges of AOSLO imaging

Although AOSLO has enabled visualisation of individual cells in the living human eye, there are a number of challenges associated with introducing AOSLO to routine imaging of the retina. This is particularly true when imaging patients with retinal disease who often have loss of central vision and unstable fixation, which makes visual tasks difficult and tiring.

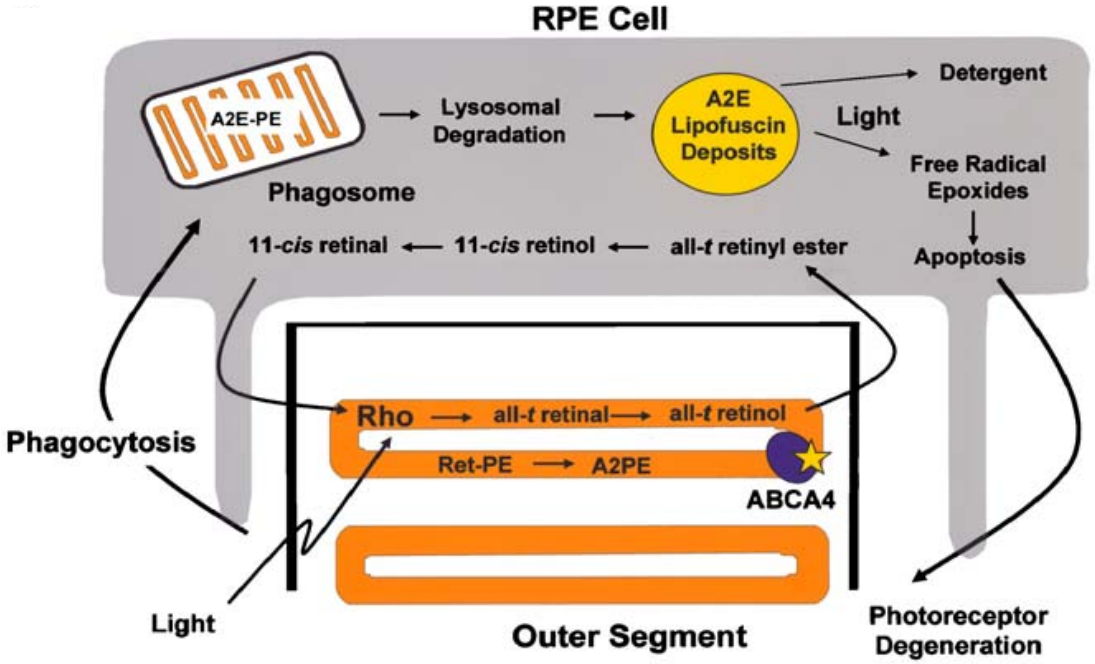


Figure 1.15: Role of ABCA4 in Stargardt disease (adapted from [55]). The photobleaching of rhodopsin (Rho) leads to the conversion of 11-*cis*-retinal to all-*trans*-retinal. All-*trans*-retinal is converted to all-*trans*-retinol, which crosses the subretinal space into the retinal pigment epithelial (RPE) cell where it is recycled back to 11-*cis*-retinal. All-*trans*-retinal also reacts with phosphatidylethanolamine to produce N-retinylidene-phosphatidylethanolamine (Ret-RPE). The abnormal ABCA4 leads to a build up of Ret-RPE within the outer segment disc lumen, and the production of diretinoid-pyridinium-phosphatidylethanolamine (A2PE). Phagocytosis of photoreceptor outer segments leads to a build up of diretinoid-pyridinium-ethanolamine (A2E) within RPE cells, a major component of lipofuscin, leading to apoptosis and subsequent photoreceptor degeneration.

1.6.1 Signal to noise ratio

Increasing the intensity of light illuminating the retina improves image quality by improving the image signal to noise ratio (higher signal intensity). However, the corneal light exposure, and therefore the signal returning from the retina, is restricted by light safety limits. This limits the signal to noise ratio of single unprocessed AOSLO frames (Figure 1.16A), making it necessary to align and average multiple frames (Figure 1.16B).

1.6.2 Ocular media opacities

Ocular media opacities (described in Section 1.3.5.1) reduce signal intensity by absorbing and scattering light, which reduces image contrast. Due to the already

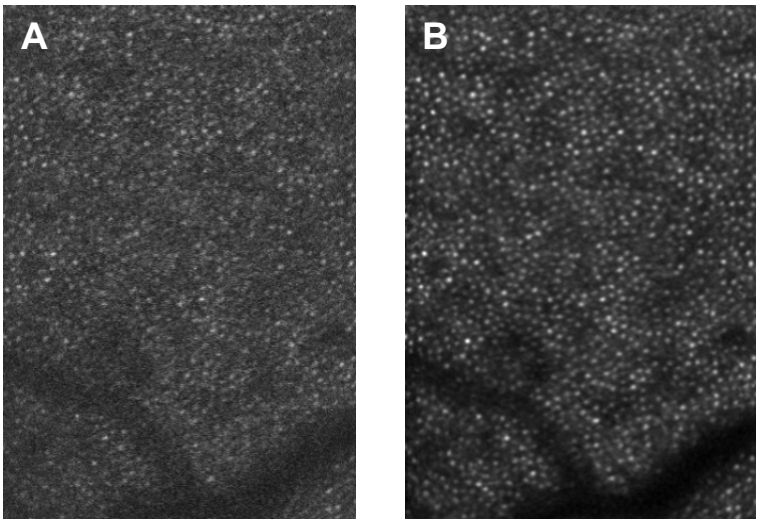


Figure 1.16: An example showing the improvement in signal to noise ratio of AOSLO images with averaging. **A** - a single raw frame, **B** - an image produced by averaging 30 spatially aligned raw frames.

limited signal intensity of single frames, ocular media opacities have a substantial effect on AOSLO image quality. Without the presence of good retinal structure in single frames, eye movement correction and spatial alignment is not possible. So there is a threshold level of signal required in single frames to be able to benefit from averaging.

1.6.3 Deformable mirror stroke

The stroke of a deformable mirror is the maximum possible displacement of its surface and limits the maximum amplitude of the wavefront correction. In participants with high refractive error, the stroke of the deformable mirror may be inadequate to fully correct the ocular wavefront aberrations, which may prevent diffraction-limited imaging and provide insufficient resolution for cellular measurements.

1.6.4 Retinal anatomy

The wavefront sensor is strongly driven by the greatest signal intensity, which in a normal retina is thought to originate from the photoreceptor outer segments [45]. In patients with retinal disease, areas of degeneration can be hyper-reflective and therefore drive the wavefront measurements and correction away from the

layer of interest. These reflections may originate from a retinal plane other than the photoreceptor inner and outer segment junction, causing the resultant images to be away from the desired focal plane (leading to images of retinal structure from an unknown retinal layer) or making the AO control unstable (leading to images with substantial residual aberration).

1.6.5 Small field of view

Due to the small field of view of single images, which is required to achieve the sampling rates for cellular imaging, images from a number of partially overlapping retinal locations are usually required to allow subsequent montaging. This increases the total time taken for an imaging session, and demands additional image processing. Fixational eye movements, which occur during image acquisition on a similar time scale to the scanning rate, cause distortion of AOSLO images and are exacerbated by the small field of view.

1.6.6 Imaging range

When imaging retinal structures at increasing distances from the visual axis, the pupil appears increasingly elliptical to the imaging system and light enters the eye and reflects from the retina at an increasingly oblique angle, making accurate wavefront sensing and correction difficult.

1.7 Challenges of AOSLO image processing

In addition to the challenges associated with image acquisition, there are also a number of challenges associated with the processing of AOSLO images, particularly from patients with retinal disease.

The eye is continuously moving, even during fixation of a stationary target. These fixational eye movements, which occur on a time scale similar to the AOSLO scanning rate, lead to distortion of the images (Figure 1.17). They are especially pronounced because of the small field of view (approximately 1° to 3°) required to achieve sufficient sampling rates for cellular retinal imaging [56]. The removal

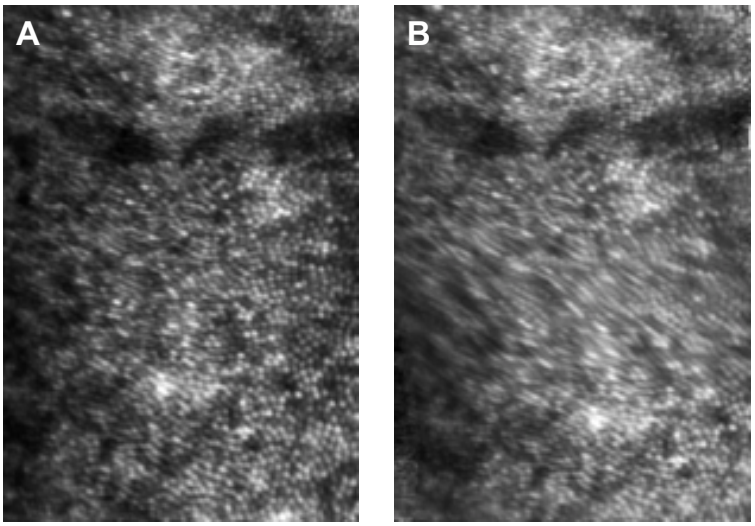


Figure 1.17: Eye movement related AOSLO image distortion. **A** - minimally distorted image. **B** - image from the same retinal location showing eye movement related distortion centrally in the image.

of these distortions is an essential precursor to obtaining an accurate visualisation of retinal structures.

The accurate measurement and subsequent correction of eye movement related distortions requires a reference frame to be selected. Identifying the best reference frame from an image sequence is challenging. Furthermore, as with any image processing pipeline, errors at any stage will propagate along the pipeline impairing image quality.

The need to average images to improve their signal to noise ratio, along with the magnification of eye movement related distortion that occurs due to the small field of view means that participants with stable fixation are preferred for cellular retinal imaging with an AOSLO. Patients with macular disease are known to have unstable fixation [57–59] compared to normal observers, making AOSLO imaging more challenging in this population. In addition, comprehensive assessment of a retinal region requires montaging several averaged images, which is also challenging in diseased samples due to degraded retinal structure.

1.8 Challenges of AOSLO image analysis

Quantifying changes in the photoreceptor mosaic is useful for providing diagnostic and prognostic information about retinal diseases, for understanding disease mechanisms, and for evaluating novel treatments. Photoreceptor-based biomarkers, such as density, geometry, reflectivity and size, have been proposed for analysing changes in the photoreceptor mosaic [45]. Identification of individual photoreceptors is often the first step in generating these quantitative metrics of the photoreceptor mosaic. Photoreceptor identification is typically performed manually, which is both time consuming and subjective. It is challenging in patients with retinal disease as photoreceptors are not as readily identifiable as in healthy eyes [60] and can result in low repeatability without a grader experienced with images of retinal disease [45]. The significant effort required for the analysis of these images is a major rate-limiting step in the clinical application of AO retinal imaging [61, 62] and has been the motivation behind the development of automated analytical tools.

A number of algorithmic approaches for automated cone identification in AO images have been proposed [63–67]. However, while they demonstrate good performance for the specific task they were designed for, their reliance on a predefined set of rules limits their generalisability. Many of these approaches have been developed for cone identification from normal retinae and their performance in diseased retinae has not been validated. Automated cone identification is less reliable in the presence of rod photoreceptors or when RPE cells are visible in eyes with retinal disease [45]. Many algorithms developed for automated cone detection still require accurate supervision when detecting cones in diseased eyes or low-quality images [62]. Automated algorithms that are able to identify rods and cones within the same image are required [67]. Algorithms that can use multimodal images of the photoreceptor mosaic may help to improve reliability and repeatability of photoreceptor identification in images from patients with retinal disease [45].

Unlike traditional programming that uses explicit instructions to perform a task, machine learning algorithms build statistical models to learn patterns within sample data in order to perform a specific task. Deep learning, a subfield of machine

learning, provides an effective way of automatically extracting and learning features directly from the image data, potentially increasing generalisability.

1.9 Research aims and thesis outline

The overarching aim of this work is to develop AOSLO imaging as a clinical tool for patients with retinal disease. In order to achieve this, the relationship between conventional multimodal retinal imaging and AOSLO in participants with Stargardt disease was investigated, and fully automated algorithms for reference frame selection and cone identification were developed and validated.

This thesis comprises five experimental chapters (Chapters 2 to 6) and a summary and conclusion (Chapter 7) as follows:

1.9.1 Chapter 2

Chapter 2 describes a cohort of participants with Stargardt disease that have undergone AOSLO imaging (the results of which are presented in Chapter 6), and the analysis and interpretation of their conventional clinical retinal images [68].

1.9.2 Chapter 3

Chapter 3 describes the participant selection criteria outlined in Chapter 2 that were chosen to maximise the likelihood of successful AOSLO imaging. It also discusses the development of these criteria and of image acquisition protocols that occurred throughout the study and presents recommendations for future participant recruitment. These were developed through experience and expertise, to optimise AOSLO image quality in patients with retinal disease.

1.9.3 Chapter 4

Chapter 4 describes the development and validation of a fully automated reference frame selection algorithm for confocal AOSLO images of the photoreceptor mosaic.

1.9.4 Chapter 5

Chapter 5 describes the development and validation of a fully automated cone identification algorithm for confocal AOSLO images using deep learning.

1.9.5 Chapter 6

Chapter 6 describes the analysis and interpretation of AOSLO images from participants with Stargardt disease. It also explores the application of the reference frame selection and cone identification algorithms (described in Chapter 4 and Chapter 5, respectively) to AOSLO images from participants with Stargardt disease.

1.9.6 Chapter 7

Chapter 7 summarises the major results from the previous experimental chapters, discusses the work presented in this thesis within the context of the wider field and presents directions for future study.

2

The participant cohort

Contents

2.1 Contributions 26
2.2 Background 27
2.3 Methods 28
2.3.1 Participants 28
2.3.2 Clinical assessment 29
2.3.3 Clinical data analysis 31
2.3.4 Statistical analyses 34
2.4 Results 34
2.4.1 Demographics 34
2.4.2 Retinal phenotype of participants with Stargardt disease 37
2.4.3 Preferred retinal locus 47
2.5 Discussion 49
2.5.1 Hyper-reflective foci 49
2.5.2 Leading edge of disease 52
2.5.3 Clinical retinal imaging 53
2.5.4 Macular microperimetry 53
2.5.5 Limitations 54

2.1 Contributions

My contributions to this work include: study design; obtaining ethical approvals; participant recruitment; acquisition, analysis and interpretation of data. The automated tool used for multimodal clinical image registration was provided by

Iain Wilson. I had full access to all of the data presented in this chapter and take responsibility for the integrity of the data and the accuracy of the data analysis.

2.2 Background

The use of conventional clinical retinal imaging has become standard care for the diagnosis and management of patients with retinal disease. The improvement in resolution using an adaptive optics scanning laser ophthalmoscope (AOSLO) has already revealed evidence of disease progression before visible changes on conventional retinal imaging or a measurable reduction in visual function [5].

Stargardt disease results in the progressive loss of central vision. The genetic underpinnings of the disease have been well documented and are localised to the *ABCA4* gene. The heterogeneity of the retinal phenotype parallels the over 900 disease causing variants within this locus that have been reported to date [52]. Stargardt disease represents an interesting target population for AOSLO imaging because it is an inherited disease prioritised for intervention. The potential for new therapies creates a demand for sensitive and reliable cell-specific biomarkers (such as photoreceptor density or reflectivity) that can be provided by AO retinal imaging systems.

Despite being first described over 100 years ago [69], the exact sequence of pathophysiological events that eventually leads to photoreceptor and retinal pigment epithelium (RPE) cell death in Stargardt disease is still not entirely understood. Natural history studies and models of temporal disease progression suggest centrifugal expansion from mild disease confined to the macula to an extreme of potentially severe retina-wide disease [70–72]. The leading disease edge represents the transition from an area of relatively normal to degenerated retina and the advancing front of retinal atrophy. Given the centrifugal nature of disease progression, studying the leading disease edge and surrounding retina may provide a temporal cross section into the sequence of events that lead to photoreceptor and RPE cell death.

The purpose of this study was to generate high-resolution *in vivo* retinal images from participants with Stargardt disease. This chapter aims to characterise the phenotypic heterogeneity and investigate the leading disease edge in a cohort of participants with Stargardt disease using multimodal conventional retinal imaging. Results of the AOSLO imaging will be presented and discussed in Chapter 6. The cohort of participants presented in this chapter represents the first group of individuals with retinal disease to successfully undergo retinal imaging using the custom-built confocal AOSLO described by Young *et al.* [26].

2.3 Methods

National Health Service Research Ethics Committee approval was obtained from the South Central – Oxford B Research Ethics Committee (reference - 15/SC/0144) and the Yorkshire & The Humber - Sheffield Research Ethics Committee (reference - 16/YH/0265). All participants provided written informed consent prior to participation in this study.

2.3.1 Participants

Patients from the Oxford Eye Hospital with a clinical diagnosis of Stargardt disease and at least one pathogenic variant in the *ABCA4* gene, and normal age-matched controls, were included in this study. Participants younger than 16 years of age, with a refractive error greater than six dioptres of sphere, who were unable to maintain stable fixation on a visual target, with evidence of ocular media opacities, or with any concurrent retinal disease were excluded. These participant selection criteria were chosen to maximise the likelihood of successful AOSLO imaging and are discussed further in Chapter 3.

2.3.2 Clinical assessment

All participants underwent a comprehensive ophthalmic clinical assessment, which included the measures listed in Table 2.1. Further details on the clinical assessment undertaken are provided below where required.

Table 2.1: Ophthalmic clinical assessment.

Measure	Instrument
BCVA	LogMAR visual acuity chart
Colour photography	TRC-50DX Mydriatic Retinal Camera, Topcon
SW-FAF	Spectralis, Heidelberg Engineering
NI-FAF	Spectralis, Heidelberg Engineering
SD-OCT	Spectralis, Heidelberg Engineering
Macular MP	Macular Integrity Assessment, CenterVue
Axial length	Zeiss IOL Master, Carl Zeiss Meditec

BCVA - best corrected visual acuity, LogMAR - Logarithm of the Minimum Angle of Resolution, MP - microperimetry, NI-FAF - near-infrared fundus autofluorescence, SD-OCT - spectral-domain optical coherence tomography, SW-FAF - short-wavelength fundus autofluorescence.

2.3.2.1 Best corrected visual acuity (BCVA)

The minimum angle of resolution is the smallest visual angle between two point sources in order for them to be resolved on the retina. A participant who is able to resolve details as small as 1 minute of visual angle will record a Logarithm of the Minimum Angle of Resolution (LogMAR) visual acuity of 0.0 as the base-10 logarithm of 1 is 0. Similarly, a participant with reduced visual acuity who is able to resolve details as small as 10 minutes of visual angle will record a LogMAR visual acuity of 1.0 as the base-10 logarithm of 10 is 1.

2.3.2.2 Autofluorescence imaging

Excitation wavelengths of 488 nm or 790 nm were used for short-wavelength fundus autofluorescence (SW-FAF) and near-infrared fundus autofluorescence (NI-FAF), respectively, with cut-off filters at 500 nm or 830 nm, respectively, to separate excitation and fluorescence light. Images were captured with a 55° and 30° field of view centred on the macula.

NI-FAF imaging was not initially included in the clinical imaging protocol. During the early phase of the study NI-FAF became more routinely used in clinical practice. It was felt that NI-FAF could provide additional phenotyping

information and so it was subsequently added to the clinical imaging protocol for all relevant study visits.

2.3.2.3 Spectral-domain optical coherence tomography imaging

Spectral-domain optical coherence tomography (SD-OCT) imaging was performed with a centre wavelength of 870 nm and images were captured in high speed mode (each B-Scan consisting of 768 A-Scans), which provided a lateral resolution of 11 µm/pixel.

2.3.2.4 Macular microperimetry

Macular microperimetry uses fundus tracking to provide data with improved spatial localisation that is independent of eye movements and is used to evaluate macular function by assessing a participant's threshold sensitivity and fixation behaviour. This enables macular microperimetry to provide an objective assessment of a participant's fixation stability. During microperimetry testing a series of stimuli of varying intensity are projected and the participant is instructed to press a handheld button to indicate when they have detected a stimulus. Macular microperimetry was performed under mesopic (low) light conditions, prior to pupil dilation, and after 20 minutes of dark-adaptation. Mesopic lighting conditions were chosen as the microperimeter operates with a background illumination of 1.3 cd/m^2 [73]. Dark-adaptation was performed to enable photoreceptors to recover their maximum light sensitivity by allowing photopigments to regenerate. The duration of dark-adaptation was chosen based on previously published reports of photoreceptor physiology [73, 74]. All examinations were performed monocularly, with patching of the fellow eye. A test run in each eye was performed to familiarise participants with the microperimeter and to minimise any effect of learning on the results. A 10-2 grid was used with a Goldmann III stimulus, whose diameter is 0.43° of visual angle, and a 4-2 projection strategy (explained below). The 10-2 testing grid consists of 68 points covering the central 20° of the visual field (10° temporal and nasal to fixation) with testing points spaced 2° apart. The 4-2 projection strategy (a staircase method) changes the stimulus light intensity in 4 dB steps

until there is a perceived change (from seen to not seen or vice versa) at which point the stimulus intensity changes in 2 dB steps.

2.3.3 Clinical data analysis

An automated tool, developed using a Generalized Dual Bootstrap-ICP Algorithm [75], was used for multimodal clinical image registration (SD-OCT, SW-FAF, NI-FAF and macular microperimetry).

2.3.3.1 Fundus autofluorescence

In healthy individuals, SW-FAF and NI-FAF signal intensity shows a characteristic pattern (Figure 2.1). The SW-FAF signal within the central macula and especially at the fovea is reduced due to absorption by macular pigments. Conversely, the NI-FAF signal is increased within the central macula due to increased optical density of melanin [17]. See Section 1.3 for a discussion on the relationship between the retinal fluorophores lipofuscin and melanin, and the SW-FAF and NI-FAF signals. Outside the central macula, the SW-FAF and NI-FAF signals remain relatively uniform.

Fundus autofluorescence imaging provides a density map of ocular fluorophores (lipofuscin and melanin) in the retina. Knowing the distribution of these ocular fluorophores in the normal retina is key to understanding fundus autofluorescence images in retinal disease. The relative intensity of fundus autofluorescence images corresponds to the accumulation (or loss) of ocular fluorophores that can occur in retinal disease.

All fundus autofluorescence images in this study were analysed qualitatively by an experienced clinical grader. Qualitative descriptions of the changes in autofluorescence signal intensity in many retinal diseases have made substantial contributions to their diagnosis and monitoring [76–79]. Quantitative analysis of short-wavelength or near-infrared autofluorescence in fundus images requires custom-built instruments [79] that are not currently available for routine clinical use.

2.3.3.2 Spectral-domain optical coherence tomography

The assignment of SD-OCT features used in this study (described in Figure 2.1) to retinal anatomic landmarks (described in Figure 1.4) is in accordance with published nomenclature for normal posterior segment SD-OCT [80]. A consensus regarding the anatomical correlate of the ellipsoid zone has not been reached and it has been attributed either to the interface between inner and outer photoreceptor segments and/or to the photoreceptor ellipsoid component with a high mitochondrial density [80, 81]. The two hyper-reflective bands that make up the RPE/Bruch's membrane complex correspond to the RPE and Bruch's membrane, respectively, but are usually not separable and are seen as a single hyper-reflective band.

SD-OCT provides a non-invasive technique for high-resolution, cross-sectional imaging of the retina. Knowing the anatomic landmarks that are identifiable in the normal retina is essential to understanding SD-OCT images in retinal disease. Deviations in the normal reflectivity of SD-OCT images correspond to structural alterations within the retina that occur in retinal disease.

All SD-OCT images in this study were analysed qualitatively by an experienced clinical grader. Analysis of SD-OCT images in clinical practice is typically performed manually with qualitative descriptions of image features [82]. This is due to a lack of the robust automated analysis tools that are required. For example, automated techniques for segmentation analysis are a source of error, particularly in patients with retinal disease, and require manual supervision [83].

2.3.3.3 Macular microperimetry

Fixation points, the points on the retinal surface used by the participant to fixate on the target, were recorded throughout each macular microperimetry examination. The preferred retinal locus (PRL) is the area of the retina that is used for visual tasks. Patients with retinal disease that has affected their central vision usually develop a peripheral PRL that is outside the fovea or macula. The centre of all fixation points recorded after the initial 10 seconds of a macular microperimetry examination was used to determine the PRL position on the retinal surface. A

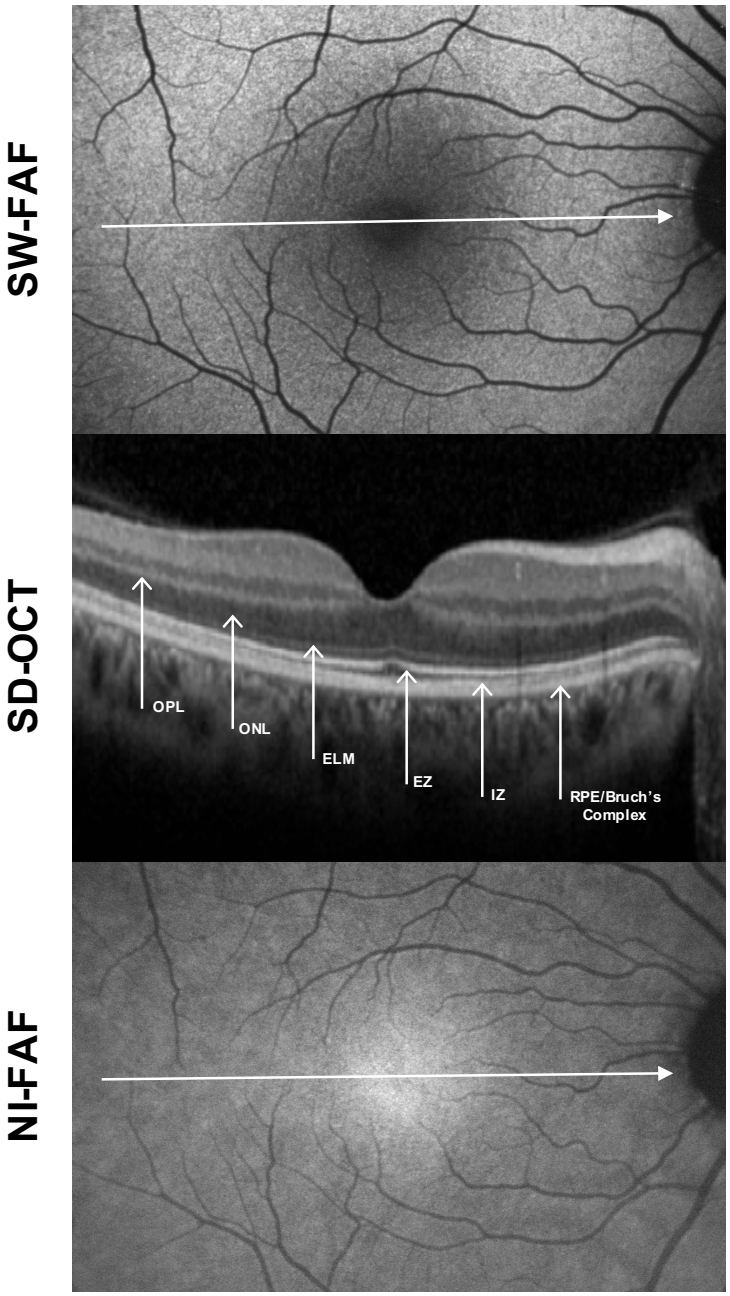


Figure 2.1: Examples of normal short-wavelength fundus autofluorescence (SW-FAF), spectral-domain optical coherence tomography (SD-OCT) and near-infrared fundus autofluorescence (NI-FAF) images from a right eye. The horizontal white arrows in the SW-FAF and NI-FAF images indicate the retinal location of the SD-OCT image. OPL - hyper-reflective outer plexiform layer. ONL - hypo-reflective outer nuclear layer. ELM - hyper-reflective external limiting membrane. EZ - hyper-reflective ellipsoid zone. The interdigitation zone (IZ), a hyper-reflective band, has been proposed to represent photoreceptor outer segment tips and/or the interdigitation of the retinal pigment epithelium (RPE) apical processes with the cone outer segments [80]. RPE/Bruch's membrane complex - hyper-reflective band(s) that make up the RPE/Bruch's membrane complex.

measure of fixation stability, the bivariate contour ellipse area represents an ellipse (expressed in square degrees) that encompasses a proportion of all fixation events [84]. A 63% bivariate contour ellipse was used in this study, which represents an ellipse describing the distribution of 63% of all fixation points. Average retinal sensitivity, which will be referred to as retinal sensitivity, was determined by calculating the mean threshold of all 68 points assessed. Bivariate contour ellipse area and retinal sensitivity were automatically computed and outputted by the microperimeter software. PRL sensitivity was determined by calculating the average threshold of the four macular microperimetry test points centred on the PRL position.

2.3.4 Statistical analyses

Welch's t-test, an independent samples test that does not assume equality of variances, was used to compare between participants with Stargardt disease and control participants for significant differences in PRL distance from the foveal centre, bivariate contour ellipse area and retinal sensitivity. A p-value of less than 5% was considered statistically significant. Bonferroni correction was used for multiple-comparison correction.

2.4 Results

2.4.1 Demographics

Sixteen participants (7 male, 9 female) from 16 families with Stargardt disease (Table 2.2), and 7 controls (1 male, 6 female; Table 2.3) underwent clinical assessment and confocal AOSLO imaging and were included in this study. Of these, 4 participants with Stargardt disease and 3 controls were followed up after their initial visit every 6 months for between 6 and 18 months.

In participants with Stargardt disease the age at symptom onset varied from 5 to 72 years, median best corrected visual acuity was 0.90 LogMAR (interquartile range (IQR) 0.56 LogMAR, range 0.02 LogMAR - 1.08 LogMAR), and 4 participants had hyperopia and 12 had myopia (Table 2.2). Molecular genetic testing identified

Table 2.2: Demographics and baseline clinical characteristics of participants with Stargardt disease.

Participant No./Sex/Age	Age at symptom onset	LogMAR BCVA		Refraction						PRL distance from fovea (°)		63% BCEA (deg^2)		Average sensivity (dB)		Axial length (mm)	
				OD			OS										
		OD	OS	SPH	CYL	AX	SPH	CYL	AX	OD	OS	OD	OS	OD	OS	OD	OS
S01/F/79	72	0.60	0.10	3.00	-1.50	105	3.00	-0.25	075	3.11	0.01	7.00	1.20	6.70	13.10	22.86	22.70
S02/M/31	23	0.40	0.30	-0.25	-0.50	168	∞	-0.25	000	0.04	0.03	0.30	0.20	21.50	23.30	24.53	24.41
S03/M/53	35	0.34	0.04	-2.75	-1.00	155	-2.75	-0.75	175	0.04	0.03	1.60	1.30	11.00	13.60	25.45	25.51
S04/M/46	41	0.02	0.40	0.50	0.25	116	3.00	-1.50	042	0.03	0.09	0.50	0.60	26.40	24.60	23.30	22.66
S05/F/28	15	0.90	0.94	-2.25	-0.50	022	-1.75	-0.75	156	4.42	1.97	0.80	3.60	16.60	18.10	23.68	23.49
S06/F/23	21	0.12	0.66	-0.50	-0.25	010	-0.75	-0.50	020	0.02	1.91	0.70	1.40	23.70	21.20	22.69	22.66
S07/M/20	11	0.80	0.80	-3.07	-0.08	090	-3.09	-0.11	090	4.13	2.94	4.70	3.10	20.90	21.20	25.54	25.58
S08/M/21	11	0.96	0.94	-3.32	-1.09	033	-3.03	-1.95	153	8.06	8.66	5.50	5.70	16.20	15.80	25.26	25.17
S09/F/34	33	0.52	0.08	-2.25	-0.50	085	-2.25	-	-	2.56	0.30	1.80	1.20	23.80	23.40	24.45	24.29
S10/F/33	5	1.04	0.96	1.25	-	-	2.00	-0.50	114	5.30	5.90	4.30	3.80	18.60	20.50	22.15	21.89
S11/F/18	6	0.90	0.96	-1.00	-2.00	180	-1.00	-2.00	180	21.35	16.81	14.90	6.40	11.10	-	23.33	23.16
S12/M/20	15	1.02	0.98	0.25	-1.00	002	1.50	-1.00	178	5.53	5.96	4.50	4.90	19.30	29.70	22.61	22.13
S13/F/46	20	1.08	0.94	-2.25	-0.75	021	-2.25	-0.75	138	10.30	9.64	7.60	6.20	19.60	18.40	23.74	23.58
S14/F/45	33	0.96	1.08	-1.00	-0.75	041	-1.25	-1.75	011	4.79	6.62	7.60	3.10	12.20	12.00	23.12	23.31
S15/F/41	17	0.90	0.96	-0.50	-	-	-1.00	-	-	9.06	8.26	1.50	3.40	14.20	14.50	22.96	22.95
S16/M/39	26	1.00	0.94	-2.25	-0.50	009	-1.25	-0.75	007	10.12	8.36	10.00	12.20	16.60	13.80	23.99	24.08

AX - axis, BCEA - bivariate contour ellipse area, BCVA - best corrected visual acuity, CYL - cylinder, LogMAR - Logarithm of the Minimum Angle of Resolution, OD - right eye, OS - left eye, PRL - preferred retinal locus, SPH - sphere.

Table 2.3: Demographics and baseline clinical characteristics of control participants.

Participant No./Sex/Age	LogMAR BCVA		Refraction						PRL distance from fovea (°)		63% BCEA (deg^2)		Average sensivity (dB)		Average PRL sensitvity [SD] (dB)		Axial length (mm)	
			OD			OS												
	OD	OS	SPH	CYL	AX	SPH	CYL	AX	OD	OS	OD	OS	OD	OS	OD	OS	OD	OS
C01/F/30	-0.22	-0.16	-	-	-	-	-	-	0.77	0.35	0.1	0.1	28.0	28.7	30.0 [1.4]	31.0 [3.0]	22.0	21.9
C02/F/21	0.00	0.00	-	-	-	-	-	-	0.59	0.45	0.2	0.2	28.9	29.0	31.3 [1.8]	31.0 [1.4]	22.4	22.6
C03/F/28	-0.36	-0.36	-1.50	-0.75	148	-1.75	-0.25	179	0.49	0.33	0.3	0.3	26.8	27.0	29.3 [1.1]	29.0 [1.4]	24.4	24.3
C04/F/34	-0.12	-0.10	-3.00	-	-	-3.00	-	-	0.44	0.32	0.5	0.3	25.7	27.3	26.5 [1.7]	30.5 [0.9]	25.2	25.2
C05/F/47	-0.10	0.02	-2.25	-	-	-0.25	-	-	0.05	0.16	0.4	0.6	26.6	27.5	28.5 [1.7]	27.0 [0]	24.0	23.3
C06/M/45	-0.32	-0.32	0.50	-0.25	090	0.25	-	-	0.17	0.51	0.2	0.1	26.8	26.6	29.0 [1.0]	28.5 [0.9]	23.7	23.9
C07/F/36	-0.20	-0.22	∞	0.50	075	∞	0.25	088	0.53	0.47	2.2	1.2	26.2	27.3	28.5 [0.9]	29.5 [0.9]	22.0	22.0

AX - axis, BCEA - bivariate contour ellipse area, BCVA - best corrected visual acuity, CYL - cylinder, LogMAR - Logarithm of the Minimum Angle of Resolution, OD - right eye, OS - left eye, PRL - preferred retinal locus, SPH - sphere.

36 variants in *ABCA4* (Table 2.4). Variants in genes other than *ABCA4* were identified in one participant (S09; Table 2.4).

2.4.2 Retinal phenotype of participants with Stargardt disease

The retinal phenotypes of participants with Stargardt disease are summarised in Table 2.5 and illustrated in Figures 2.2, 2.3, 2.4 and 2.5.

2.4.2.1 Short-wavelength fundus autofluorescence

All participants showed a common phenotype of SW-FAF abnormalities with central macular hypo-autofluorescence surrounded by a ring of hyper-autofluorescence. Thirteen participants showed additional SW-FAF abnormalities with a reticular pattern consisting of hyper-autofluorescence flecks and areas of hypo-autofluorescence giving a 'salt and pepper' appearance. These additional SW-FAF abnormalities ranged in extent from within the central macula to external to the vascular arcades. All participants except one (S01) exhibited peripapillary (around the optic disc) sparing of retinal abnormalities.

2.4.2.2 Near-infrared fundus autofluorescence

Eight participants underwent NI-FAF imaging and the observed pattern of autofluorescence abnormalities were similar to those seen with SW-FAF (Figure 2.6). However, the extent of the NI-FAF abnormalities were greater over the same retinal area in comparison to the SW-FAF images.

2.4.2.3 Spectral-domain optical coherence tomography

A common phenotype on SD-OCT of retinal thinning with loss of outer retinal layers (outer nuclear layer to RPE/Bruch's membrane complex) with or without loss of the inner retinal layers was observed in all participants with Stargardt disease. This phenotype ranged in severity and extent from within the central macula to throughout the macula, and in some cases extended to the temporal vascular arcades and beyond. Early features of outer retinal degeneration include thinning of the outer

Table 2.4: Genetic testing results of participants with Stargardt disease.

Participant No.	Gene	Variants*			
S01	ABCA4	c.3004C>T p.(Arg1002Trp)	c.5714+5G>A		
S02	ABCA4	c.5461-10T>C			
S03	ABCA4	c.2588G>C (p.Gly863Ala)	c.4326C>A (p.Asn1442Lys)		
S04	ABCA4	c.5461-10T>C	c.6718A>G (p.Thr2240Ala)		
S05	ABCA4	c.4537dupC (p.Gln1513fs)	c.6391G>A (p.Glu2131Lys)	c.1411G>A (p.Glu471Lys)	c.455G>A (p.Arg152Gln)
S06	ABCA4	c.3259G>A p.(Glu1087Lys)	c.6089G>A p.(Arg2030Gln)		
S07	ABCA4	c.1230A>G (p.Ile410Met)	c.4216C>T (p.His1406Tyr)	c.6148G>C (p.Val2050Leu)	
S08	ABCA4	c.2588G>C (p.Gly863Ala)	c.1715G>A (p.Arg572Gln)	c.5461-10T>C	
S09	ABCA4, RP1	c.5882G>A p.(Gly1961Glu)	c.3259G>A p.(Glu1087Lys)	c.5693G>A p.(Arg1898His)	RP1: c.4603C>G p.(Pro1535Ala)
S10	ABCA4	c.5018+2T>C	c.6316C>T p.(Arg2106Cys)		
S11	ABCA4	c.3064G>A p.(Glu1022Lys)	c.6709A>C p.(Thr2237Pro)		
S12	ABCA4	c.3085C>T p.(Gln1029*)	c.5882G>A p.(Gly1961Glu)		
S13	ABCA4	c.2588G>C p.(Gly863Ala)	c.1A>G p.(Met1Val)		
S14	ABCA4	c.6089G>A p.(Arg2030Gln)	c.5461-10T>C		
S15	ABCA4	c.634C>T (p.Arg212Cys)	c.4457C>T (p.Pro1486Leu)		
S16	ABCA4	c.6079C>T (p.Leu2027Phe)	c.3322C>T (p.Arg1108Cys)		

* all variants are heterozygous.

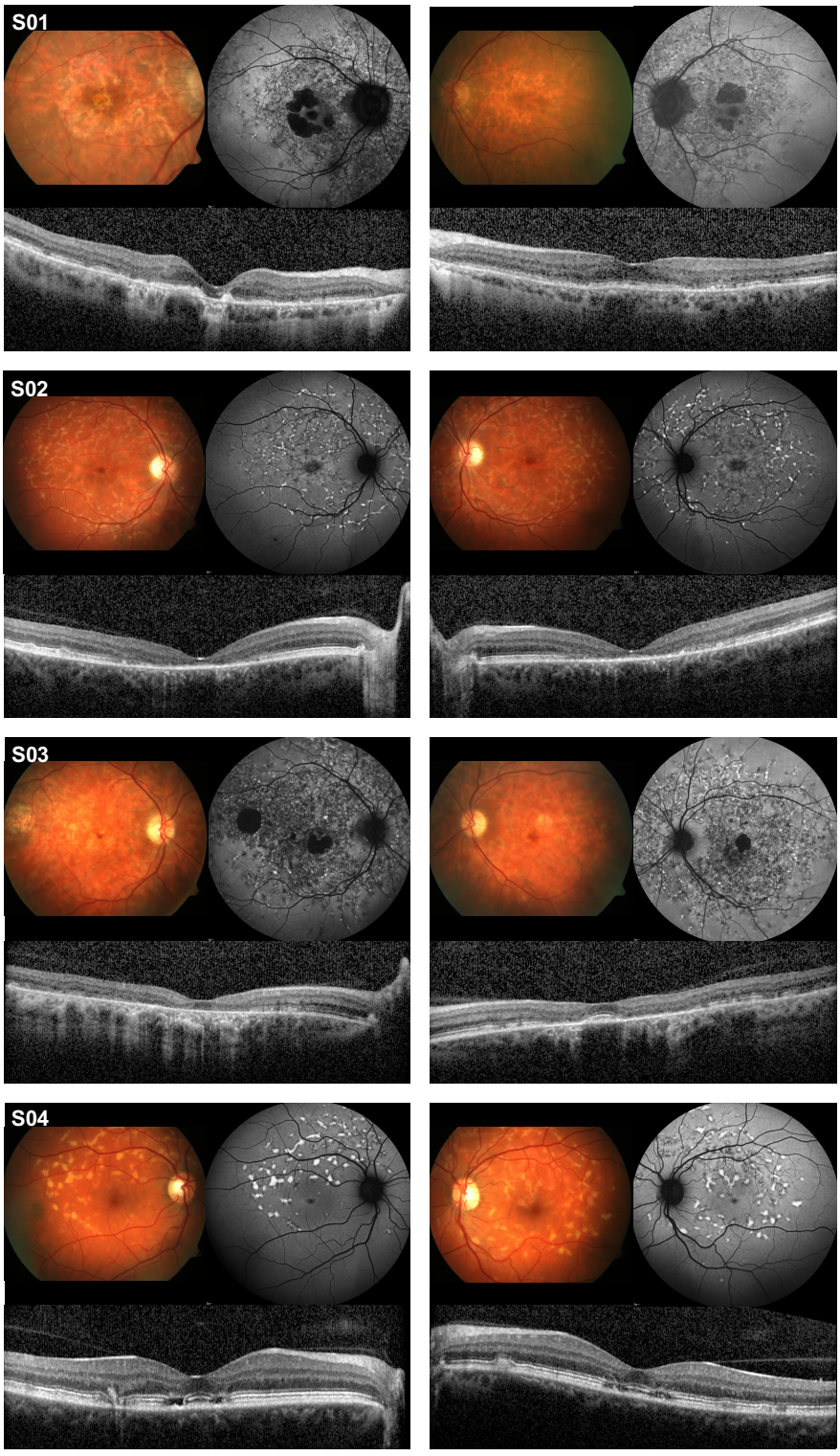


Figure 2.2: Retinal phenotype of participants with Stargardt disease (S01 - S04). For each participant the panels show: 55° short-wavelength fundus autofluorescence (top left); fundus colour (top right); and spectral-domain optical coherence tomography providing a cross-section through the fovea (bottom). Images of both eyes (left column - right eyes; right column - left eyes) are shown.

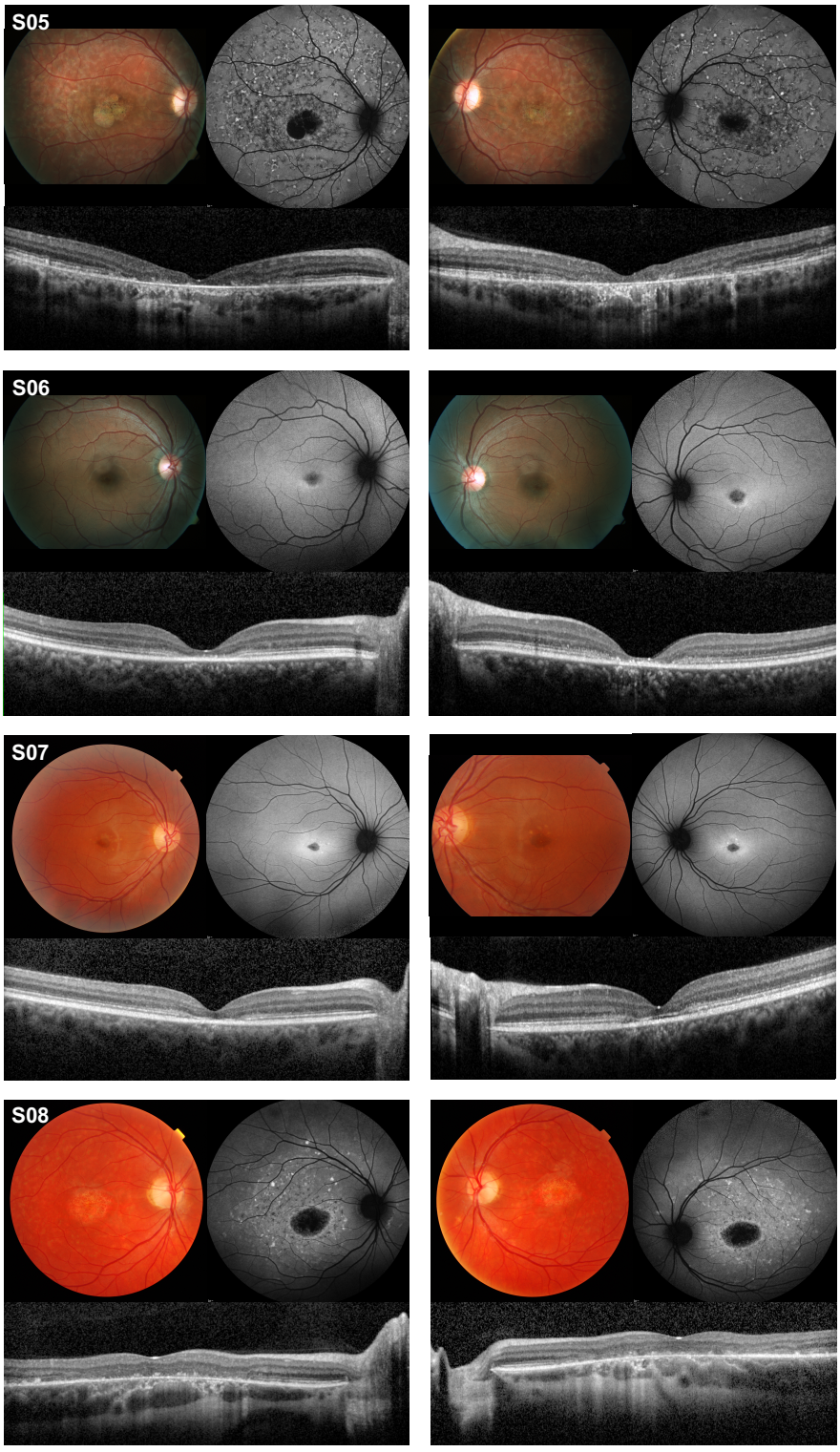


Figure 2.3: Retinal phenotype of participants with Stargardt disease (S05 - S08). For each participant the panels show: 55° short-wavelength fundus autofluorescence (top left); fundus colour (top right); and spectral-domain optical coherence tomography providing a cross-section through the fovea (bottom). Images of both eyes (left column - right eyes; right column - left eyes) are shown.

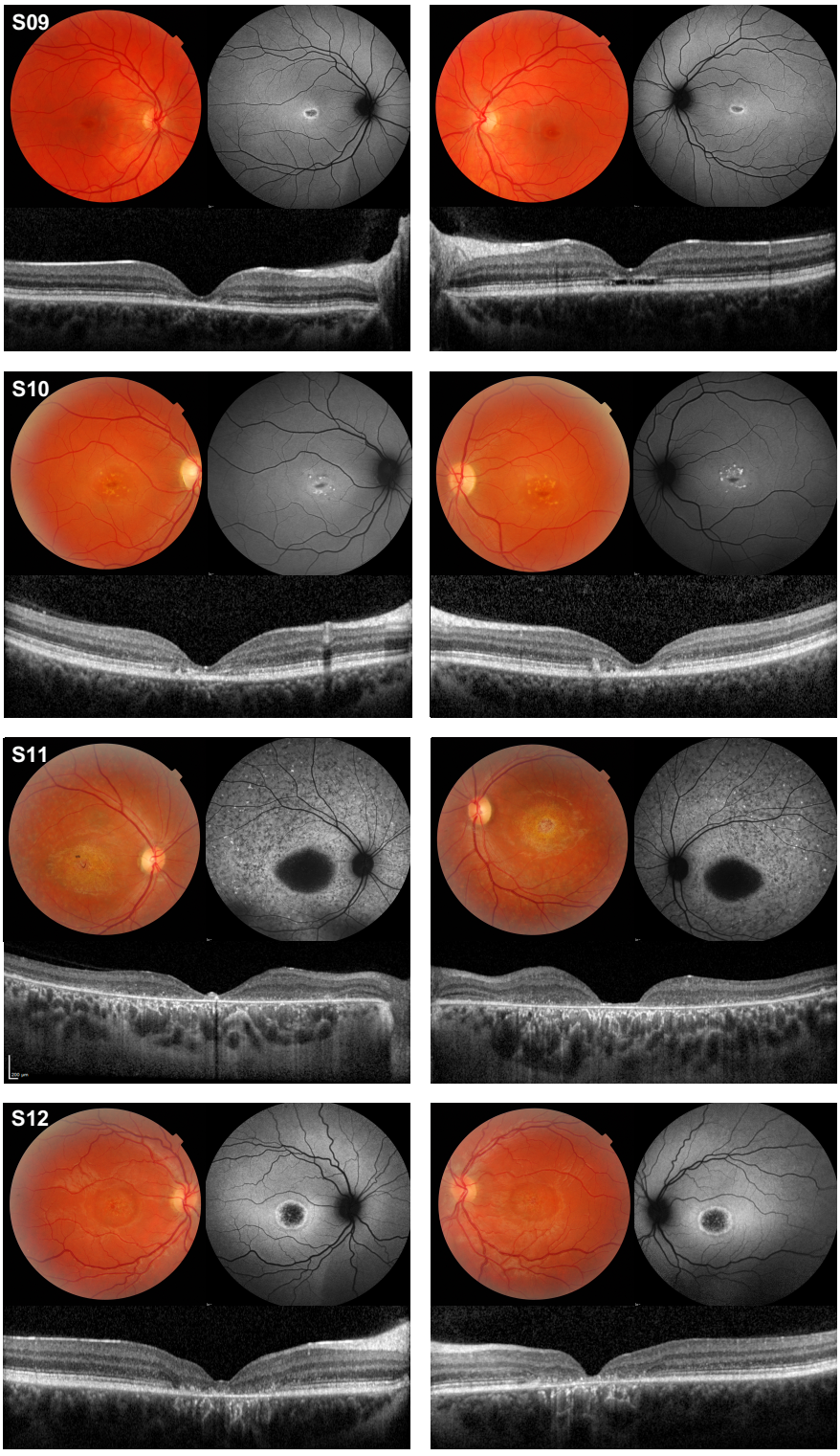


Figure 2.4: Retinal phenotype of participants with Stargardt disease (S09 - S12). For each participant the panels show: 55° short-wavelength fundus autofluorescence (top left); fundus colour (top right); and spectral-domain optical coherence tomography providing a cross-section through the fovea (bottom). Images of both eyes (left column - right eyes; right column - left eyes) are shown.

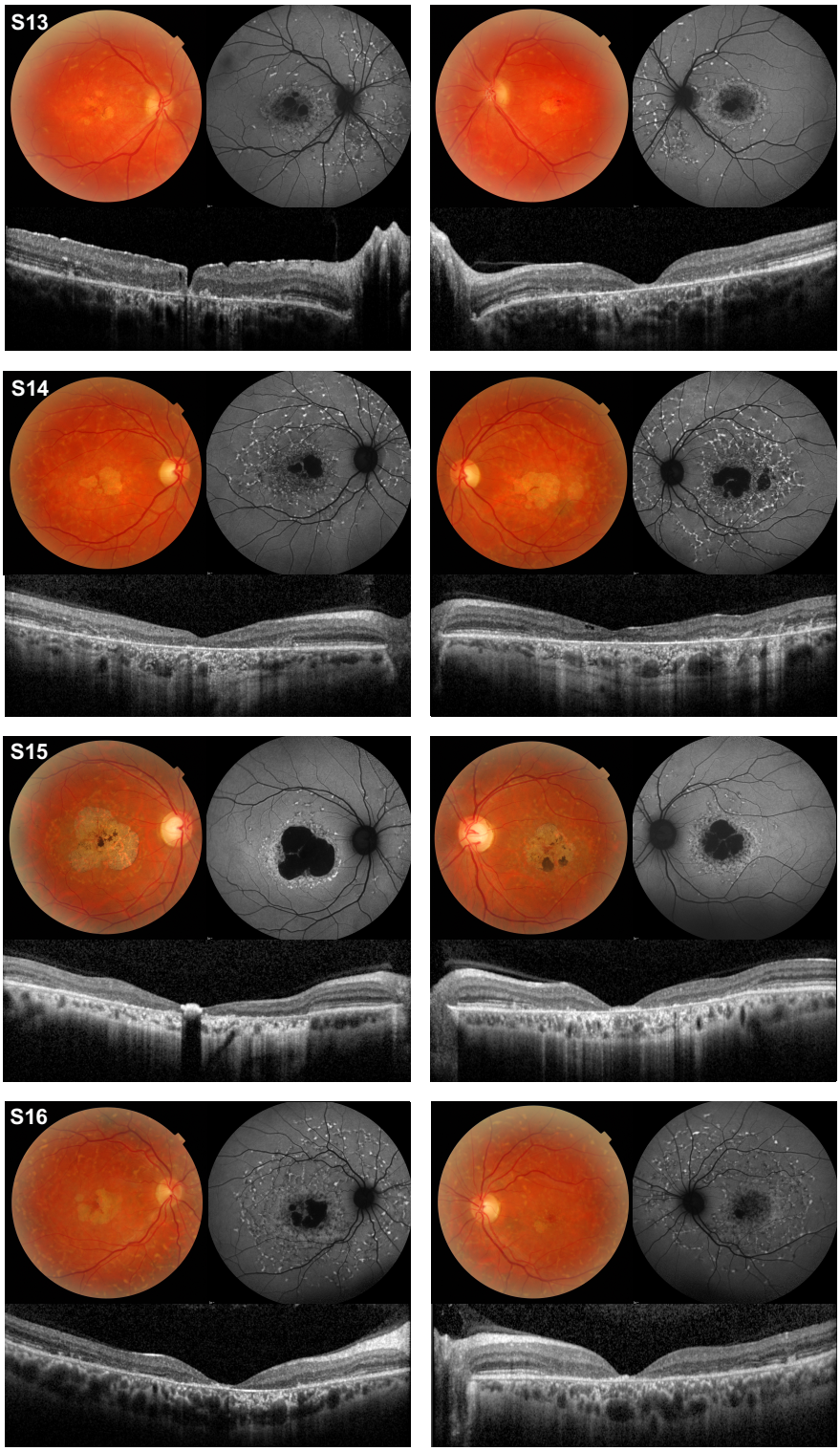


Figure 2.5: Retinal phenotype of participants with Stargardt disease (S13 - S16). For each participant the panels show: 55° short-wavelength fundus autofluorescence (top left); fundus colour (top right); and spectral-domain optical coherence tomography providing a cross-section through the fovea (bottom). Images of both eyes (left column - right eyes; right column - left eyes) are shown.

Table 2.5: Retinal phenotype in participants with Stargardt disease.

Participant	Colour	SW-FAF	NI-FAF	SD-OCT
S01	OU: central macular and peripapillary chorioretinal atrophy.	OU: 'salt and pepper' AF abnormalities extending external to the arcades, peripaillary hypo-AF.	-	OU: common phenotype of SD-OCT abnormalities. OS: foveal sparing.
S02	OU: irregular central macular RPE reflex, retinal flecks.	OU: 'salt and pepper' AF abnormalities extending external to the arcades.	-	OU: common phenotype of SD-OCT abnormalities.
S03	OU: chorioretinal atrophy within the vascular arcades and central macula.	OU: 'salt and pepper' AF abnormalities extending external to the arcades.	-	OU: common phenotype of SD-OCT abnormalities, foveal sparing.
S04	OU: retinal flecks extending external to the arcades.	OU: common phenotype of AF abnormalities, hyper-AF retinal flecks.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities, foveal sparing.
S05	OU: central macular chorioretinal atrophy.	OU: 'salt and pepper' AF abnormalities extending external to the arcades.	-	OU: common phenotype of SD-OCT abnormalities, early features of ORD.
S06	OU: irregular central macular RPE reflex.	OU: common phenotype of AF abnormalities.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities, hyper-reflective foci predominantly within the ONL, early features of ORD.
S07	OU: central macular irregular RPE reflex, retinal flecks.	OU: common phenotype of AF abnormalities.	-	OU: common phenotype of SD-OCT abnormalities, early features of ORD.
S08	OU: central macular chorioretinal atrophy.	OU: 'salt and pepper' AF abnormalities extending external to the arcades.	-	OU: common phenotype of SD-OCT abnormalities, early features of ORD.
S09	OU: irregular central macular RPE reflex.	OU: common phenotype of AF abnormalities.	-	OD: common phenotype of SD-OCT abnormalities, early features of ORD. OS: optical gap phenotype.
S10	OU: central macular irregular RPE reflex, retinal flecks.	OU: common phenotype of AF abnormalities, hyper-AF retinal flecks.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities, hyper-reflective foci predominantly within the ONL, early features of ORD.
S11	OU: central macular chorioretinal atrophy.	OU: 'salt and pepper' AF abnormalities extending external to the arcades.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities.
S12	OU: irregular central macular RPE reflex.	OU: common phenotype of AF abnormalities.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities, hyper-reflective foci predominantly within the ONL, early features of ORD.
S13	OU: central macular chorioretinal atrophy, retinal flecks.	OU: common phenotype of AF abnormalities, hyper-AF retinal flecks.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities. OD: epiretinal membrane.
S14	OU: central macular chorioretinal atrophy, retinal flecks.	OU: common phenotype of AF abnormalities, hyper-AF retinal flecks.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities.
S15	OU: central macular chorioretinal atrophy, retinal flecks.	OU: common phenotype of AF abnormalities, hyper-AF retinal flecks.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities, early features of ORD.
S16	OU: central macular chorioretinal atrophy, retinal flecks.	OU: common phenotype of AF abnormalities, hyper-AF retinal flecks.	OU: common phenotype of AF abnormalities.	OU: common phenotype of SD-OCT abnormalities.

AF - autofluorescence, NI-FAF - near-infrared fundus autofluorescence, OD - right eye, ONL - outer nuclear layer, ORD - outer retinal degeneration, OS - left eye, OU - both eyes, RPE - retinal pigment epithelium, SD-OCT - spectral-domain optical coherence tomography, SW-FAF - short-wavelength fundus autofluorescence.

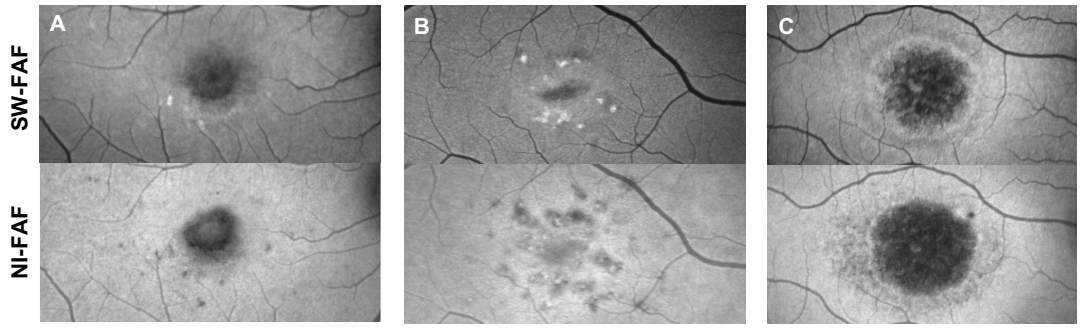


Figure 2.6: An example showing comparison between near-infrared fundus autofluorescence (NI-FAF) and short-wavelength fundus autofluorescence (SW-FAF) images of the same eye. The patterns of autofluorescence abnormalities are similar between the NI-FAF and SW-FAF images, however, in the NI-FAF images they are greater in their extent. All images are of the right eye. **A** - participant S06, **B** - participant S10, **C** - participant S12

nuclear layer (Figure 2.7A); followed by disruption (Figure 2.7B), and subsequent loss (Figure 2.7C) of the ellipsoid zone and/or external limiting membrane. These early degenerative changes were observed in eight participants before any evidence of RPE atrophy on SW-FAF or SD-OCT. Hyper-reflective foci were clearly detected in three participants (S06, S10 and S12), predominantly located within the outer nuclear layer (Figure 2.8), and with preservation of the underlying outer retinal layers (external limiting membrane to RPE/Bruch's membrane complex). Hyper-reflective foci were also observed in areas of severe degeneration and RPE atrophy (Figure 2.8). Three participants (S01, S03 and S04) exhibited foveal sparing in one or both eyes. One participant (S09) exhibited an optical gap phenotype in their left eye with loss of the ellipsoid zone and interdigitation zone and preservation of the underlying RPE (Figure 2.4).

2.4.3 Preferred retinal locus

In participants with Stargardt disease, median PRL eccentricity was 4.60° from the foveal centre (IQR 7.45°, range 0.09° - 21.35°), median 63% bivariate contour ellipse area was 3.5 deg^2 (IQR 4.6 deg^2, range 0.2 deg^2 - 14.9 deg^2), the median retinal sensitivity (for all 68 points assessed) was 18.4 dB (IQR 7.35 dB, range 6.7 dB - 29.7 dB), and median PRL sensitivity was 17.0 dB (IQR 9.0 dB, range

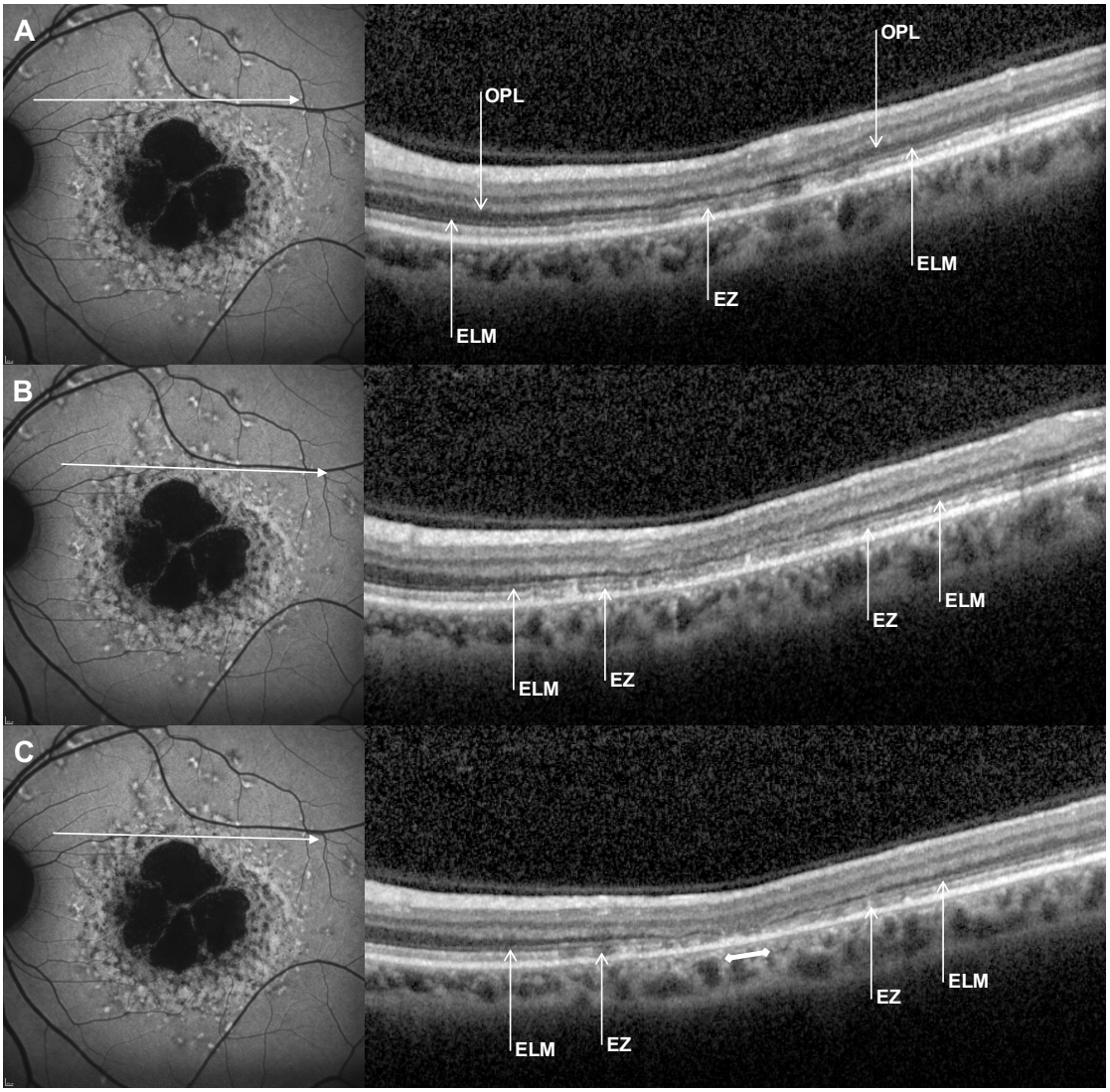


Figure 2.7: Early spectral-domain optical coherence tomography (SD-OCT) features of outer retinal degeneration. Images are from the left eye of participant S15. 30° short-wavelength fundus autofluorescence image (left) with scan lines (indicated by white horizontal arrows) for each SD-OCT image (right). **A** - early features of outer retinal degeneration (furthest from the leading disease edge) showing thinning of the outer nuclear layer (ONL), the hyporeflective zone between the outer plexiform layer (OPL) and the external limiting membrane (ELM). **B** - closer to the leading disease edge showing additional disruption of the ELM and ellipsoid zone (EZ). **C** - closest to the leading disease edge showing loss of the ONL, ELM and EZ overlying the white horizontal double-headed arrow.

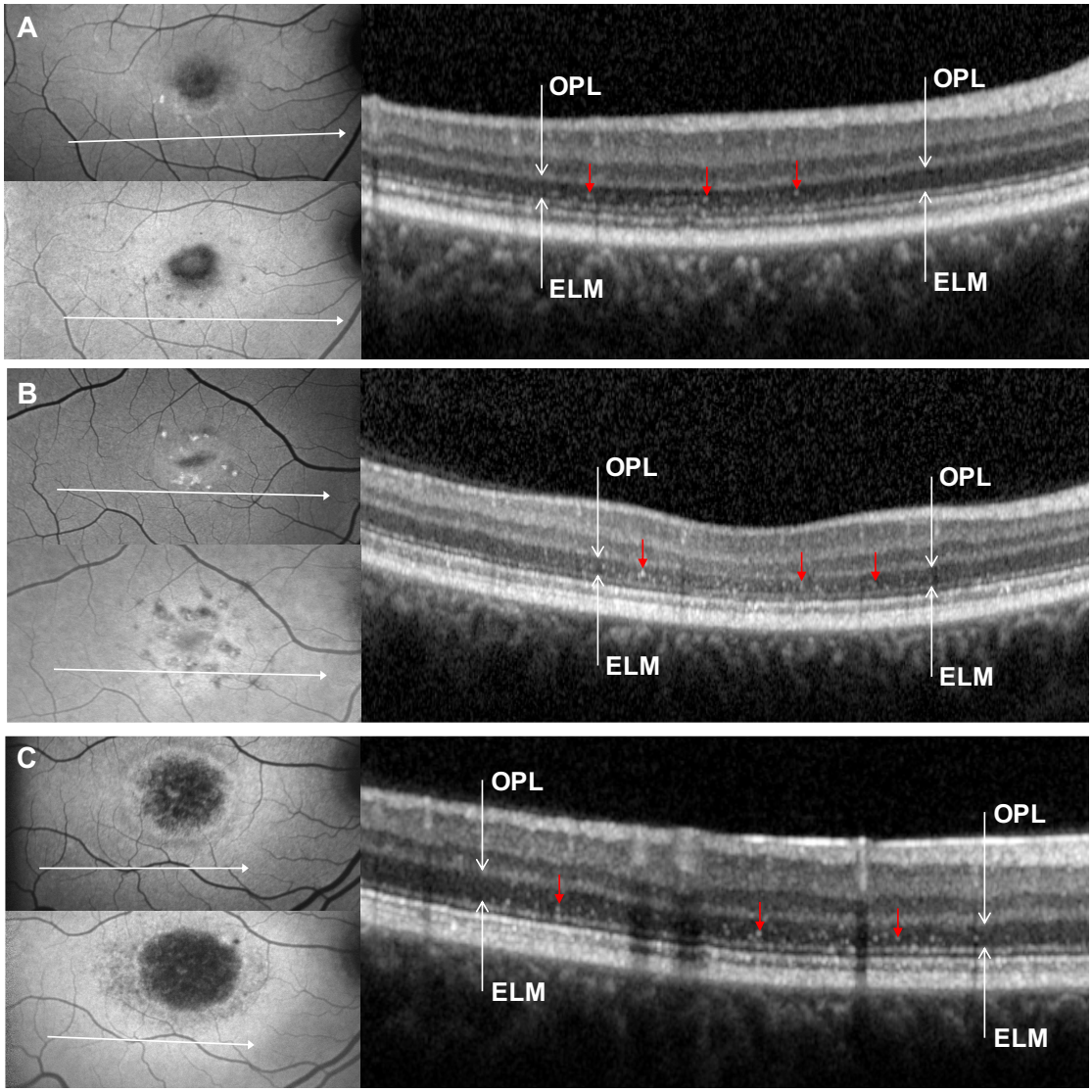


Figure 2.8: Hyper-reflective foci can be seen within the outer nuclear layer (hypo-reflective zone between the outer plexiform layer (OPL) and the external limiting membrane (ELM)) on spectral-domain optical coherence tomography (SD-OCT). Vertical red arrows indicate some examples of the hyper-reflective foci visible. For each participant the panels show: short-wavelength fundus autofluorescence (top left), near-infrared fundus autofluorescence (bottom left), and SD-OCT (right). Horizontal scan lines (white) show the position of each SD-OCT image. All images are from the right eye. **A** - participant S06, **B** - participant S10, **C** - participant S12.

0.3 dB - 31.5 dB). Spearman's correlation coefficients between participants' right and left eyes were 0.90 ($p < 0.001$) for PRL eccentricity, 0.69 ($p = 0.003$) for bivariate contour ellipse area, 0.89 ($p < 0.001$) for retinal sensitivity, and 0.62 ($p = 0.013$) for PRL sensitivity.

2.4.3.1 Retinal structure

The retinal structure at the PRL was variable (Table 2.6). The RPE/Bruch's membrane complex was present in all eyes except one where the PRL position on the retinal surface fell within an area of RPE atrophy. Autofluorescence signal (short-wavelength and near-infrared) was abnormal in most eyes and there was variable presence of the external limiting membrane and ellipsoid zone on SD-OCT. However, in areas of normal autofluorescence signal the external limiting membrane and ellipsoid zone (on SD-OCT) were present.

2.4.3.2 Control participants

In control participants, median PRL eccentricity was 0.45° from the foveal centre (IQR 0.18°, range 0.05°- 0.99°), median 63% bivariate contour ellipse area was 0.3 deg^2 (IQR 0.3 deg^2, range 0.1 deg^2 - 2.2 deg^2), the median retinal sensitivity (for all 68 points assessed) was 27.2 dB (IQR 1.2 dB, range 25.7 dB - 29.0 dB), and the median PRL sensitivity was 29.2 dB (IQR 1.9 dB, range 26.5 dB - 31.3 dB). Preferred retinal locus position, bivariate contour ellipse area, retinal sensitivity and PRL sensitivity differed significantly ($p < 0.001$) between the Stargardt disease and control participant groups. An adjusted p-value of less than 1.25%, after Bonferroni correction, was considered to be statistically significant.

2.5 Discussion

This chapter describes the retinal phenotype of a cohort of 16 participants with Stargardt disease who successfully underwent retinal imaging using a custom-built confocal AOSLO [26]. Results of the AOSLO imaging are presented and discussed in Chapter 6. By combining information from structural and functional

Table 2.6: Retinal structure and function at the preferred retinal locus.

Participant No./Sex/Age	Eye	PRL distance from fovea (°)	Average PRL sensitivity [SD] (dB)	SW-FAF	NI-FAF	SD-OCT		
						ELM	EZ	RPE/ Bruch's
S01/F/79	OD	3.11	1.5 [2.6]	Hypo-AF	-	Absent	Absent	Absent
	OS	0.01	17.0 [2.2]	Hypo-AF	-	Present	Present	Present
S02/M/31	OD	0.04	2.0 [3.5]	Hypo-AF	-	Absent	Absent	Present
	OS	0.03	1.0 [1.7]	Hypo-AF	-	Absent	Absent	Present
S03/M/53	OD	0.04	0.3 [0.4]	Hypo-AF	-	Absent	Absent	Present
	OS	0.03	5.3 [6.9]	Normal	-	Present	Present	Present
S04/M/46	OD	0.03	26.0 [0]	Normal	Hypo-AF	Present	Present	Present
	OS	0.09	24.5 [6.1]	Normal	Hypo-AF	Present	Present	Present
S05/F/28	OD	4.42	21.8 [2.4]	Hypo-AF	-	Absent	Absent	Present
	OS	1.97	10.5 [10.5]	Hypo-AF	-	Absent	Absent	Present
S06/F/23	OD	0.02	3.8 [6.5]	Hypo-AF	Hypo-AF	Absent	Absent	Present
	OS	1.91	13.5 [13.5]	Hypo-AF	Hypo-AF	Present	Present	Present
S07/M/20	OD	4.13	23.0 [1.4]	Normal	-	Present	Present	Present
	OS	2.94	21.0 [2.8]	Hypo-AF	-	Present	Present	Present
S08/M/21	OD	8.06	20.0 [2.2]	Hypo-AF	-	Present	Present	Present
	OS	8.66	20.5 [3.0]	Hypo-AF	-	-	-	-
S09/F/34	OD	2.56	20.0 [7.1]	Hypo-AF	-	Present	Present	Present
	OS	0.30	16.0 [4.2]	Hypo-AF	-	Present	Absent	Present
S10/F/33	OD	5.30	19.0 [1.0]	Normal	Normal	Present	Present	Present
	OS	5.90	19.5 [2.6]	Normal	Normal	Present	Present	Present
S11/F/18	OD	21.35	13.0 [3.7]	Hypo-AF	Hypo-AF	Present	Absent	Present
	OS	16.81	-	Hypo-AF	Hypo-AF	Present	Absent	Present
S12/M/20	OD	5.53	17.3 [10.1]	Normal	Normal	Present	Present	Present
	OS	5.96	31.5 [4.5]	Normal	Normal	Absent	Present	Present
S13/F/46	OD	10.30	22.0 [3.7]	Normal	Normal	Present	Present	Present
	OS	9.64	23.5 [1.7]	Hyper-AF	Hyper-AF	Present	Present	Present
S14/F/45	OD	4.79	0.5 [0.9]	Hypo-AF	Hypo-AF	Absent	Absent	Present
	OS	6.62	14.0 [5.1]	Hypo-AF	Hypo-AF	Absent	Absent	Present
S15/F/41	OD	9.06	15.0 [7.1]	Normal	Normal	Present	Present	Present
	OS	8.26	18.5 [4.3]	Hypo-AF	Hypo-AF	Present	Present	Present
S16/M/39	OD	10.12	17.0 [1.0]	Hypo-AF	Hypo-AF	Present	Present	Present
	OS	8.36	16.5 [0.9]	Hyper-AF	Hypo-AF	-	-	-

AF - autofluorescence, ELM - external limiting membrane, EZ - ellipsoid zone, NI-FAF - near-infrared fundus autofluorescence, OD - right eye, OS - left eye, PRL - preferred retinal locus, RPE - retinal pigment epithelium, SD-OCT - spectral-domain optical coherence tomography, SW-FAF short-wavelength fundus autofluorescence.

testing (SD-OCT, SW-FAF, NI-FAF, microperimetry) a number of early disease features have been identified.

2.5.1 Hyper-reflective foci

Hyper-reflective foci were clearly identified on SD-OCT within the outer nuclear layer in three participants. These features were identified within the transition zone, at a spatial location furthest from the leading disease edge, and in association with qualitative reductions in intensity of the external limiting membrane and ellipsoid zone. The hyper-reflective foci were observed in areas of normal SW-FAF signal, and in two of the three participants, NI-FAF signal was also normal.

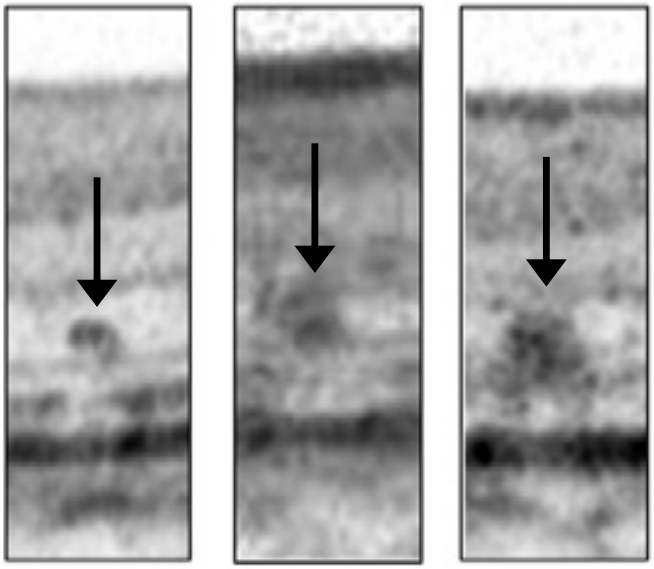
The well-defined yellowish-white lesions known as 'retinal flecks' are a well documented macroscopic feature in patients with Stargardt disease. In 2006, Querques *et al.* used time-domain OCT (with an axial resolution of 8-10 µm) to analyse retinal flecks in patients with Stargardt disease and defined two types of hyper-reflective deposits [85, 86]. Type 1 deposits were in continuum with the RPE and located within the inner (closest to the front of the eye) part of the RPE layer and type 2 deposits were located at the level of the outer nuclear layer and clearly separated from the RPE. This classification was revised in 2010 by Voigt *et al.* using SD-OCT (with an axial resolution of 7 µm) to distinguish five different types of hyper-reflective deposits (Groups A - E) [86, 87]. Groups A - C described lesions that resembled what had previously been classified as type 1 lesions with progressive protrusion of the hyper-reflective material towards and into the outer nuclear layer. Group D lesions resembled those previously classified as type 2 lesions with hyper-reflective material limited to the outer nuclear layer and clearly separate from the RPE. Group E lesions were described as drusen-like retinal pigment epithelial detachments. Similar hyper-reflective deposits, observed with SD-OCT, were also described in the choroid of patients with Stargardt disease by Piri *et al.* in 2015 [86]. In contrast to the retinal flecks described by Voigt *et al.* [87], the hyper-reflective foci identified in this study have different morphologic characteristics - they are smaller in size and do not cast a shadow on deeper retinal

structures on SD-OCT (Figure 2.9). They were also found in areas of near normal retina with uniform SW-FAF signal and an absence of macroscopic retinal flecks.

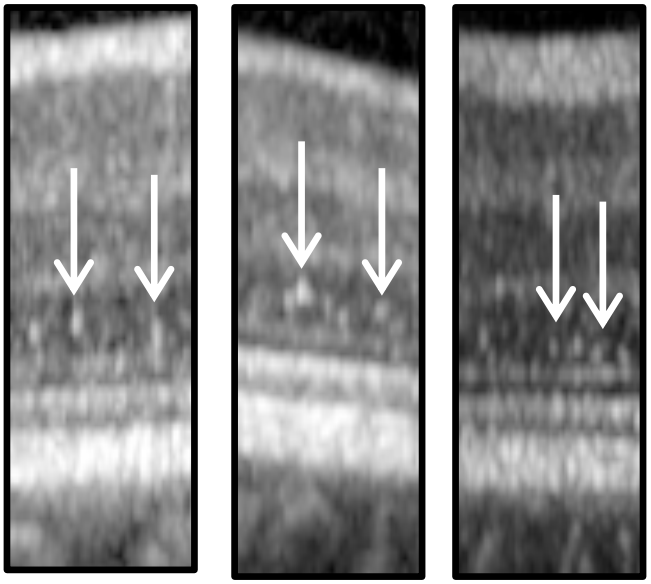
Hyper-reflective foci have also been reported in association with age-related macular degeneration [88], diabetic macular oedema [89, 90], retinal vein occlusion [91] and type 2 idiopathic macular telangiectasia [92]. A number of hypotheses have been suggested regarding the aetiology of these hyper-reflective foci including aggregates of activated microglial cells [88], degenerated photoreceptors [93], precursors of hard exudates [89], or migrating RPE cells [94]. These features have not been previously reported in patients with Stargardt disease.

A number of studies have suggested that retinal flecks are located in the outer retina at the level of the RPE [79, 95–98], making it less likely that the hyper-reflective foci observed in this study are retinal flecks. Multimodal imaging, specifically autofluorescence imaging, allows additional hypotheses to be considered unlikely. For example, the observation of hyper-reflective foci in areas of uniform autofluorescence signal makes the possibility of these being lipofuscin or melanin containing deposits less likely as abnormalities in autofluorescence signal would be expected. The presence of lipofuscin and melanin containing granules within RPE cells makes migrating RPE cells less likely as a source of the hyper-reflective foci. The formation of bisretinoid fluorophores (such as lipofuscin) occurs within photoreceptor outer segments and is accelerated in Stargardt disease [99]. This, coupled with the photoreceptor bisretinoid accumulation serving as a source of SW-FAF signal [99] makes degenerated photoreceptors as the source of these hyper-reflective foci less likely. However, as the maximum lateral resolution of the fundus autofluorescence images is approximately 10 µm/pixel, it is possible that the hyper-reflective foci may be too small for a change in autofluorescence signal to be detected.

Microglia are the resident immune cells of the retina and are restricted to the inner retinal layers [100]. Light-induced photoreceptor death in Abca4$^{-/-}$ mice has been shown to cause infiltration of microglia to the outer retina [101]. It is therefore possible to hypothesise that the hyper-reflective foci identified in this



(a) Retinal flecks described by Voigt *et al.* [87]



(b) Hyper-reflective foci described in this study

Figure 2.9: Comparison of the hyper-reflective foci described in this study with Group D retinal flecks on spectral-domain optical coherence tomography (SD-OCT) images. **(a)** Group D retinal flecks (indicated by the black arrows) described by Voigt *et al.* [87] that most resemble the hyper-reflective foci described in this study show an accumulation of hyper-reflective material limited to the outer nuclear layer (adapted from [87]). It is important to note that the SD-OCT images in (a) are inverted, therefore hyper-reflective areas will appear dark. **(b)** Hyper-reflective foci described in this study (indicated by the white arrows) located within the outer nuclear layer. Left panel - participant S06, middle panel - participant S10, right panel - participant S12. Note the different morphology of the hyper-reflective foci in **(b)** compared to the retinal flecks in **(a)**.

study may represent photoreceptor mediated activation of immune (microglial) cells as a consequence of neurosensory retinal degeneration.

2.5.2 Leading edge of disease

Examination of the leading disease edge and surrounding retina has provided insights into the sequence of events leading to atrophy within the macula. Thinning of the outer nuclear layer, detected at spatial locations furthest from the leading disease edge (i.e. towards normal retina but not retinal atrophy; Figure 2.7A) and therefore likely to represent some of the earliest degenerative changes, is probably due to loss of significant numbers of photoreceptor cell nuclei. At spatial locations closer to the leading disease edge (Figure 2.7B), in addition to the outer nuclear layer thinning, qualitative reductions in reflectivity of the external limiting membrane and ellipsoid zone were observed suggesting subtle outer segment pathologic features. At the leading disease edge (closest to the area of active disease), collapse of the outer nuclear layer and loss of the external limiting membrane and ellipsoid zone were observed (Figure 2.7C). Although these features were identified in areas of abnormal autofluorescence signal, the absence of increased choroidal transmission of the SD-OCT signal, a characteristic feature of RPE atrophy [102], would suggest that the RPE remains intact. Identification of these changes on confocal AOSLO imaging was not possible due to the ambiguous nature of cone photoreceptors within the transition zone (see Chapter 6). Our findings are consistent with previously reported studies investigating early patterns of macular degeneration in Stargardt disease [103], which suggested that outer nuclear layer thickness determined using SD-OCT may be the preferred biomarker for monitoring early patterns of degeneration in patients with Stargardt disease.

2.5.3 Clinical retinal imaging

The observed autofluorescence abnormalities were in agreement with previously published reports. SW-FAF and NI-FAF were reduced within the central macula, however, NI-FAF demonstrated a greater central area of reduced autofluorescence

signal compared to SW-FAF [70, 104–106]. In this cohort of patients, autofluorescence abnormalities appeared earlier and were more widespread with NI-FAF compared to SW-FAF. This is in keeping with prior studies that have reported NI-FAF imaging enables visualisation of degenerative changes that precede SW-FAF abnormalities [104, 107]. The increased sensitivity of NI-FAF in addition to potential safety concerns over blue light exposure in patients with Stargardt disease [108, 109] suggests that NI-FAF may be the imaging modality of choice for monitoring these patients in routine clinical practice.

The SD-OCT and fundus autofluorescence data presented here support the hypothesis that cone photoreceptor degeneration is a primary phenomenon and does not only occur secondary to RPE atrophy [103, 110–113]. Greater accumulation of the toxic bisretinoid diretinoid-pyridinium-ethanolamine (A2E) occurs in the ABCA4-deficient cone-dominant mouse model compared to the ABCA4-deficient rod-dominant mouse model [110]. Not only do ABCA4-deficient cones generate more A2E than rods, but the generated A2E may remain associated with the photoreceptor for longer allowing it to exert its toxic effects [110].

2.5.4 Macular microperimetry

Macular microperimetry provides a better assessment of visual function than best corrected visual acuity testing alone [114]. Correlating functional information from macular microperimetry with structural information from multimodal conventional clinical retinal imaging demonstrated that the PRL was located in areas with an intact RPE/Bruch's membrane complex. However, in most participants with Stargardt disease, the PRL was associated with abnormalities in autofluorescence signal intensity.

2.5.5 Limitations

Limitations of this study include the small sample size and cross-sectional nature of the data. Participants with any conditions affecting image acquisition quality were excluded (see Chapter 3) and therefore this cohort will not be fully representative

of a condition that is both clinically and genetically highly heterogeneous. In the future, larger studies will be required to validate these findings.

3

Participant selection and image acquisition with the AOSLO

Contents

3.1 Contributions 55
3.2 Background 56
3.3 Participant selection and image acquisition 56
3.3.1 Fixation stability 56
3.3.2 Fundus-guided retinal imaging 57
3.3.3 Refractive error 58
3.3.4 Ocular media opacity 59
3.3.5 Assessment of retinal structure 59
3.3.6 Ocular surface 60
3.3.7 Maintaining active individual participation 60
3.3.8 Avoiding dark adaptation 61
3.3.9 Variations in local retinal topography 62
3.4 Recommendations for future participant recruitment . 62
3.4.1 Participant selection criteria 62
3.4.2 AOSLO image acquisition protocol 63

3.1 Contributions

Through the experience of imaging participants with Stargardt disease and healthy controls I gained expertise in optimising AOSLO image quality at the acquisition stage and developed the participant selection criteria and image acquisition protocols

described below. The graphical user interface described in Section 3.3.2 was developed by Laura Young.

3.2 Background

Despite the benefits, there are a number of challenges associated with *in vivo* AOSLO imaging, particularly in patients with retinal disease. To mitigate these challenges and optimise AOSLO image quality in patients with retinal disease, specific participant selection criteria and image acquisition protocols are required.

The study described in Chapter 2 represents some of our first experiences of AOSLO imaging in participants with retinal disease. The selection of participants for inclusion into the study required participant selection criteria (described in Section 2.3.1) to be established *a priori*. Substantial learning occurred throughout the study with the development of experience and expertise in AOSLO imaging of participants with retinal disease.

Synthesising the learning that has taken place over the course of this study has enabled recommendations for the future recruitment of participants to optimise AOSLO image quality in patients with retinal disease. This chapter aims to describe the recommendations for future participant selection criteria and AOSLO image acquisition protocols that were learned and developed through experience, to optimise AOSLO image quality in patients with retinal disease.

3.3 Participant selection and image acquisition

3.3.1 Fixation stability

Early experiences of imaging a limited number of participants with Stargardt disease to assess the imaging capability of the AOSLO did not yield useful images due to poor participant fixation stability and the selection of participants with substantial central chorioretinal atrophy. Furthermore, there was no way of reliably identifying which retinal region was being imaged. Following these early experiences macular microperimetry was introduced to provide information on participants' fixation

stability. Fixation stability is important during AOSLO imaging due to the raster scanning nature of the AOSLO, and eye movements cause characteristic artefacts within AOSLO images. Due to the small field of view, eye movements corresponding to a large fraction of the size of an individual frame will impair subsequent image processing steps to improve image signal to noise ratio.

The participant task during AOSLO imaging, although similar to macular microperimetry, only requires participants to fixate on a stationary target for short periods of up to 3 seconds. A typical macular microperimetry examination consisting of 68 points covering the central 20° of the visual field and using a 4-2 projection strategy would take a healthy subject approximately 7-10 minutes to complete. Fixation stability measured during macular microperimetry is calculated over the duration of an examination and this measurement is therefore subject to the effect of participant fatigue. Fixation stability was classified by determining the percentage of fixation points located within either a 2° or 4° circle centred on the gravitational centre of all fixation points measured during a single exam. If greater than 75% of all fixation points were within 2°, fixation was classified as stable. If less than 75% of all fixation points were within 2° but greater than 75% were within 4°, fixation was classified as relatively unstable. If less than 75% of all fixation points were within 4°, fixation was classified as unstable.

Fixation stability measured during macular microperimetry is a useful screening tool to identify suitable participants with retinal disease for AOSLO imaging. Further studies with larger participant numbers will be required to accurately determine a minimum threshold for fixation stability that will yield useful AOSLO images from participants with retinal disease.

3.3.2 Fundus-guided retinal imaging

The addition of macular microperimetry to provide information on participants' fixation stability represented a significant step forward in selecting suitable participants for AOSLO imaging. In order to accurately determine the spatial localisation of the AOSLO scan patch and target retinal regions of interest a graphical user interface

was used that accounted for the preferred retinal locus (PRL) position determined from macular microperimetry with an e*n face* fundus image obtained during the clinical assessment. The position of the scan patch within the AOSLO is fixed so retinal areas are targeted by altering the participant's gaze relative to the scan patch, which is achieved by moving the fixation target. Accurate identification of the PRL position on an *en face* fundus image within the interface allowed fundus-guided movement of the fixation target to select retinal regions of interest.

In participants with an eccentric PRL, light passes through the eye lens off-axis, inducing optical aberrations, which could be too large to correct adequately with the deformable mirror. When imaging the PRL and surrounding retina the further the PRL position is from the foveal centre the greater this effect will be [115, 116]. Identifying participants with a PRL sufficiently close to the foveal centre will minimise the adverse effect of this phenomenon on image quality.

3.3.3 Refractive error

The stroke of a deformable mirror represents the possible range of displacement of the mirror surface, which is directly proportional to the amount of aberration it is able to correct. Optical aberrations of the human eye can broadly be classified into low order (defocus and astigmatism) and high order. The stroke of the deformable mirror used in the Oxford AOSLO is 3.5 µm, which enables the correction of approximately 1 dioptre of defocus. To preserve sufficient stroke for the deformable mirror to correct high order aberrations, low order aberrations were corrected using a participant's prescription glasses, contact lenses or trial lenses. However, lenses used for the correction of low order aberrations absorb light and can introduce scattering and back reflections, degrading image quality. Trial lenses of greater refractive power, which are required to correct larger refractive errors, typically have a greater adverse effect on image quality. Therefore, in order to adequately correct ocular aberrations, participants with significant refractive error were excluded due to the need for trial lenses and their effect on image quality.

The presence of astigmatism has a disproportionate effect on image quality compared to defocus. This is because the stroke of a deformable mirror can only correct approximately half the amount of astigmatism compared to defocus, which for the Oxford AOSLO would equate to the correction of approximately 0.5 dioptres of astigmatism. When correcting a participant's refractive error with trial lenses in the Oxford AOSLO, their refraction can be used as an initial estimate of defocus and astigmatism, however, the final refractive correction required should be guided by image quality.

3.3.4 Ocular media opacity

Ocular media opacities cause scattering and absorption of light and will degrade the quality of AOSLO images. Even ocular media opacity that is not clinically significant (such as early cataract) may cause sufficient light scattering and absorption to adversely affect AOSLO image quality. Opacification of the eye lens (cataract) is a normal age-related process, and although it can be associated with some retinal diseases it is not typically associated with Stargardt disease. Participants with clear ocular media, which includes those who have undergone cataract surgery with placement of an artificial intraocular lens, are more likely to yield useful AOSLO images.

3.3.5 Assessment of retinal structure

Confocal AOSLO requires waveguiding photoreceptors in order to visualise photoreceptor outer-segment structures. Spectral-domain optical coherence tomography (SD-OCT) allows the presence of photoreceptors to be determined in areas of significant retinal degeneration through visualisation of the ellipsoid zone (which corresponds to the photoreceptor layer). However, SD-OCT has limited lateral resolution that precludes identification of individual photoreceptor outer-segment structures. SD-OCT can therefore be used to identify the presence and integrity of the ellipsoid zone to determine the likelihood of visualising photoreceptor outer-segment structures with confocal AOSLO imaging. By determining the PRL position

and using this to guide the AOSLO imaging location with respect to the fundus image (see Section 3.3.2), the severity of degeneration within the surrounding retina can be ascertained.

3.3.6 Ocular surface

The tear film is crucial for supporting the optical integrity of the eye by maintaining a smooth refracting surface on the cornea. During an AOSLO imaging session non-continuous viewing of a fixation target on a visual display unit is required and a session can take up to one or two hours. Visual display unit use is associated with a reduced blink rate, reducing the quality of the tear film [117], and retinal image quality has been shown to decrease with tear film instabilities due to the inducement of optical aberrations [118, 119]. Lubricating eye drops prevent drying of the ocular surface and the associated decline in image quality. Regular and early application of lubricating eye drops guided by image quality rather than the subjective symptoms of dry eyes is important as pre-symptomatic dry eyes can have a significant effect on image quality. Lubricating eye drops with a high viscosity would take too long to disperse across the ocular surface causing blurred vision and adversely affecting AOSLO image quality. Therefore, lubricating eye drops with a viscosity similar to water are preferred.

3.3.7 Maintaining active individual participation

Acquiring good quality AOSLO images requires active subject participation. If a participant is uncomfortable and not able to concentrate on the fixation target, image quality will be compromised. Ensuring that participants are seated comfortably with a straight back and neck, have an appropriately positioned bite bar and dental impression (to maintain a stable head position), and maintain engagement and concentration throughout the imaging session is important. This is especially true for participants with reduced visual acuity who may require higher levels of concentration in order to maintain stable fixation. The size of the fixation target can affect accommodation and fixation stability. Accommodation, the process of

refocussing light to enable the eye to focus on objects at different distances, is accomplished by controlled changes in the geometry (shape and thickness) of the eye lens. The ability of the eye to accommodate decreases with age (presbyopia) while in young adults it contributes up to 15 dioptres of refraction [22]. Accommodation of the eye lens controls accurate focusing in response to blur cues. Poorer visual acuity may be associated with a less accurate accommodative response [120]. Pharmacological cycloplegia (the use of eye drops to paralyse the muscles of the eye that control accommodation) is not routinely used during imaging with the Oxford AOSLO. However, if accommodation fluctuates too rapidly it will introduce large and rapidly changing wavefront aberrations that the AOSLO system is unable to compensate for, impairing the quality of AOSLO images. Participants with poor vision will struggle to see a small fixation target and it is therefore important to select an appropriately-sized fixation target relative to the participant's visual acuity.

3.3.8 Avoiding dark adaptation

The resolution of the AOSLO is proportional to $\frac{\lambda}{D}$, where λ is the wavelength of the imaging light and D is the pupil diameter. As participants are not routinely dilated, bright ambient illumination will cause excessive pupil constriction, reducing resolution. Pupil constriction will also reduce the signal intensity by reducing the amount of light passing through the pupil. For these reasons, imaging was performed in a dim environment. However, changes in cone reflectance can occur for a number of reasons, one of which is photopigment dependent light absorption [121]. During dark adaptation cone photopigment density increases as photopigments regenerate, reducing cone reflectively and the signal returning from cone photoreceptors. It was observed that AOSLO image quality reduced over the course of a continuous imaging session. It is possible that a reduction in cone reflectivity caused by dark adaptation may have led to the observed reduction in image quality. However, it may also be due to a break up in the tear film caused by reduced blinking during a continuous imaging session. Dark adaptation can be prevented by encouraging participants to take regular breaks in a well lit environment. Regular breaks also help to prevent

excessive drying of the ocular surface by encouraging normal blinking, and help maintain participant engagement and concentration during the imaging session.

3.3.9 Variations in local retinal topography

Patients with Stargardt disease can develop retinal flecks — macroscopic lipofuscin containing lesions that originate from the RPE [79, 95–98]. These flecks cause local elevations of the overlying retinal structures, which can cause the photoreceptor mosaic to move away from the focal plane set by the adaptive optics system. By adding a small offset to the deformable mirror, changes due to the local topography of the retina can be overcome and the focal plane corrected.

3.4 Recommendations for future participant recruitment

Participant factors make a substantial contribution to the quality of AOSLO images and represent a number of the challenges associated with *in vivo* AOSLO imaging, particularly in patients with retinal disease. Appropriately selected participant selection criteria and image acquisition protocols may help to mitigate these challenges and optimise AOSLO image quality in patients with retinal disease. The participant selection criteria and AOSLO image acquisition protocol outlined below have been developed through the experience and expertise gained in AOSLO imaging of participants with retinal disease. They relate to the AOSLO system and setup currently in use and may change as the system continues to be developed.

3.4.1 Participant selection criteria

Recommended participant selection criteria for future studies include:

1. *Absence of significant refractive error.* Exclude participants with a spherical equivalent greater than 4 dioptres or astigmatism greater than 2 dioptres. The spherical equivalent is a useful measure to provide an estimate of the eye's overall refractive error. It considers defocus and astigmatism, and is calculated by adding the sphere power with half of the cylinder power.

2. *Absence of ocular media opacity.* Exclude participants with evidence of ocular media opacities determined on clinical examination, even at levels that may not be clinically significant.

3. *Fixation stability.* Exclude participants without stable fixation, determined using macular microperimetry (Section 3.3.1). If greater than 75% of all fixation points were within 2° of the gravitational centre of all fixation points measured during a single microperimetry exam, fixation was classified as stable.

4. *PRL within 10° of the foveal centre.*

5. *The presence of sufficient retinal structure.* Exclude participants with an absent or substantially disrupted ellipsoid zone on SD-OCT within the retinal region(s) of interest.

3.4.2 AOSLO image acquisition protocol

Recommended image acquisition procedures for future studies include:

1. *Ensure the participant is seated comfortably with a straight back and neck.*

2. *Correct defocus and astigmatism if necessary.*

3. *Use lubricating eye drops prophylactically to prevent drying of the corneal surface.*

4. *Match the size of the fixation target to the participant's visual acuity.*

5. *Avoid dark adaptation.* Encourage participants to take regular breaks in a well lit environment.

6. *Maintain active participation during the imaging session to improve fixation stability.*

7. *Fundus-guided imaging.* Use the fundus image to select retinal regions of interest and target them by moving the fixation cross.

3. Participant selection and image acquisition with the AOSLO

8. *Account for variations in the local retinal topography.* Change the plane of focus to account for variations in the local retinal topography by using a focus offset on the deformable mirror.

4

Automated reference frame selection

Contents

4.1 Contributions 65
4.2 Background 66
4.3 Methods 67
4.3.1 Total intensity 69
4.3.2 Interframe grouping 71
4.3.3 Hierarchical clustering 75
4.3.4 Image quality metrics 76
4.3.5 Synthetic images 80
4.3.6 Synthetic datasets 82
4.3.7 Estimating residual aberration 83
4.3.8 Estimating signal to noise ratio 86
4.3.9 Estimating eye motion 86
4.3.10 Reference frame selection 93
4.4 Results 95
4.4.1 Algorithm evaluation 95
4.5 Discussion 99

4.1 Contributions

My contributions to this work include: study concept and design; algorithm development and coding; collection, analysis and interpretation of data. I collected and obtained ethical approvals for the AOSLO images from participants with Stargardt disease and healthy controls used to inform development of the algorithm

(see Section 2.3). The simulation code described in Section 4.3.5 was developed and written by Laura Young. I had full access to all of the data presented in this chapter and take responsibility for the integrity of the data and the accuracy of the data analysis.

4.2 Background

Eye movements are fundamental to seeing and even during fixation of a stationary target the eye is moving. These eye movements, which occur on a time scale similar to the scanning rate, lead to distortion of images from the adaptive optics scanning laser ophthalmoscope (AOSLO) and are especially pronounced because of the small field of view required to achieve the sampling rates for cellular imaging. Correction of these distortions is imperative to achieve highly accurate visualisations of retinal structures.

Due to the scanning nature of image capture with an AOSLO and the presence of fixational eye movements, it is not possible to obtain a distortion free representation of the retina with *in vivo* imaging. Furthermore, averaging a spatially-aligned and eye-motion-corrected sequence of images is required to improve the signal to noise ratio as the signal of a single image is restricted by light safety limits. Reference frames are required to measure and correct intraframe eye motion in order to produce averaged images with high signal to noise ratios. The perfect reference frame would contain no residual aberration or eye movement related distortion. Reference frame selection, the first step in the image processing pipeline following image capture (Figure 1.13), has traditionally been performed manually by an experienced image grader. However, this is both time consuming and subjective. Following the acquisition of AOSLO images, reference frame selection represents the first opportunity to minimise any errors that may propagate along the image processing pipeline.

To date, one algorithm has been described for semi-automated selection of reference frames from AOSLO image sequences of the cone photoreceptor mosaic [122]. However, user input is still required to verify that the selected reference

frame is an accurate representation of the participant's photoreceptor mosaic and does not contain significant distortions. Approaches that utilise no-reference image quality metrics such as sharpness [123–126], contrast [123, 125, 127] or entropy [125, 128] to determine image quality have been described for adaptive optics (AO) telescopes [124], scanning laser ophthalmoscopes [123, 127] and flood illumination systems [125, 126]. However, these image metrics are unlikely to be suitable for identifying the intraframe eye motion artefacts inherent to AOSLO images due to the interaction of these artefacts with image noise and residual aberration. The default image processing pipeline for the Oxford AOSLO selects reference frames automatically by identifying the frame from an image sequence that is most similar to all other frames within that sequence of images. This method assumes that eye motion is random and averages to zero, and therefore that the most similar frame to all others will contain the least motion.

Due to the significant volume of data generated from each AOSLO imaging session, processing and analysis of the data has become a rate limiting step. Current image processing pipelines require significant manual input for the processing and analysis of these images. In order for the clinical use of AOSLO imaging to become widespread, a robust and reliable fully automated image processing pipeline is required that addresses the unique challenges associated with high-resolution *in vivo* imaging of patients with retinal disease. This chapter describes the development and validation of an unsupervised automated reference frame selection algorithm that considers the unique challenges associated with processing AOSLO images from patients with retinal disease.

4.3 Methods

The algorithm was developed and implemented in the Python programming language (Python Software Foundation. Python Language Reference, version 2.7) and is illustrated in Figure 4.1 and described below.

Briefly, following the acquisition of an AOSLO image sequence, bad frames, such as those captured during a blink, are rejected based on their total intensity

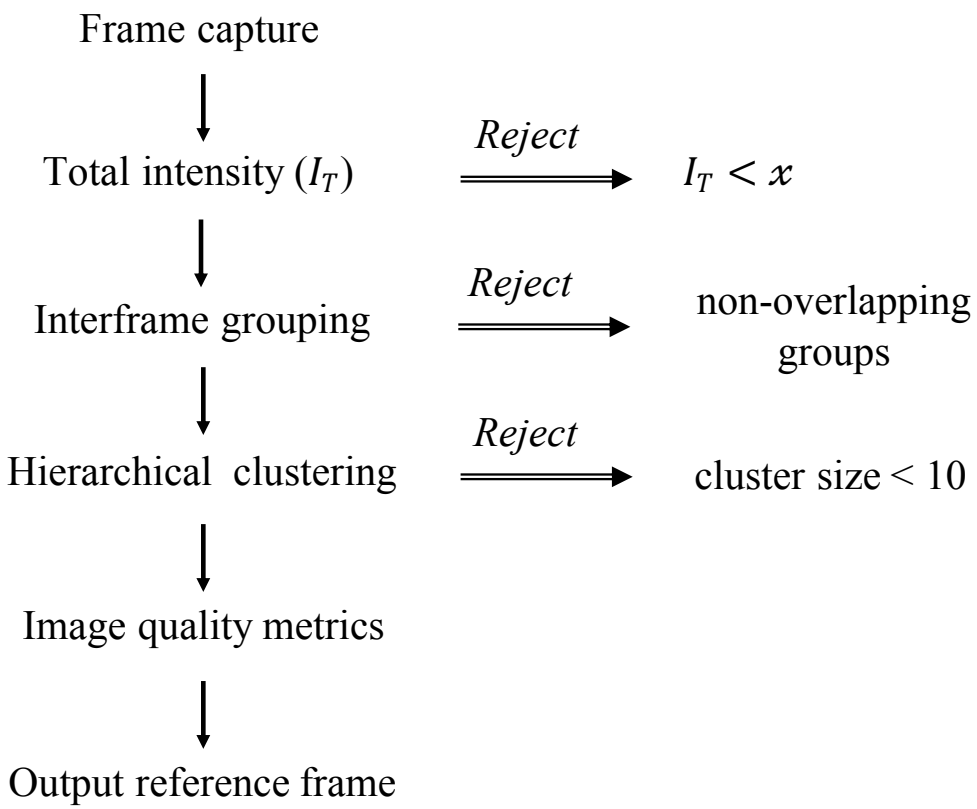


Figure 4.1: Automated reference frame selection algorithm workflow. Following acquisition of an AOSLO image sequence, frames were rejected based on total intensity ($I_T < x$), where x is the minimum intensity threshold (Section 4.3.1). Next, all frames were grouped based on their retinal spatial location and non-overlapping groups were rejected. Groups were then clustered using hierarchical clustering and image quality metrics calculated for frames in each cluster (cluster size $\geq$ 10). The best frame in each cluster (as determined by the quality metrics) was output as the reference frame.

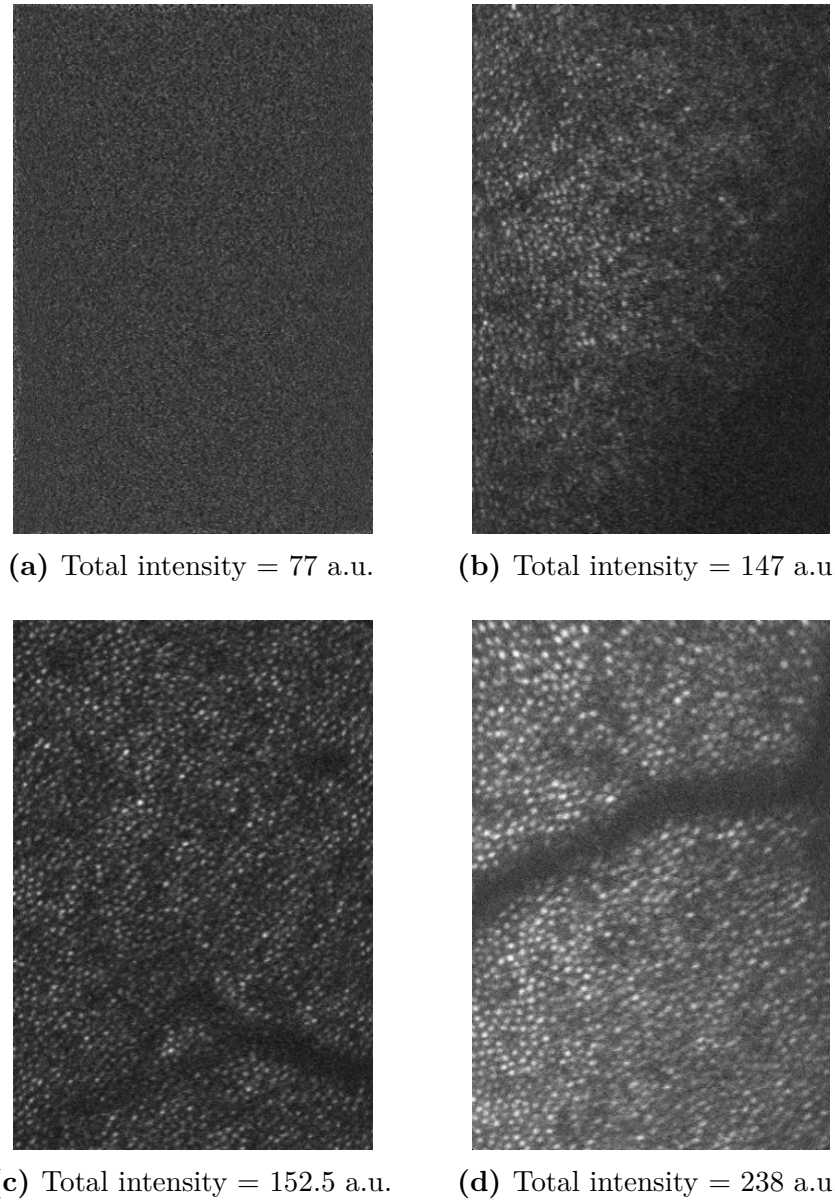
(Section 4.3.1). Next, all frames are grouped based on their spatial location on the retina (Section 4.3.2). Groups of frames with no spatial overlap to other groups (non-overlapping groups) are rejected at this stage. The remaining groups are then clustered using hierarchical clustering (Section 4.3.3) to produce clusters of frames with a minimum amount of overlap based on their spatial location on the retina. Following hierarchical clustering, all clusters containing less than 10 frames are rejected. The penultimate step in the algorithm is to determine the image quality of all frames within a cluster, one cluster at a time (Section 4.3.4). The image quality metrics are used to determine the best frame within each cluster, which is output as the reference frame for that cluster (Section 4.3.10). The automated

reference frame selection algorithm was evaluated using synthetic datasets with known ground-truths (Section 4.4.1).

4.3.1 Total intensity

Frames with reduced image quality (intensity), such as those captured during a blink, require removal from the image processing pipeline. Image intensity also changes with retinal eccentricity, photoreceptor reflectivity and image acquisition settings. Current methods to identify these bad frames utilise a relative intensity threshold based on the mean and standard deviation of image intensity from an image sequence captured at one retinal location [122]. However, this method is sensitive to outlier intensity values due to their effect on the mean and standard deviation. Images from participants with retinal disease exhibit greater variability in image intensity compared to images from control participants and this may be due to differences in retinal structure (image intensity will vary between areas of normal, mildly and severely degenerated retina) and the ocular surface (imaging sessions can take longer in participants with retinal disease leading to a breakdown of the tear film). Due to the greater variability of image intensity with patient imaging, a method that utilises a relative threshold of average image intensity may fail to reject all bad frames without adjusting the threshold for each patient. An absolute threshold of total image intensity will overcome these limitations of a relative threshold.

During a blink, the light reflected to the detector is substantially reduced, and is essentially due to noise inherent in the system. As images will have been captured at the same point in time (at a frame rate of 30 Hz, it takes 3 seconds to capture 90 frames) and will have imaged the same retinal location, the noise present in each image can be considered the same. The differences between the total intensity of two frames where one is captured during a blink will therefore be related to the signal corresponding to the retinal features (Figure 4.2). The following formula was used to provide a non-linear measure of total image intensity (I_T), where i corresponds to the i^{th} pixel in the image:



(a) Total intensity = 77 a.u. **(b)** Total intensity = 147 a.u.

(c) Total intensity = 152.5 a.u. **(d)** Total intensity = 238 a.u.

Figure 4.2: The total intensity of four raw frames. Frames with a total intensity less than 153 a.u. were rejected. Frame **(a)** was captured during a blink. Frames **(b)** and **(c)** illustrate examples of images containing retinal structure but with a total intensity just below threshold. Frame **(d)** illustrates a good image containing clear structure with a total intensity above threshold.

$$\mathrm{I}_T = \sqrt{\sum_{i=1}^{N} \mathrm{I}_i^2} \tag{4.1}$$

An absolute minimum total intensity threshold, to reject images based on their intensity, was manually selected. The threshold was determined by calculating the total intensity of 277,992 frames from 16 participants with Stargardt disease and 7

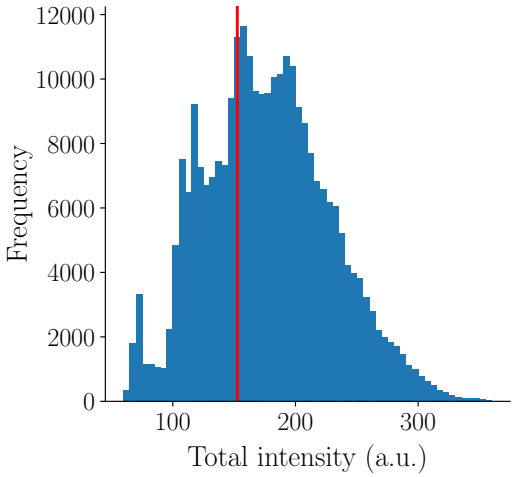


Figure 4.3: Histogram of the total intensity (calculated using Equation 4.1) from 277,992 frames of 16 participants with Stargardt disease and 7 normal controls. An absolute minimum total intensity threshold of 153 a.u. (represented by the vertical red line) was manually selected.

normal controls (using Equation 4.1) and manually selecting a suitable threshold based on image quality (Figure 4.3). Frames used to select the absolute threshold were not pre-filtered and were from a range of retinal eccentricities. A conservative lower intensity threshold was selected in order to reject poor quality frames and those containing partial blinks. Although this threshold will reject some frames containing sufficient retinal structure to contribute information to a reference frame, the adverse effect of a few poor quality frames on the averaged image was felt to be more significant as it will impede the subsequent step of image analysis. Furthermore, due to the relatively high rate of data-capture, a short imaging session provides sufficient high intensity frames to justify a conservative lower intensity threshold. An upper intensity threshold was not selected as images of highly reflective retinal areas were not routinely captured from this participant cohort.

4.3.2 Interframe grouping

Fixational eye movements are important for normal vision but have a substantial effect on the spatial relationship of sequentially captured frames. Identifying a single reference frame for all images collected at a given retinal location is problematic because some images will not overlap with the chosen reference frame,

impairing the subsequent step of eye motion correction. Clustering frames by their relative retinal location allows a reference frame for each cluster to be identified and therefore substantially greater overlap of all frames within a cluster to their chosen reference frame.

To identify the relative retinal location of each frame, image based registration using normalised cross correlation was carried out with a custom peak detection algorithm. Briefly, the first step grouped consecutive, overlapping frames; the second step connected overlapping groups of frames; and the final step selected a reference group, used to determine the spatial location of all frames on the retina.

4.3.2.1 Peak detection

The normalised cross correlation of two overlapping frames will contain a distinct peak, whose spread, shape and height encodes the similarity between two frames. The linear translation providing maximal signal overlap between these frames is indicated by the position of the global maximum in the normalised cross correlation matrix. The absolute value of the maximum (the peak's height) is related to the proportion of overlap and inter-frame distortions, while the spread and shape of the peak is related to the spatial frequency content of the two frames. The presence and magnitude of inter-frame distortions will vary between patients due variations in fixation stability, which will affect the value of the normalised cross correlation maximum. Using an absolute threshold for the maximum will therefore require adjustment of the threshold for each patient.

A custom peak detection algorithm was implemented that used the spread (2D spatial standard deviation) of the normalised cross correlation peak to determine if two frames overlapped. An autocorrelation of the source image will account for variations in the spatial frequency content of frames and was used to provide an estimate of the expected spatial extent of the peak. A threshold for the maximum acceptable dimensions of the peak was determined by using the standard deviation obtained by fitting a 2D Gaussian to the peak from an autocorrelation of the source image. Maximum projection was used across four axes (vertical, horizontal and the

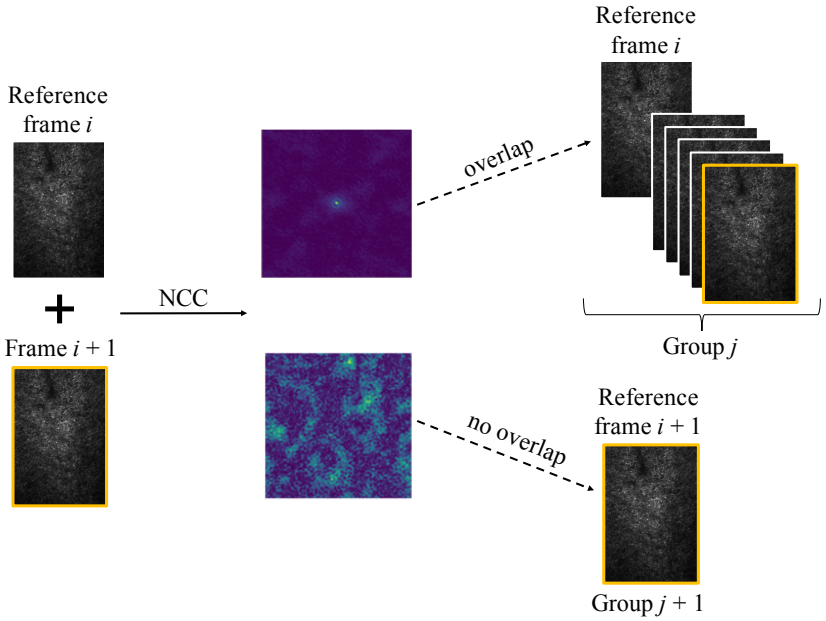
diagonals) to estimate the full-width half-maximum of a peak from the normalised cross correlation of two frames. If the full-width half-maximum was less than the threshold (determined by fitting a 2D Gaussian as described above), the position of the peak was used to generate coordinates of the frame's position relative to the source image, as it implies a high certainty on the position offset. If the full-width half-maximum was greater than the threshold, no connection between the frames was made as it implies insufficient overlap.

4.3.2.2 Step 1: grouping consecutive frames

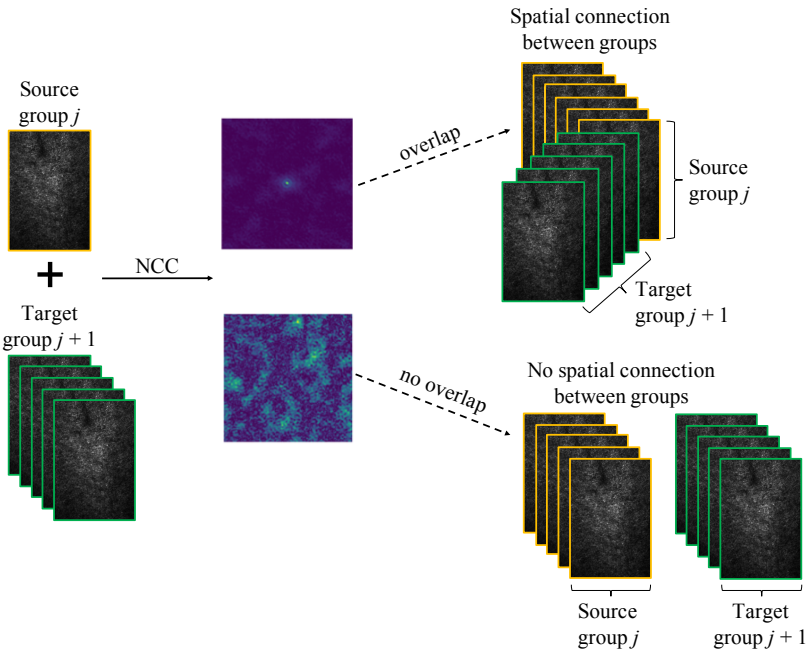
The first step in identifying the relative retinal location of each frame utilised the custom peak detection algorithm to group overlapping, consecutive frames (Figure 4.4a). Frames were grouped if a peak was successfully found, and a new group started if no peak could be identified. Groups could range in size from 1 to the total number of frames. This allows single frames with no overlapping consecutive frames to be connected to other frames within the imaging sequence, and is of particular relevance to imaging patients with retinal disease where due to retinal pathology, fixation stability may be reduced leading to a greater frequency of larger eye movements. In this situation connecting frames across the imaging sequence becomes more important.

4.3.2.3 Step 2: connecting overlapping groups of frames

The second step used the same peak detection algorithm to determine a spatial connection between the groups created in step one and identify all overlapping groups (Figure 4.4b). Groups were connected by taking the first frame in the source group and sequentially performing a normalised cross correlation with each frame in the target group. If a peak was detected (overlap), the source and target groups could be spatially connected and the position of the peak was used to generate coordinates of the target group's position relative to the source group. In order to maximise the number of groups that are connected, if no peak was detected after the first frame in the source group had been compared with every frame in the target group, then each subsequent frame in the source group would be



(a) Step 1 - create groups of overlapping consecutive frames



(b) Step 2 - identify all overlapping groups

Figure 4.4: Two-step imaged-based registration method used to detect inter-frame position offsets and identify the relative spatial location of frames. **(a)** Step 1 - uses normalised cross correlation (NCC) to create groups of overlapping consecutive frames. If the next frame in the sequence (frame $i+1$) overlaps with the reference frame (frame i) of the group (group j), it is added to group j. However, if the frames do not overlap, a new group (group $j+1$) is created. **(b)** Step 2 - uses NCC to identify all overlapping groups. If a target group (group $j+1$) does not overlap with the source group (group j), it is considered to have no spatial connection to the source group.

used in turn as the source image for a normalised cross correlation. If no peak was detected (no overlap), then the target group was not considered to have any spatial connections to the source group.

4.3.2.4 Step 3: selecting the reference group

The final step identifies a reference group from which all other groups will derive their relative retinal locations. The group with the greatest number of spatial connections to other groups was selected as the reference group and retinal coordinates were generated for all frames relative to the first frame in the reference group. In order to maximise the number of frames available for hierarchical clustering, any group that could be spatially connected to the reference group by up to one intermediary group was included.

4.3.3 Hierarchical clustering

Eye motion correction using a reference frame works better the greater the overlap between the frames being corrected and the selected reference frame. Our experience shows that there is a reduction in the quality of the corrected frames if the amount of overlap falls below 75% and that the overlap is direction dependent. Vertical overlap is more important than horizontal overlap as all pixels along a single row (in the direction of the fast scanner) should move together. Vertical non-overlap therefore means missing data for eye motion correction. Clustering frames by their retinal location allows the spatial overlap of all frames in a cluster to be determined. Identifying a reference frame for each cluster ensures substantially greater overlap of all frames within a cluster to their chosen reference frame, improving the subsequent step of eye motion correction.

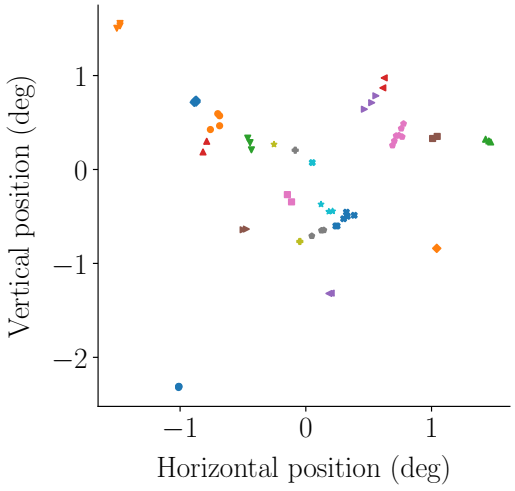
The previous step of interframe grouping (Section 4.3.2) provides a map of each overlapping frame's retinal location. In the next step of the processing pipeline, the retinal locations were clustered using hierarchical agglomerative clustering (Figure 4.5). Hierarchical agglomerative clustering uses a 'bottom up' approach whereby each observation is assigned its own cluster and then based on a similarity

metric, which in this instance was the average distance between each frame within a cluster to every other frame within the same cluster, the most similar clusters are iteratively merged. The 'average' linkage method was used for hierarchical agglomerative clustering. In this method, the distance between two clusters (a linkage) is defined as the average distance of each element in one cluster with every element in the other cluster. Clusters that are greater than a user defined average distance threshold are not merged.

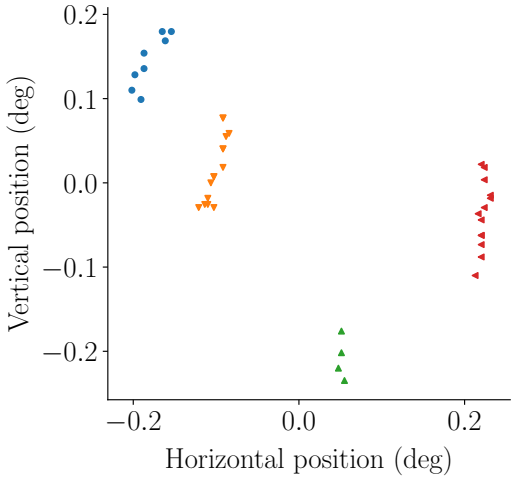
This approach was used rather than a 'top down' approach (e.g. K-Means), where the number of clusters is fixed, in order to allow the data to inform the number of clusters to be selected. As the retinal region captured while imaging at a single retinal location varies based on the fixation stability of the participant, it is not possible to know the optimal cluster number *a priori*. If too many clusters are selected using a 'top down' clustering method, additional merging of clusters will be required. Due to the importance of sufficient overlap between the selected reference frame and the frames to be eye-motion corrected, a distance metric (specifying 75% overlap as discussed above) was used to limit the average distance of every frame within a cluster to every other frame within the same cluster.

4.3.4 Image quality metrics

Artefacts that are inherent in images captured using an AOSLO can broadly be categorised into three groups: residual aberration, eye motion and noise. Since a ground truth of the retinal area imaged is not available when performing *in vivo* imaging, the residual aberration, eye motion and noise present in these images is not known. Therefore, in order to determine which frame within each cluster should be chosen as the reference frame, the relative presence of these artefacts needs to be quantified. Four image quality metrics were considered for this purpose: RMS contrast, sharpness, entropy and the Grey Level Co-occurrence Matrix. The utility of these image quality metrics in identifying the best reference frame was evaluated using synthetic datasets.



(a) Participant S10



(b) Participant S12

Figure 4.5: An example of hierarchical agglomerative clustering results from two participants. Each colour and marker combination represent a different cluster of frames with 75% overlap.

4.3.4.1 RMS contrast

The contrast of an image can be simply explained as the difference between the pixel intensity of two areas of interest. For a greyscale image that contains only two intensity values, this would be the difference between the maximum and minimum pixel intensities. Some common definitions of contrast include Weber contrast, Michelson contrast and RMS contrast, and different definitions are used for different situations. Weber contrast is defined as $C_{Weber} = \frac{L - L_b}{L_b}$, where L and L_b represent luminance of the features and background, respectively. It is

commonly used when small features are present on a large uniform background, where the background luminance is approximately equal to average luminance. Michelson contrast is defined as $C_{Michelson} = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$, where L_{max} and L_{min} represent the maximum and minimum luminance, respectively. It is commonly used for patterns with equivalent bright and dark features, such as gratings, but captures little of the contrast variations across an image as it depends on luminance levels at only two locations. RMS contrast (defined below) is commonly used to estimate the contrast of images and was used by Ramaswamy and Devaney [125] as an image quality metric for the selection of retinal images from a commercial AO flood illumination system. It does not depend on the spatial distribution of contrast or frequency content of the image. The following definition of RMS image contrast, C_{RMS}, was used where σ_I and μ_I represent the standard deviation and mean pixel intensity of the image, respectively:

$$C_{RMS} = \frac{\sigma_I}{\mu_I} \tag{4.2}$$

4.3.4.2 Sharpness

A set of eight image sharpness metrics were originally defined by Muller and Buffington [129] for the real-time correction of astronomical images by an optical element that is used to maximise image sharpness. Fienup and Miller [130] investigated aberration correction in Synthetic Aperture Radar images by maximising generalised sharpness metrics, including those originally defined by Muller and Buffington. Understanding that the second derivative of a nonlinear function (defining a sharpness metric) determines its action on the reconstructed image allowed them to design a sharpness metric to take advantage of the properties of a specific image type and invent 'designer metrics'. One of the simplest of these designer sharpness metrics that enhances the contrast of both bright and dark pixels is described by:

$$S_i = \frac{1}{12}(I_i - \gamma\bar{\mathrm{I}})^{\beta} \tag{4.3}$$

where I represents pixel intensity, $\bar{I}$ represents the average pixel intensity of the image, and $i = 1, 2, ...n$ for n pixels in the image. Average sharpness for the image is derived by calculating the mean sharpness of each pixel (S_i). Ramaswamy and Devaney [125] investigate the use of this designer metric for the selection of retinal images from a commercial AO flood illumination system prior to averaging and determined that for their dataset the optimal parameters for this metric were $\gamma = 1.0$ and $\beta = 4$.

4.3.4.3 Entropy

Entropy can be simply described as a statistical measure of disorder and can be used to indicate the amount of information in a system. Only one bit of information is contained in a perfectly ordered image where pixel intensities are all either 0 or 1. An image with increasingly variable pixel intensities will contain more information and have a higher entropy. The following definition of entropy (E) was used [131]:

$$E = -\sum P_i \log_2(P_i) \tag{4.4}$$

where P_i is the probability distribution of pixel intensities within the image, which is calculated using the grey level histogram.

4.3.4.4 Grey Level Co-occurrence Matrix

A Grey Level Co-occurrence Matrix (GLCM) and its derived statistical measures are commonly used for image texture analysis. The GLCM calculates the frequency of pixel pairs with specific values and a specific spatial relationship within an image. The spatial relationship of pixel pairs to be examined can be altered by changing the distance or angle between a selected pair of pixels. Sampling pixel pairs across eight axes (45° spacing) allows sampling of any underlying textures that may have a specific directional component. The optimal distance between a pair of pixels will be related to the scale of the retinal features and the optical angle to any underlying orientation of the retinal features (e.g. blood vessels). Due to the increasing size and spacing of photoreceptors as a function of eccentricity from the fovea, it will

not be possible to use a fixed pixel pair distance for all retinal images. Additionally, in patients with retinal degenerative changes, the sizes of photoreceptors may exceed the normal variability expected due to retinal eccentricity, and gaps in the photoreceptor mosaic will alter the effective spacing. Therefore the GLCM and its derived statistical measures were not investigated further as an image quality metric.

4.3.5 Synthetic images

The testing and validation of metrics and algorithms developed for the analysis of *in vivo* AOSLO retinal images requires a ground truth. However, when performing *in vivo* retinal imaging, we do not have a ground truth for the retinal structure. An alternative approach is to use synthetic images of specified retinal structures produced via a model of the AOSLO imaging system. The image processing and analysis techniques presented in this thesis have been tested and validated using an end-to-end simulation of the AOSLO data acquisition and image reconstruction process [132], which is briefly described here.

There are three components to the simulation:

1. First, a realistic cone photoreceptor mosaic (the ground truth for cone locations) is created. As the photoreceptor mosaic is not perfectly hexagonal [133, 134], the simulation generates a pseudo-hexagonal matrix of cells using a self-organising algorithm. The spacing in the cell matrix is based on *ex vivo* histologic data from Curcio *et al.* [25].

2. Next, natural eye movements that would be made by a control participant are simulated. A stream of simulated eye position data are produced based on the three main components of fixational eye movements: drift (a low frequency component), microsaccades (small jerk-like movements occurring approximately once per second), and tremor (a high frequency component with a bandpass spectrum of approximately 65 Hz). The parameters used for these eye movements are based on data from the literature.

3. Finally, images are reconstructed based on the reflection of light from the simulated photoreceptor mosaic. The reflectance profile of individual cones is modelled as a Gaussian intensity distribution, the parameters of which are chosen from a random distribution that is selected to represent the natural variation in cone reflectances in human eyes. An individual photoreceptor mosaic is specified as the sum of these Gaussian functions placed at the cone locations. The position on the retina of each pixel in the simulated image is specified as the sum of the scanner position (based on real measurements from the AOSLO) and the eye position at that point in time. Therefore, an image is reconstructed by 'scanning' over the photoreceptor mosaic and the normal desinusoiding procedures followed.

The simulation produces a set of frames of the cone photoreceptor mosaic (rod photoreceptors and blood vessels are not simulated) that are simulated without optical effects considered. The effects of diffraction, which depends on the pupil diameter and wavelength of light, residual aberrations and noise can be added to the image data.

4.3.5.1 Residual aberration

Wavefront aberrations were obtained from a database consisting of a continuous 3.4 second time sequence (without blinks) of 50 healthy eyes, measured with a Shack-Hartmann aberrometer at a temporal frequency of 236 Hz [135]. These measurements were performed in a dark room with natural pupil dilation and no cycloplegia. Wavefront aberrations are quantified by Zernike coefficients. The mean Zernike coefficient was calculated across the time series for each eye and all 65 of the Zernike modes that had been measured (excluding piston). To simulate residual aberration a wavefront composed of a proportion of the mean Zernike coefficients was generated and its point spread function was used to render the synthetic frames via convolution.

4.3.5.2 Signal to noise ratio

Noise introduced at any point in the system was considered as constant in its characteristics over time (see Section 4.3.8). Gaussian distributed white noise (randomly generated for each frame) was added with the same characteristics as the noise measured through the system. By considering noise power as a constant, differences in signal to noise ratio between successive frames will be related to the signal present in each frame, representing the case in real imaging where a number of factors can affect the signal. To vary the signal to noise ratio of frames, the signal, measured in nanowatts at the pinhole, was varied in the simulation. Information about several uncontrolled variables (such as laser power at the cornea, detector gain and temperature, although the detector temperature is controlled) is required in order to accurately derive power at the detector (signal) in nanowatts. In the simulation, these variables were specified based on measurements from the Oxford AOSLO. In order to facilitate the interpretation of this work for readers using a different AOSLO system design, the signal of synthetic images has been specified in both absolute (nanowatts) and relative (signal to noise ratio) terms.

4.3.6 Synthetic datasets

Two separate synthetic datasets were used for evaluating the image quality metrics (Dataset 1) and the reference frame selection algorithm (Dataset 2).

4.3.6.1 Dataset 1: image quality metric evaluation

A total of 90 synthetic frames from a retinal eccentricity of 3°, 480 × 300 pixels in size, containing normal simulated eye movements were generated. To simulate raw frames, the signal to noise ratio was varied and residual aberration was added to each synthetic frame. In order to vary the signal to noise ratio of frames, the signal, measured in nanowatts (nW) at the pinhole, was varied from 0.2 nW to 0.5 nW in steps of 0.1 nW. To simulate residual aberration, fixed percentages (50%, 40%, 30%, 20%, 10% and 0%) of the mean Zernike coefficients from 10 randomly

selected eyes were applied to each of the synthetic frames. This resulted in a total of 21,600 synthetic frames (Figure 4.6).

4.3.6.2 Dataset 2: reference frame selection algorithm evaluation

Ten image sets were used for algorithm evaluation. Normal simulated eye movements (that included between one and three microsaccades) and wavefront aberrations were unique to each image set. Each image set consisted of 90 synthetic images (each 480 × 300 pixels in size) from a retinal eccentricity of 3°, representing an image sequence captured over approximately 3 seconds at one retinal location. Each frame within an image set was randomly assigned wavefront aberrations of between 0% and 10% (in 1% steps) and signal between 0.2 nW and 0.3 nW (in steps of 0.01 nW). Wavefront aberrations were selected from a different healthy eye for each image set.

4.3.7 Estimating residual aberration

The Strehl ratio, a measure of wavefront error, was used to quantify the residual aberration present in each image. Strehl ratio is defined as the ratio of the maximum value in the aberrated point spread function to the equivalent value in a diffraction-limited point spread function. The Strehl ratio ranges between 0 and 1, with a perfect optical system having a Strehl ratio of 1.

The goal was to estimate residual aberration from synthetic images. Image quality metrics were tested using Dataset 1 (described in Section 4.3.6.1), and evaluated against how well the simulated aberration was predicted. Figure 4.7 illustrates the relationship between each image quality metric tested and the Strehl ratio. RMS contrast ($r = 0.948$, $p < 0.001$) and sharpness ($r = 0.535$, $p < 0.001$) are correlated with Strehl ratio, whereas entropy ($r = 0.008$, $p = 0.236$) is not. All image quality metrics tested are sensitive to a frame's signal to noise ratio with higher signal to noise ratios associated with lower RMS contrast, sharpness and entropy values (Figure 4.7).

A least squares multiple linear regression model was used to test if RMS contrast and sharpness predict Strehl ratio (Table 4.1). Based on theoretical considerations,

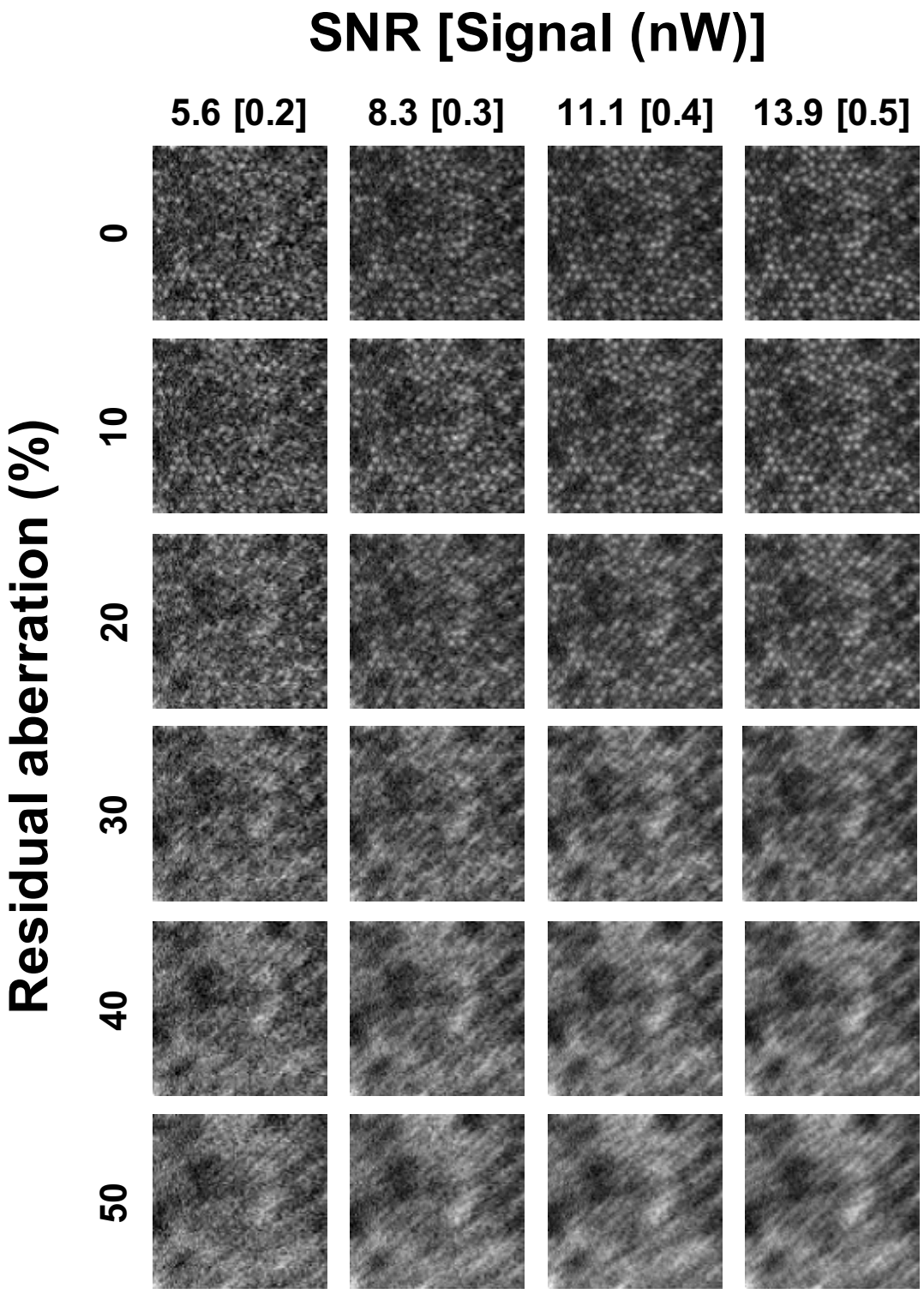
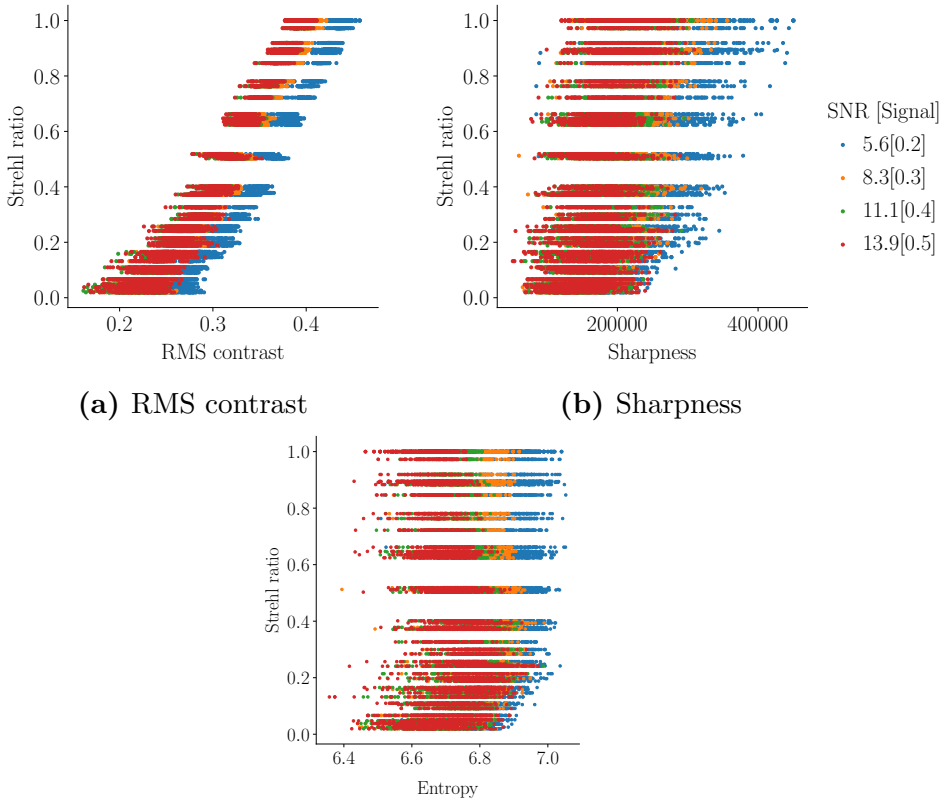


Figure 4.6: An example of the synthetic cone photoreceptor mosaic images with normal simulated eye movements used for evaluating image quality metrics. The signal to noise ratio (SNR) i.e. the image signal (measured in nanowatts (nW) at the pinhole), and the wavefront aberrations added to each image (to simulate residual aberration) have been varied.



(a) RMS contrast **(b)** Sharpness

(c) Entropy

Figure 4.7: Scatter plots illustrating the relationship between each image quality metric tested and the Strehl ratio. Each datum represents a single synthetic frame. Data points are colour coded by the frame's signal to noise ratio (SNR) and the image signal, measured in nanowatts at the pinhole, is given in parentheses. Pearson correlation coefficients (r) and p-values are given for each image quality metric tested. (a) $r = 0.948$, $p < 0.001$. (b) $r = 0.535$, $p < 0.001$. (c) $r = 0.008$, $p = 0.236$.

as the signal to noise ratio of an image decreases the regression slope would be expected to flatten i.e. the model would be expected to poorly predict Strehl ratio from images with a low signal to noise ratio. Using the definitions of RMS contrast (Equation 4.2) and sharpness (Equation 4.3) discussed above, this would occur as an image with a low signal to noise ratio, where image noise substantially outweighs image signal, could increase both metrics. Entropy was excluded from the regression model as it is not correlated with Strehl ratio. The results of the regression analysis indicate that the two predictors explain 91.6% of the variance within the model ($R^2 = 0.916$, $F(2, 21597) = 1.18 \times 10^5$, $p < 0.001$) and RMS contrast is the main predictor.

Table 4.1: Multiple linear regression to predict Strehl ratio.

					95% CI	
Variable	B	$SE\ B$	t	p	**0.025**	**0.975**
RMS contrast	5.63	0.014	402	<0.001	5.60	5.66
Sharpness	$-1.12 \cdot 10^{-6}$	$1.69 \cdot 10^{-8}$	-65.9	<0.001	$-1.15 \cdot 10^{-6}$	$-1.08 \cdot 10^{-6}$

B - unstandardised beta (correlation coefficient), CI - confidence interval, SE - standard error, t - t-test statistic, p - probability.

4.3.8 Estimating signal to noise ratio

As all candidate frames being evaluated for reference frame selection will have been captured in the same 3 second time period and at the same retinal location, the total amount of noise present within each image can be considered as a constant power. It is important to note that although the noise characteristics of each synthetic frame are the same, the noise is randomly generated for each frame. By considering the noise present in each frame as a constant power, the signal present within each frame can be estimated from the frame's total intensity, calculated during the first step of reference frame selection (see Section 4.3.1). The frame containing the most signal is therefore the image with the highest total intensity.

4.3.9 Estimating eye motion

The Fourier transform of a photoreceptor mosaic image contains an annulus (Figure 4.8E), originally described by Yellott [136] and commonly referred to as Yellott's ring, which corresponds to the dominant spatial frequency of the photoreceptor mosaic [137]. The reciprocal of the spatial frequency corresponding to the radius of this annulus is defined as the modal spacing of the photoreceptors from the original image [137]. Due to the raster scanning nature of AOSLO imaging, different parts of the retinal surface are imaged at different times. Eye movements manifest as expansions or compressions in the images when the eye moves parallel to the slow scanning axis and as shear when the eye moves parallel to the fast scanning axis (Figure 4.8A and C). By detecting disruptions to the modal spacing

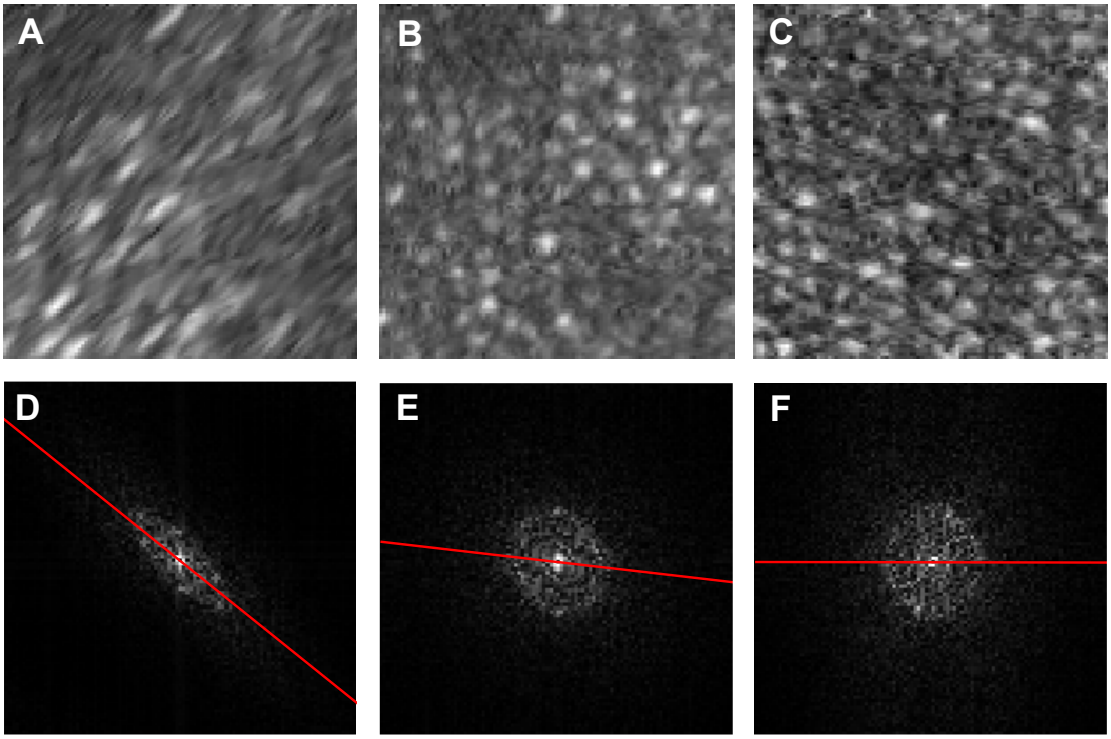


Figure 4.8: Estimating intraframe eye motion. Top row - each image represents an image patch, 100 × 100 pixels in size, from a control participant taken at 5° from fixation. 57° and 10° of shear was applied to **C** in a clockwise direction to create **A** and **B**, respectively. Bottom row - discrete Fourier Transform corresponding to the image in the row above after thresholding. The red line represents the linear least-squares regression model calculated from the coordinates of the remaining maxima in the discrete Fourier Transform after thresholding. **D** - Pearson correlation coefficient (r) = 0.811, p = <0.001. **E** - r = 0.116, p = <0.001. **F** - r = -0.0259, p = 0.351.

of the photoreceptor mosaic in 2D within the frequency domain, the presence of eye movement related image distortions can be quantified.

Each synthetic frame from Dataset 1 (described in Section 4.3.6.1) was divided into non-overlapping image patches, 40 × 40 pixels in size, and each image patch was transformed using the discrete Fourier Transform. A size of 40 pixels was selected to ensure sufficient cone photoreceptors (approximately 36) were present in each image patch to adequately represent the local photoreceptor mosaic. Although having a larger image patch would increase the spatial frequency resolution of the discrete Fourier Transform, small fixational eye movements, which cause small areas of distortion, may become averaged out within a larger image patch thereby underestimating the eye movement related distortion. To identify samples

representing frequencies with a significant contribution from the photoreceptor mosaic a threshold was applied to the discrete Fourier Transform using Otsu's thresholding method [138], which assumes an image contains pixels with a bimodal distribution (foreground and background) and calculates the optimal threshold to maximise variance between pixel classes (or minimise variance within pixel classes). After thresholding, the discrete Fourier Transform was binarised and the Pearson correlation coefficient (r_{DFT}) calculated from the coordinates of the remaining maxima. In a frame with no eye motion, the maxima would be expected to form a disc around the zero frequency component (Yellott's ring), giving a correlation coefficient value of zero (Figure 4.8E). Eye movement related image distortions would result in an elliptical (assuming constant motion over the image patch) rather than circular arrangement of the maxima and a positive or negative correlation coefficient (Figure 4.8D and F).

Dataset 1 (described in Section 4.3.6.1), containing normal simulated eye movements, was used to evaluate intraframe eye motion detection. A vector of eye positions was added to the vector of raster scan positions and each pixel intensity is simulated as though recorded from those retinal positions. The amount of induced eye motion for a synthetic frame was quantified by summing the standard deviations of the vector of eye positions that were introduced during the generation of each constituent non-overlapping image patch. The prediction of the amount of eye motion for each synthetic frame was calculated by summing the absolute values of the r_{DFT} from each constituent non-overlapping image patch. There was no linear association between the induced and predicted eye motion within a frame for natural simulated eye movements (r = -0.0008123, p = 0.905, Figure 4.9).

Next, the relationship between r_{DFT} and induced shear (horizontal eye motion) was investigated. Although both horizontal and vertical eye movements cause image distortions, due to the aliasing artefacts caused by vertical eye movements these were not simulated. A fixed amount of shear (0°, 1°, 2°, 3°, 4°, 5°, 6°, 7°, 8°, 9°, 10°, 20°, 30°, 40°, 50°, 60°and 70°) was added to synthetic images from a retinal eccentricity of 3° (see Section 4.3.5), prior to the addition of residual aberration

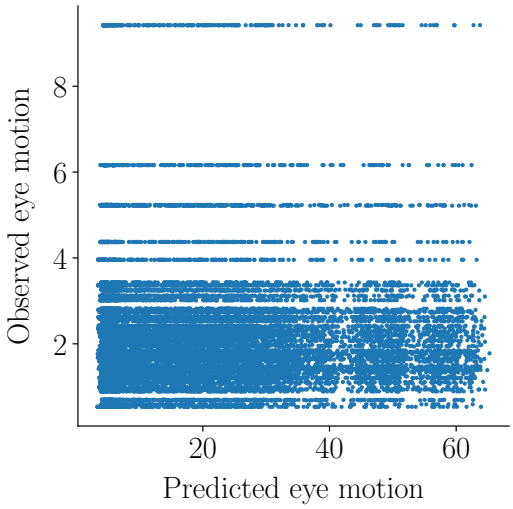


Figure 4.9: Predicting normal simulated eye movements. Each datum represents a single synthetic frame containing normal simulated eye movements with added noise and residual aberration (n = 21,600). Predicted and observed (induced) eye motion was calculated for each frame. Observed eye motion was determined by summing the standard deviation of the eye motion added to each image patch within a synthetic frame. Predicted eye motion was calculated by summing the absolute value of the Pearson correlation coefficient (r_{DFT}) of each image patch within a synthetic frame. r = -0.000813, p = 0.905.

and noise. The maximum shear of 70° applied to synthetic images is equivalent to a speed of 15.9°/s, which when compared with typical microsaccade speeds is biologically plausible [139]. To vary the image signal to noise ratio, the signal (measured in nW at the pinhole) was varied from 0.2 nW to 0.5 nW in steps of 0.1 nW. To simulate residual aberration, fixed percentages (50%, 40%, 30%, 20%, 10% and 0%) of the mean Zernike coefficients from 10 randomly selected eyes were applied. Residual aberration and noise were added to a single synthetic image following addition of the induced shear. This resulted in a dataset of 4,080 synthetic frames, 480 × 300 pixels in size (Figure 4.10). These synthetic frames were divided into 40 × 40 pixel patches and the r_{DFT} calculated for each image patch.

A greater positive correlation exists between r_{DFT} and shear values greater than or equal to 10° (r = 0.955, p < 0.001, Figure 4.11b) compared to shear values less than or equal to 10° (r = 0.385, p < 0.001, Figure 4.11a). Stratifying the relationship between r_{DFT} and induced shear by the image signal to noise ratio demonstrates a greater positive correlation for images with a higher signal

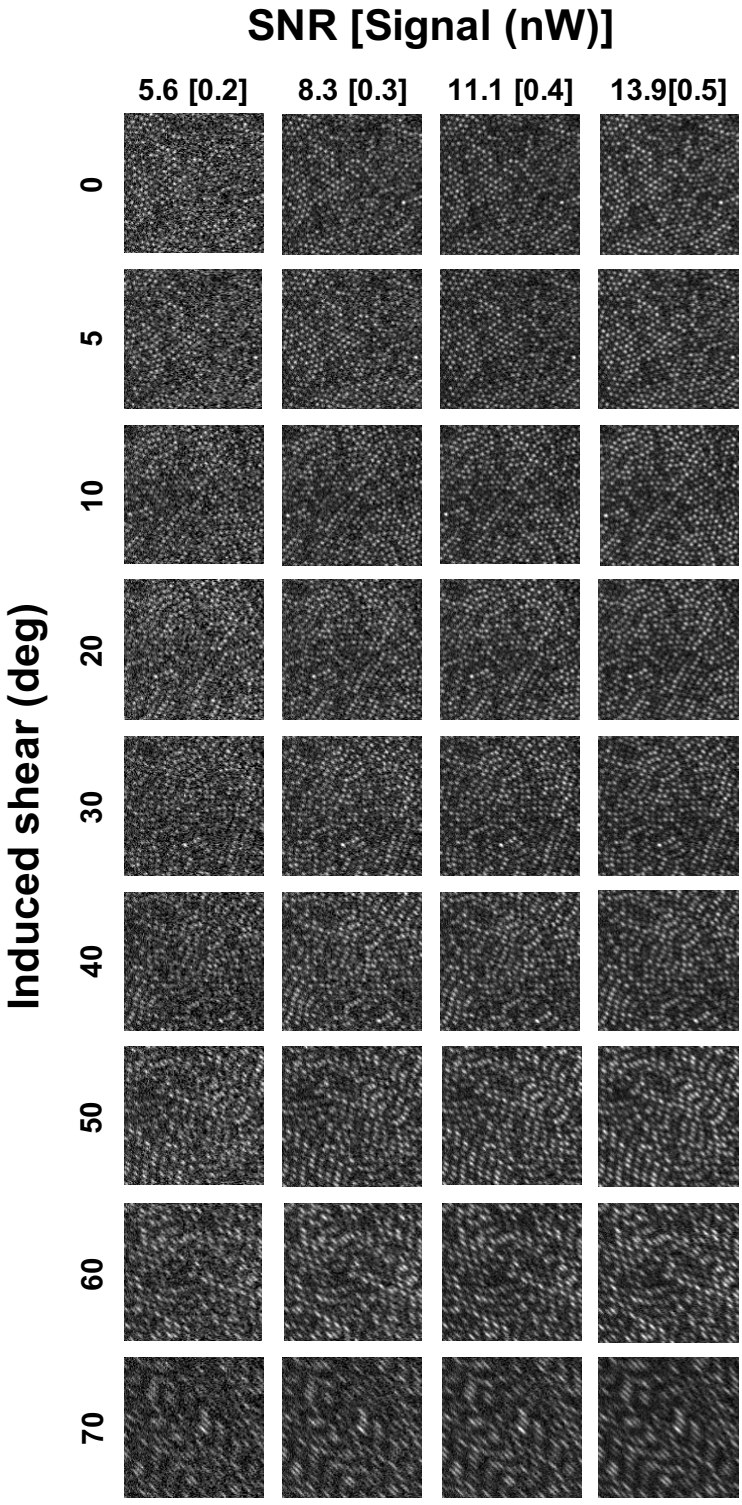
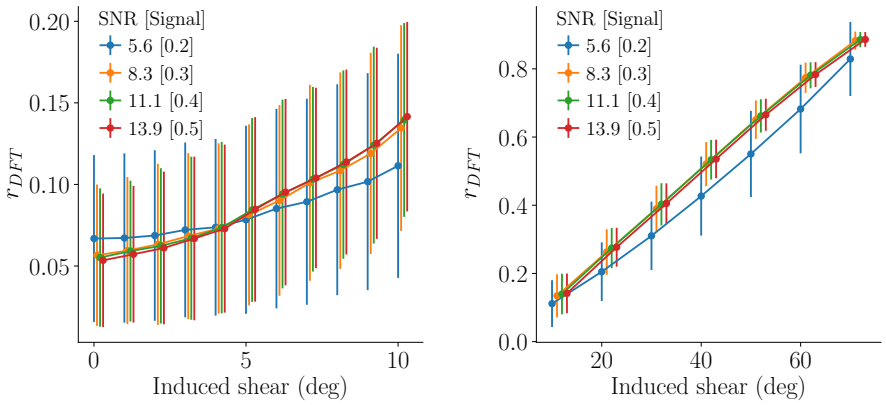


Figure 4.10: An example of the synthetic frames with induced shear used for evaluating intraframe eye motion detection. In this example all images have the same added residual aberration (10% of the wavefront aberrations from a single eye). The image signal to noise ratio (SNR) i.e. the image signal (measured in nanowatts (nW) at the pinhole) has been varied.



(a) Induced shear $\leq 10°$ **(b)** Induced shear $\geq 10°$

Figure 4.11: The relationship between r_{DFT} (the Pearson correlation coefficient of the maxima in a discrete Fourier transform of an image patch) and induced shear. Each datum represents the mean r_{DFT} for all image patches of a given signal to noise ratio (SNR) or image signal, measured in nanowatts (nW) at the pinhole. Error bars represent the standard deviation. All data points for each image signal value have been offset horizontally for clarity. **(a)** Correlation between r_{DFT} and induced shear (0° to 10°) for all image patches: $r = 0.385$, $p < 0.001$. There was a greater positive correlation between r_{DFT} and induced shear for images with a SNR of 13.9 ($r = 0.465$, $p < 0.001$) compared to those with a SNR of 11.1 ($r = 0.442$, $p < 0.001$), 8.3 ($r = 0.403$, $p < 0.001$) or 5.6 ($r = 0.232$, $p < 0.001$). **(b)** Correlation between r_{DFT} and induced shear (10° to 70°) for all image patches: $r = 0.955$, $p < 0.001$. There was a greater positive correlation between r_{DFT} and induced shear for images with a SNR of 13.9 ($r = 0.983$, $p < 0.001$) compared to those with a SNR of 11.1 ($r = 0.982$, $p < 0.001$), 8.3 ($r = 0.979$, $p < 0.001$) or 5.6 ($r = 0.911$, $p < 0.001$).

to noise ratio when compared to images with a lower signal to noise ratio (Figure 4.11). Residual aberration does not have a substantial effect on the relationship between r_{DFT} and induced shear.

A linear regression model, which can be used to predict image shear as a continuous variable, was initially used to test whether r_{DFT} predicts image shear, however the residuals were not uniformly distributed. A binary logistic regression model, which can be used to predict image shear as a categorical variable, was therefore used to investigate r_{DFT} as a predictor of image shear. Modelling the probability of a frame belonging to one of two classes, i.e. frames containing more or less shear than a minimum shear threshold (T_{min}), will identify all frames containing less shear than the minimum threshold. However, in order to identify a single frame with the least shear to be used as the reference frame, iteratively predicting

image shear using more than one T_{min} will be required. For example, if two frames contain 10° and 30° of shear respectively, using a T_{min} of 40° will not be sufficient to distinguish between these two frames. By iterating through a range of T_{min} values, it will be possible to identify a frame with the least shear. A range of T_{min} values, where $0° < T_{min} \leq 70°$, was used. Based on theoretical considerations, as the signal to noise ratio of an image decreases the gradient of the regression slope would be expected to decrease i.e. the model would poorly predict image shear from images with a low signal to noise ratio. As noise is distributed randomly across an image, it will not contain a directional component. In images with a low signal to noise ratio, directionless noise will dominate the repeating pattern of a photoreceptor mosaic (the signal). Therefore, the model would be expected to perform poorly when predicting image shear in images with a low signal to noise ratio.

Due to the problem of class imbalance at high or low minimum shear thresholds the true positive rate (TPR) and false positive rate (FPR) were used to evaluate the model's accuracy using the following definitions:

$$TPR = \frac{TP}{TP + FN} \tag{4.5}$$

$$FPR = \frac{FP}{FP + TN} \tag{4.6}$$

where TP represents the number of true positives, FP the number of false positives, TN the number of true negatives and FN the number of false negatives. Both the true positive rate and the false positive rate range between 0 and 1, with a true positive rate of 1 and a false positive rate of 0 signifying perfect identification of all positive and negative samples. While Precision and Recall are focussed on the positive class, true positive rate and false positive rate focus on the positive and the negative class, measuring the models ability to distinguish between classes.

The results of the logistic regression analysis are shown in Table 4.2 and Figure 4.12, which indicate that minimum shear thresholds of 10° or greater provide a suitable balance between accuracy (true positive rate and false positive rate) and the ability to distinguish between frames with different shear values. Although a

Table 4.2: Logistic regression model results investigating r_{DFT} as a predictor of induced image shear.

					95% CI			
T_{min}	B	$SE\ B$	z	p	**0.025**	**0.975**	TPR	FPR
1°	12.6	0.155	81.3	< 0.001	12.3	12.9	1.00	1.00
2°	13.1	0.114	115	< 0.001	12.9	13.3	1.00	1.00
3°	13.7	0.097	141	< 0.001	13.5	13.9	1.00	1.00
4°	14.2	0.087	163	< 0.001	14.1	14.4	0.931	0.678
5°	14.9	0.082	183	< 0.001	14.8	15.1	0.884	0.435
6°	15.5	0.078	198	< 0.001	15.3	15.6	0.855	0.286
7°	16.0	0.076	211	< 0.001	15.9	16.2	0.837	0.186
8°	16.9	0.076	221	< 0.001	16.7	17.0	0.834	0.119
9°	18.4	0.081	227	< 0.001	18.2	18.5	0.851	0.0700
10°	21.6	0.096	226	< 0.001	21.4	21.8	0.891	0.0373
20°	33.5	0.180	186	< 0.001	33.2	33.9	0.953	0.0183
30°	26.8	0.147	183	< 0.001	26.5	27.1	0.964	0.0188
40°	21.2	0.116	183	< 0.001	20.9	21.4	0.959	0.0231
50°	20.0	0.110	181	< 0.001	19.8	20.2	0.951	0.0249
60°	22.4	0.123	182	< 0.001	22.2	22.6	0.936	0.0217
70°	33.8	0.208	163	< 0.001	33.4	34.2	0.928	0.0135

B - unstandardised beta (correlation coefficient), CI - confidence interval, FPR - false positive rate, p - probability, SE - standard error, T_{min} - minimum shear threshold, TPR - true positive rate, z - z-statistic.

minimum shear threshold of 30° provides a low false positive rate with the highest true positive rate, using a single threshold of 30° will not enable frames with differing amounts of shear to be distinguished as described in the example above. As all minimum shear thresholds less than 10° have either a low true positive rate or high false positive rate, a minimum shear threshold greater than or equal to 10° was selected. A minimum shear threshold of 10° was included (low false positive rate with an acceptable true positive rate, which conservatively rejects frames that may be bad) in order to facilitate the identification of frames with low amounts of shear.

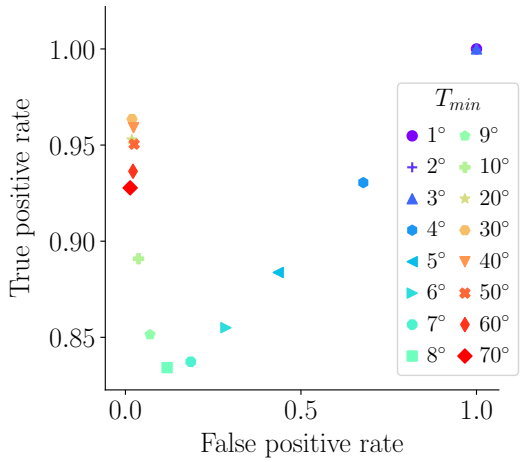


Figure 4.12: Logistic regression model results investigating r_{DFT} as a predictor of induced image shear using a range of minimum shear thresholds (T_{min}). The best T_{min} (highest true positive rate and lowest false positive rate) should be closest to the top left corner of the plot.

4.3.10 Reference frame selection

Figure 4.13 illustrates the algorithm workflow used to automatically select a reference frame from each hierarchical cluster. Clusters containing fewer than 10 frames were rejected as they will result in an average image with a low signal to noise ratio (too few images to contribute to the average). The interaction between an image's signal to noise ratio and its quality metrics when used to predict Strehl ratio or image shear can be seen in Figure 4.7 and Figure 4.11, respectively. A higher image signal to noise ratio provides a more accurate prediction of residual aberration (see Figure 4.7) or eye motion (see Figure 4.11). By considering the noise present within the images of a cluster as a constant, total intensity can be used to estimate the signal present within each image. Therefore, frames with a total intensity below the 50th percentile were rejected from consideration as possible candidates for the reference frame. To determine the residual aberration present within each image the Strehl ratio of the remaining frames was predicted and those below the 50th percentile were rejected from further consideration. Patients with macular disease exhibit an increase in fixational eye movements and are known to have unstable fixation [57–59] compared to normal observers. To identify the presence of intraframe eye motion, the probability of an image containing shear

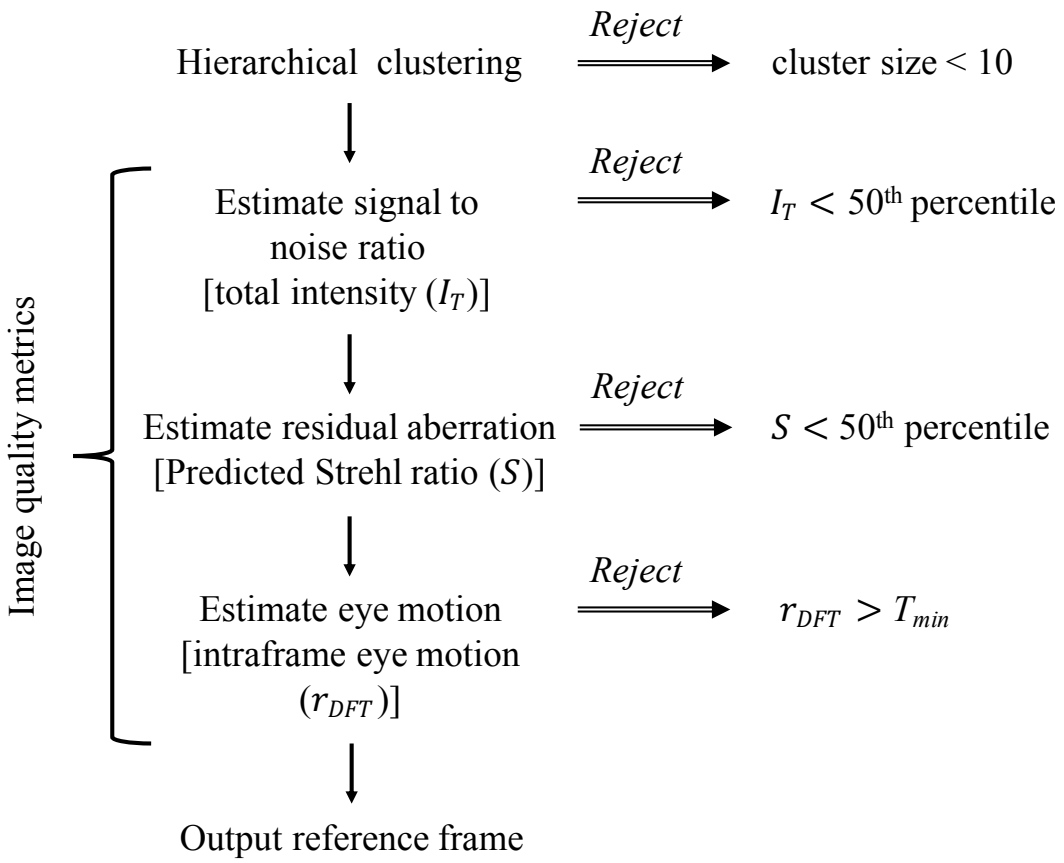


Figure 4.13: Reference frame selection module workflow. Following hierarchical clustering, clusters of less than 10 frames were rejected. Next, image quality metrics were sequentially calculated for frames from clusters that passed the first stage of rejection. Frames were rejected based on their estimated signal to noise ratio (I_t), residual aberration (S) and intraframe eye motion (r_{DFT}). The frame with the highest total intensity that passed all other stages of rejection was output as the reference frame for the cluster. T_{min} - minimum shear threshold.

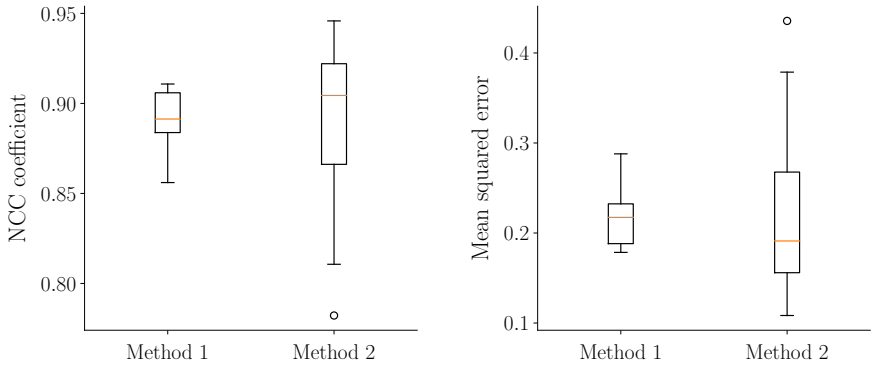
greater than or equal to T_{min} was determined (see Section 4.3.9), where T_{min} is the minimum shear threshold. To identify frames containing the least shear, an initial T_{min} of 10° was used. T_{min} was increased in steps of 10° to a maximum value of 70° until at least one frame containing shear less than T_{min} was identified. An iterative approach with a range of minimum shear thresholds was chosen to account for the variable fixation stability of patients with macular disease. Finally, of all frames not rejected at earlier stages, the frame with the highest total intensity was output as the reference frame for the cluster and the existing image processing pipeline was used to motion correct and average the frames in that cluster.

4.4 Results

4.4.1 Algorithm evaluation

The automated reference frame selection algorithm (Method 1) was evaluated using Dataset 2 (described in Section 4.3.6.2) and benchmarked against the default method of reference frame selection used with the Oxford AOSLO (Method 2) - identification of the most similar frame using normalised cross correlation. By using the most similar frame within an imaging sequence as the reference frame, Method 2 assumes that fixational eye movements are random and therefore the average motion will be zero. To assess the ability of Method 1 to select a reference frame with minimal motion-distortion and minimal residual aberration, averaged images that had been generated using the selected reference frame, and also the reference frame itself, were compared (using normalised cross correlation and the mean squared error, described below) to a perfect reference image with diffraction but no residual aberration, eye motion or noise.

Method 1 outputs one reference frame for each spatial cluster (see Section 4.3.3) with typically between two and four spatial clusters output per image set (see Section 4.3.6.2 for description of an image set), while Method 2 outputs one reference frame per image set. To maintain equal sample sizes, analysis was performed on the best spatial cluster for Method 1, which was determined by comparing the reference frame of each spatial cluster to the perfect reference image using normalised cross correlation. Averaged images were generated using the reference frame output from each method. Single frames contributed to the averaged image if the variance of the extracted intraframe eye motion was less than the mean variance of all frames within the image set. Averaged images and reference frames from the two reference frame selection methods were cropped to a common area and registered to within 0.1 pixels of the perfect reference image using linear translation and cubic spline interpolation. Averaged images and reference frames from both methods were compared to the perfect reference image by calculating the normalised cross correlation coefficient and mean squared error (after normalisation of pixel intensities by subtracting



(a) Normalised cross correlation coefficient

(b) Mean squared error

Figure 4.14: Evaluation of reference frame selection. Reference frames from both methods were compared to the perfect reference image by calculating the normalised cross correlation (NCC) coefficient and mean squared error. There was no statistically significant difference when comparing normalised cross correlation coefficients ($p = 0.959$, $n = 10$) or mean squared error ($p = 0.959$, $n = 10$) between the reference frame selection methods.

the mean pixel intensity and then dividing by the standard deviation of pixel intensities) to evaluate each method's performance. The performance metrics for both reference frame selection methods were compared using the Wilcoxon matched pairs signed rank test due to violation of the assumption that data were sampled from a normal distribution.

4.4.1.1 Evaluating reference frames

The median normalised cross correlation coefficient was 0.891 (IQR 0.0221, range: 0.856 - 0.911) and 0.904 (IQR 0.0559, range: 0.782 - 0.946), and the median mean squared error was 0.217 (IQR 0.0442, range: 0.179 - 0.288) and 0.191 (IQR 0.112, range: 0.108 - 0.436) for reference frames from Method 1 and 2, respectively (Figures 4.14 and 4.15). The normalised cross correlation coefficients ($p = 0.959$, $n = 10$) and mean squared error ($p = 0.959$, $n = 10$) of the reference frames were not significantly different between the two reference frame selection methods.

4.4.1.2 Evaluating averaged images

The median number of single frames contributing to each averaged image was 17 (IQR 4, range: 8 - 24) and 54 (IQR 8.5, range: 50 - 63) for Methods 1 and 2,

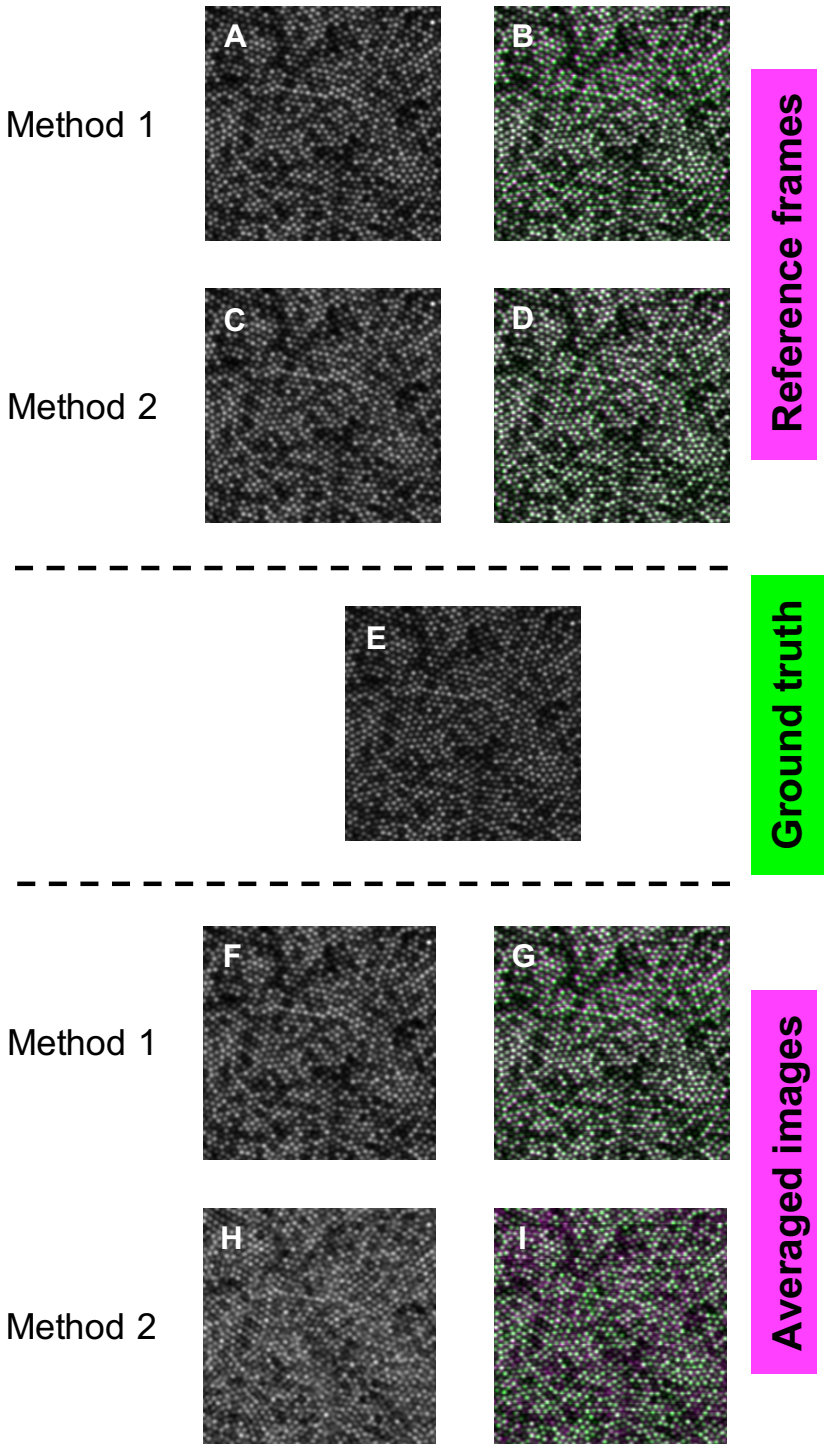
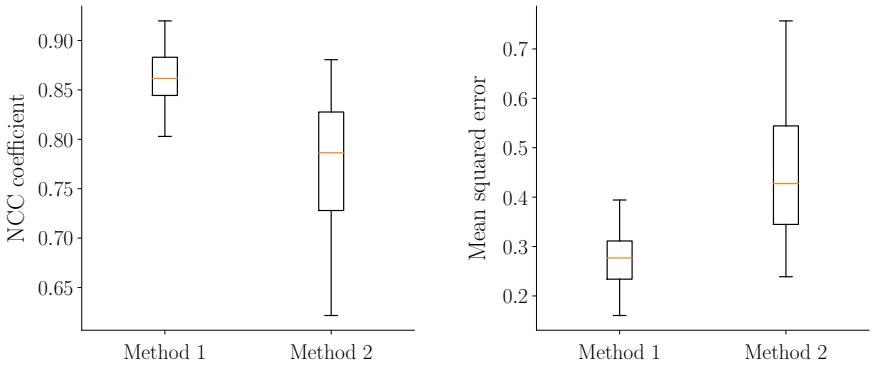


Figure 4.15: Performance of the automated algorithms for one example image set. Reference frames (A, C) and averaged images (F, H) output by the automated algorithms are shown in the left column. An overlay of the output image (pink) with the ground truth reference image (green) is shown in the right column (B, D, G, I). Where both the output image and the ground truth overlap perfectly, the overlay will appear achromatic (greyscale). Normalised cross correlation coefficients for Method 1 are 0.841 in (**A**) and 0.803 in (**F**), and for Method 2 are 0.913 in (**C**) and 0.747 in (**H**).



(a) Normalised cross correlation coefficient

(b) Mean squared error

Figure 4.16: Evaluation of averaged images. Averaged images from both reference frame selection methods were compared to the perfect reference image by calculating the normalised cross correlation (NCC) coefficient and mean squared error. The normalised cross correlation coefficients ($p = 0.0166$, $n = 10$) and mean squared error ($p = 0.0166$, $n = 10$) of the averaged images differed significantly between the two reference frame selection methods.

respectively. The median normalised cross correlation coefficient was 0.862 (IQR 0.0387, range: 0.803 - 0.920) and 0.786 (IQR 0.0997, range: 0.622 - 0.881), and the median mean squared error was 0.277 (IQR 0.0773, range: 0.160 - 0.394) and 0.427 (IQR 0.199, range: 0.239 - 0.757) for averaged images from Method 1 and 2, respectively (Figures 4.16 and 4.15). The normalised cross correlation coefficients ($p = 0.0166$, $n = 10$) and mean squared error ($p = 0.0166$, $n = 10$) of the averaged images from Methods 1 and 2 differed significantly, with averaged images from Method 1 showing higher normalised cross correlation coefficients and lower mean squared error when compared with the perfect reference image.

4.5 Discussion

This chapter describes the design and implementation of a fully automated algorithm for reference frame selection from sequences of confocal AOSLO images of the cone photoreceptor mosaic. The proposed algorithm is able to select reference frames with minimal distortion and residual aberration, which is important for increasing the accuracy and repeatability of cone mosaic metrics [140].

Pixel intensity based methods, which identify outlier frames based on their intensity, are widely used to automatically reject frames containing blinks [122]. These methods use a relative threshold of image intensity (typically a multiple of the standard deviation of image intensity from the mean) to reject frames containing a blink, which may lead to the rejection of good frames if there are no blinks within an image sequence to increase the standard deviation of image intensity. Using an absolute threshold of image intensity avoids this problem. Pixel intensity based methods of frame rejection are adequate for images of a normal photoreceptor mosaic, where the variations in image intensity across an image sequence is expected to be small. However, a greater variation in image intensity can be expected from patients with retinal disease due to degenerative structural changes within the retina combined with increased eye movements [57–59]. The proposed method of identifying low intensity frames using an absolute threshold is more robust to the variations in image intensity that can be expected in patients with retinal disease.

Patients with macular disease are known to have unstable fixation [57–59], which leads to AOSLO images being acquired over a greater retinal area during each image acquisition sequence. In the algorithm proposed by Salmon *et al.* [122], the 'top down' K-Means algorithm was used to cluster frames based on their relative spatial locations and silhouette analysis was used to predict the number of clusters to form. Silhouette analysis establishes how well a group of observations are clustered by measuring the separation distance between clusters, and is used to determine the optimum number of clusters during K-Means clustering. Over-clustering (too many clusters predicted) requires additional merging of clusters to obtain final clusters of frames with a minimum amount of overlap. This study used a 'bottom up' approach with hierarchical agglomerative clustering and the application of a minimum distance threshold, allowing the optimal number of clusters to be determined automatically in a single step. A major advantage of clustering includes the increased number of frames that can be used from an image sequence since many reference frames, rather than just one, are selected.

A perfect, distortion free reference frame would enable the estimation and correction of motion artefacts and a motion free average image with a high signal to noise ratio. Obtaining reference images with no eye motion artefacts would be possible with an AO flood illumination system, however, the images may still contain residual aberration and in extreme cases (which may be more likely with a patient) - motion blur. Evaluating reference frame selection requires a method to determine which frame contains the least eye motion and residual aberration while reflecting a true representation of the photoreceptor mosaic. Although synthetic data allow the presence of these artefacts to be accurately quantified, they have not previously been used for the evaluation of image quality metrics and reference frame selection relating to AOSLO images.

Despite being widely used for the detection of eye motion artefacts and the selection of minimally distorted frames, the relationship between the Pearson correlation coefficient of the maxima from the discrete Fourier Transform of an image patch (r_{DFT}), used to quantify intraframe motion (see Section 4.3.9), and fixational eye movements has not been comprehensively evaluated. Using synthetic data, this study has demonstrated that there is no relationship between induced and predicted normal simulated fixational eye movements (Figure 4.9) and only a weak relationship between induced and predicted eye movements introducing shear artefacts of less than 10° (Figure 4.11a). Shear of 10° is equivalent to a speed of 1.02°/s, which is approximately 15 times slower than the maximum shear of 70° applied to synthetic images (15.9°/s), and slower than a typical microsaccade [139].

The averaged images generated using Method 1 had greater similarity to the perfect reference image compared to those generated using Method 2 ($p = 0.0166$, $n = 10$). There was no statistically significant difference between reference frames selected using Method 1 or Method 2 when compared to a perfect reference image ($p = 0.959$, $n = 10$). It is likely that the averaged images generated using Method 1 were significantly better than those from Method 2 due to clustering of spatially overlapping frames, which will improve motion correction and subsequent averaging. Although there was no significant different between reference frames from Methods

1 and 2, it is possible that the small sample size coupled with the lower power of nonparametric tests may not have detected a small effect size. Method 2 uses normalised cross correlation to select a reference frame with the greatest similarity to all images from an image sequence. Normal observers have stable fixation and exhibit minimal fixational eye movements compared to patients with macular disease (who are known to have unstable fixation) [57–59]. Using Method 2 to select a reference frame from AOSLO images of a normal observer is therefore likely to identify an image containing minimal eye motion. In contrast, while Method 1 is able to identify minimally distorted images by predicting the probability of an image containing less than or equal to 10° of shear, all of the images identified are considered to contain equal eye motion (10° of shear or less) and the reference frame is determined by identifying the image with the greatest total intensity. However, in participants with unstable fixation, such as patients with macular disease, identifying a reference frame with the greatest similarity to all images from an image sequence is unlikely to yield a minimally distorted reference frame as more images will contain larger eye movements. Objectively predicting the amount of shear within an image enabled the proposed algorithm to reliably identify images with less than 70° of shear. However, experimentation has demonstrated that further work is required to reliably identify minimally distorted images containing less than 10° of shear.

Due to the feature-specific nature of the intraframe eye motion detection stage, which detects disruptions to the modal spacing of the photoreceptor mosaic in the frequency domain, this algorithm will be unable to detect intraframe motion and select an appropriate reference frame from images of the retinal nerve fiber layer or retinal vasculature. Alternative methods that are not specific to the photoreceptor mosaic, such as the steerable pyramid [141], may enable evaluation of motion artefacts within these images. This would allow generalisability of the algorithm to images of different retinal features and may also improve accuracy in areas of retinal degeneration where there is significant disruption of the photoreceptor mosaic.

Using synthetic data for the evaluation of image quality metrics enables their relationship with image artefacts to be modelled. Despite the synthetic data

providing a realistic cone photoreceptor mosaic of a healthy retina, rod photoreceptors and retinal vasculature were not simulated. Synthetic data simulating a photoreceptor mosaic affected by retinal degeneration are required to simulate the retina of patients with retinal disease. This gap between synthetic and real datasets may affect the algorithm's performance with real data. Due to the lack of access to an AO flood illumination system and the difficulty of accurately quantifying the presence of residual aberration and eye motion related artefacts in real images where there is no ground truth, it has not been possible to evaluate the algorithm's performance using real data.

5

Automated cone photoreceptor identification using deep learning

Contents

5.1 Contributions 105
5.2 Background 105
5.3 Methods 109
5.3.1 Evaluation protocol 109
5.3.2 Data 113
5.4 Experiment 1 117
5.4.1 Input images 117
5.4.2 FCRN and U-Net architectures 117
5.4.3 Results 121
5.5 Experiment 2 124
5.5.1 Results 125
5.6 Experiment 3 130
5.6.1 Input images 131
5.6.2 New U-Net architecture 132
5.6.3 Results 136
5.7 Experiment 4 140
5.7.1 Input images 142
5.7.2 U-Net transfer learning 143
5.7.3 Results 144
5.7.4 Comparison with state-of-the-art 150
5.8 Discussion 152

5.1 Contributions

My contributions to this work include: study concept and design; algorithm development and coding; collection, analysis and interpretation of data. The simulation code used to generate the synthetic data described in Section 5.3.2.1 was developed and written by Laura Young. The data described in Section 5.3.2.2 was collected using the Oxford AOSLO in collaboration with Sarah Regan and Anna-Katharina Hauperich. I had full access to all of the data presented in this chapter and take responsibility for the integrity of the data and the accuracy of the data analysis.

5.2 Background

The ability to quantify changes in the living photoreceptor mosaic is useful for providing diagnostic and prognostic information about retinal diseases, for understanding disease mechanisms, and for evaluating novel treatments. Identification of individual photoreceptors is often a prerequisite for generating quantitative metrics of the photoreceptor mosaic (such as photoreceptor density or reflectivity). Photoreceptor identification is typically performed manually, which is both subjective and time consuming. The significant effort required for the analysis of these images is a major rate-limiting step in the clinical application of AO retinal imaging [61, 62].

A number of algorithmic approaches have been developed for automated cone detection in AO images [63–67]. However, while they may show good performance for the specific problem they were designed for, their reliance on a predefined set of rules limits their generalisability. For example, the adaptive filtering and local detection method described by Cunefare *et al.* [142], which was developed using images from healthy volunteers, performed poorly when detecting cones from images of participants with Stargardt disease [143]. Many algorithms developed for automated cone detection still require manual supervision when detecting cones in diseased eyes or in low-quality images [62].

Unlike traditional algorithms that use explicit instructions to perform a task, machine learning algorithms build statistical models to learn patterns within sample data in order to perform a specific task. The present work aims to evaluate the utility of machine learning for automated cone detection from confocal adaptive optics scanning laser ophthalmoscope (AOSLO) images. Deep learning is a subfield of machine learning based on artificial neural networks - algorithms inspired by the biological neural networks of the brain. Like in the brain, the basic building block of an artificial neural network is a neuron. A neuron is a mathematical operation that has learnable parameters (Figure 5.1). Similarly to how humans learn from experience, artificial neural networks repeatedly perform a task, learning from each repetition to improve the outcome (this process is called training). During training, an artificial neural network learns by adjusting the weights and biases of the neurons to minimise the error between its output and the ground truth labels in the training dataset. The word "deep" in deep learning refers to the number of layers within the artificial neural network through which data are transformed. A layer is simply a collection of neurons that take in an input and produce an output. An artificial neural network consists of an input layer, an output layer and one or more hidden layers, so called as they are located between the input and output layers. Unlike other machine learning techniques that reach a plateau in performance as training data increases, deep learning is scalable. That means as deep neural networks are trained with more data, their performance continues to improve. Deep learning provides an effective way of automatically extracting and learning features directly from the image data, potentially increasing generalisability.

A convolutional neural network (CNN) is a class of artificial neural network that is commonly used for image recognition and classification, and has shown near human-level or surpassed human-level performance in a variety of applications including image analysis tasks [145–149]. Convolutional neural networks are very similar to ordinary artificial neural networks, however, the CNN architecture makes the explicit assumption that its inputs are images. The convolutional layer is the core building block of a CNN and consists of learnable filters (also known as

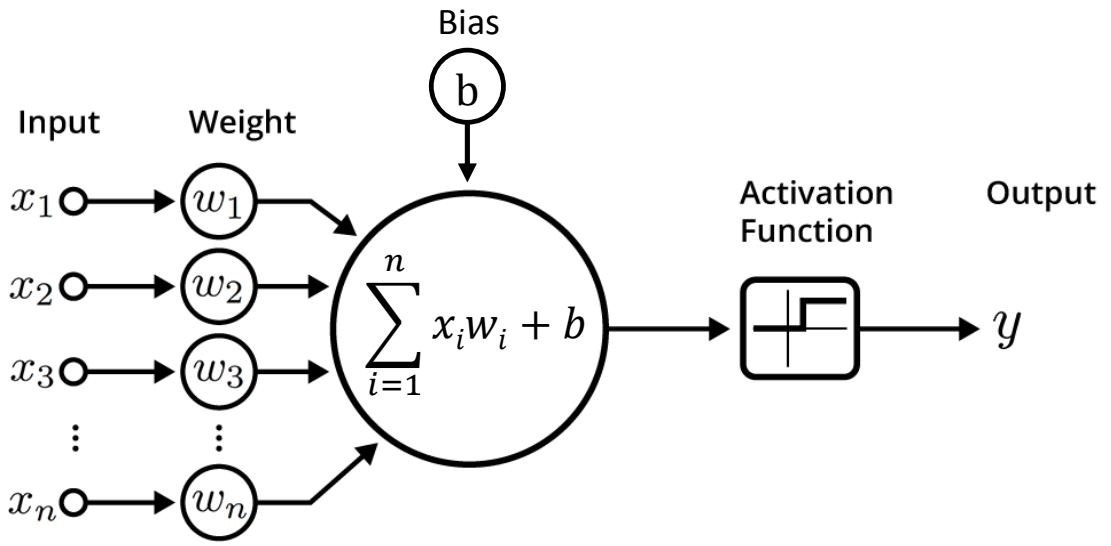


Figure 5.1: Artificial neuron (adapted from [144]). The $x_i(i = 1, 2...n)$ values refer to the neuron's inputs. The $w_i(i = 1, 2...n)$ and b values refer to the neuron's weights and biases (the learnable parameters), respectively. The activation function introduces non-linearity and controls firing of the neuron.

kernels). Each filter (made up of learnable weights and biases) is convolved with its input to produce a 2-dimensional output (feature map). Convolutional layers are typically stacked with activation and pooling layers to form CNN architectures. As a biological neuron can fire by sending a signal along its axon, the firing of an artificial neuron can be modelled with an activation function. Activation layers apply a non-linear activation function. The Rectified Linear Unit (ReLU) is a commonly used activation function that returns the value of the input, or 0.0 if the input is less than 0.0, by computing the function $f(x) = max(0, x)$. ReLU layers improve network performance by introducing non-linearity [150] and reducing training time [151]. Pooling layers reduce the spatial dimensions of the feature map by performing a downsampling operation, which reduces computation and builds invariance to small distortions [150]. The present work evaluated two types of CNN architecture chosen based on past successes in similar domains.

The problem of automated cone photoreceptor identification from AOSLO images is similar to that of automated cell counting in microscopy. A recent paper by Xie *et al.* [152] reports the use of a Fully Convolutional Regression Network (FCRN) to predict cell spatial density over an image. This method was reported as being

particularly useful for the detection of overlapping cells where traditional single-cell based segmentation methods do not work well. The problem of overlapping cells is similar to imaging near the resolution limit of an AO imaging system (such as at the fovea, with a small pupil, or with a long imaging wavelength) where convolution with the point spread function can lead to overlap of structural features making identification of individual photoreceptors challenging.

The U-Net is a CNN developed in 2015 by Ronneberger *et al.* for biomedical image segmentation [153] and consists of a contracting and expansive path. The contracting path follows the traditional architecture of a CNN and reduces the spatial information but increases the feature information by encoding the input image into feature representations at different spatial levels through successive convolutions, non-linear operations and pooling. In the expansive path pooling operations are replaced by upsampling of the feature map. It combines feature and spatial information through concatenation of higher resolution features from the contracting path with upsampled features from the expansive path in order to better localise and learn image representations. The large number of feature channels in the expansive path allows the network to propagate feature information to higher resolution layers.

Traditional deep learning approaches require large volumes of training data. Acquiring large high-quality datasets that are representative of the patient cohorts seen in real-world healthcare settings is difficult. These datasets also require accompanying annotations generated by experts, which are time consuming to produce. Not only are these large datasets difficult to acquire but they are not practical for retinal diseases with a low prevalence. In order to address the limited availability of training data, an additional aim of this study was to investigate whether synthetic data could be used to fit a deep learning model for the automated identification of cone photoreceptors from confocal AOSLO images. If this approach proves successful then it may be used with synthetic data representing photoreceptor mosaics from diseased eyes to develop tools for automated cone photoreceptor identification in patients with retinal disease. This chapter will describe two CNN models that were investigated (FCRN and U-Net).

5.3 Methods

The CNNs in this thesis were developed and implemented in the Python programming language (Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org) using the high-level neural networks Application Programming Interface Keras [154] with a Tensorflow backend [155]. All training was run on a NVIDIA Quadro P4000 graphics processing unit.

5.3.1 Evaluation protocol

The CNN models return predicted cone density maps from input images of the photoreceptor mosaic (Figure 5.2A). These output maps are 96 × 96 pixel greyscale images (the same size as the input images) where intensity represents the predicted probability of a cone being present. In order to evaluate the models' results, the predicted cones needed to be identified in these output maps. First, a minimum threshold equal to half the maximum intensity of the output map was applied to remove background pixels (Figure 5.2B). Next, the centroids of all peaks were identified using a peak local maximum function (Figure 5.2C).

Following identification of all peaks within the predicted cone density map (each peak represents the centre of a predicted cone), each predicted cone centre was paired with its corresponding ground truth. To achieve this, only predicted cone centres that were within half a cone size (described in Section 5.3.1.1 below) of a ground truth cone centre were allowed to form a cone pair (Figure 5.3A). Each ground truth cone could only be paired with one predicted cone. Half a cone size was calculated from the input image for the predicted cone density map being evaluated. If more than one predicted cone was identified within half a cone size of a ground truth cone centre then the predicted cone with the shortest euclidean distance was paired with the ground truth cone (Figure 5.3B).

Cones that are partially cropped as they are at the edge of an input image will be referred to as *edge cones*. If the centre of a ground truth edge cone does not fall within the input image, its prediction will be identified as a false positive (Figure 5.4). Conversely, if the centre of a predicted edge cone does not fall within

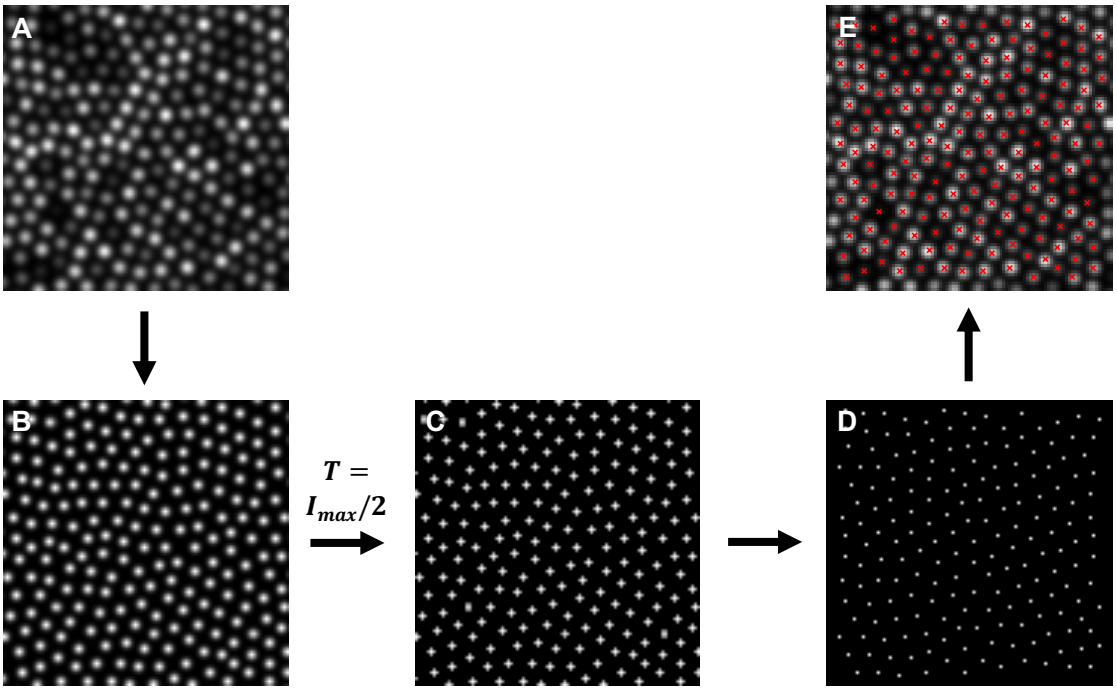


Figure 5.2: Identification of cone centres from the convolutional neural network (CNN) predicted cone density map. **A** - the input image to the CNN. **B** - the predicted cone density map output by the CNN where intensity represents the probability of a cone. **C** - the cone density map after a minimum threshold (T) has been applied. **D** - the centroids of all predicted cones after thresholding (each peak represents the centre of a predicted cone). **E** - an overlay of **D** on **A** illustrating the centre of each predicted cone with a red cross. I_{max} is the maximum intensity of the predicted cone density map.

the predicted cone density map (yet the centre of the ground truth cone does) it will be identified as a false negative (Figure 5.4). In order to overcome this problem, any cones within half a cone size of the edge of the input image or the edge of the predicted cone density map were excluded from evaluation.

5.3.1.1 Cone size

An estimate of cone size in the input image was determined by calculating the minimum of the inter-cone spacing (the distance between two cone peaks) and 6 × the cone sigma (as defined below). The cone spacing was determined by identifying Yellott's ring, an annulus in the Fourier domain representing the modal spacing of the photoreceptors [136, 137] (described in Section 4.3.9). The cone sigma (the standard deviation of a rotationally symmetric Gaussian function) was determined by fitting a 2D Gaussian to an average cone. As cones are represented as areas of

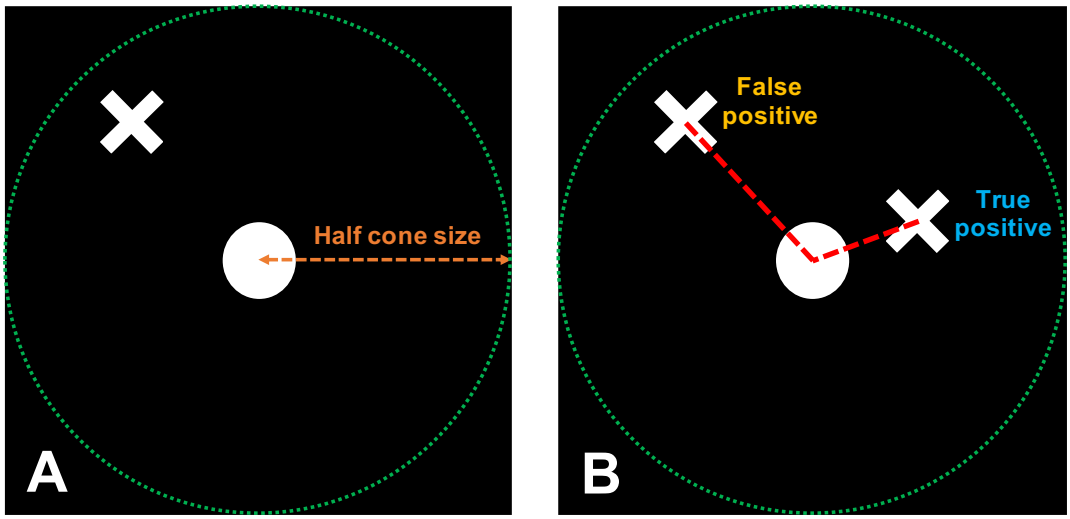


Figure 5.3: Identifying true and false positive predicted cones. **A** - if a predicted cone centre (white cross) is within half a cone size (described in Section 5.3.1.1) of a ground truth cone centre (white circle), a cone pair is formed. **B** - if the centre of more than one predicted cone (white cross) is within half a cone size of a ground truth cone centre (white circle), the predicted cone with the shortest euclidean distance (dotted red line) is paired with the ground truth cone and classified as a true positive while the other predicted cone is classified as a false positive.

increased intensity within images of the photoreceptor mosaic, an average cone was created by stacking the brightest cones and averaging through the stack. At low retinal eccentricities (with high cone density) inter-cone spacing was assumed to be equal to cone size. However, as a function of retinal eccentricity, inter-cone spacing increases at a greater rate than cone size [156]. Therefore the minimum of the inter-cone spacing and 6 $\times$ the cone sigma was taken as an estimate of cone size. The main function of estimating cone size was to assign cone pairs between ground truth and predicted cones. A multiple of 6 $\times$ the cone sigma represents approximately 99% of the total intensity of a Gaussian. This cone sigma multiple was selected as one of the inputs to estimate cone size in order to maximise the search field to find a cone's pair. To overcome the problem of incorrectly assigning a predicted cone to a cone pair, where more then one predicted cone was identified, only the predicted cone with the shortest euclidean distance was assigned to the cone pair. This is a conservative approach, designed to avoid missing a potential cone pair.

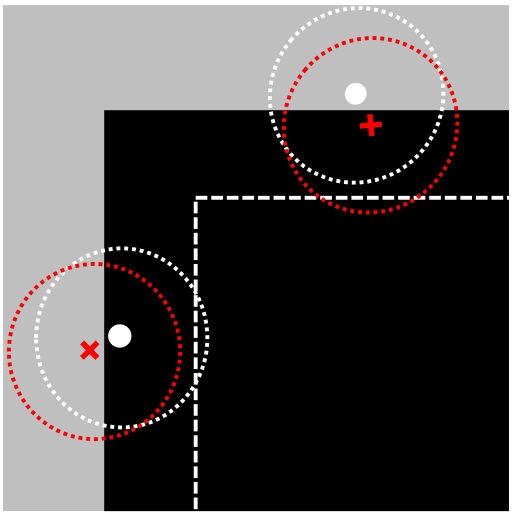


Figure 5.4: In order to overcome the problem of partially cropped cones at the edge of an input image (*edge cones*), cones within half a cone size (described in Section 5.3.1.1) of the input image or predicted density map edge were excluded from evaluation. The top of the figure shows a ground truth cone centre (white circle) that does not fall within the image (black square) but its corresponding predicted cone centre (red cross) does. If the predicted cone is not excluded, it will be identified as a false positive. Conversely, the left side of the figure shows a predicted cone centre that does not fall within the image and would be identified as a false negative if the corresponding ground truth cone was not excluded. The dotted white and red circles represent the size of the ground truth and predicted cones, respectively. The dashed white square represents half a cone size spacing from the edge of the image.

5.3.1.2 Jaccard similarity coefficient

The Jaccard similarity coefficient (also known as Intersection over Union, see Figure 5.5), a commonly used metric in image segmentation, was used to evaluate the accuracy of the CNN models. It ranges between 0 and 1, with 0 signifying no similarity and 1 signifying perfect similarity between the prediction and ground truth. The Jaccard similarity coefficient (J) of two sets, X (the prediction) and Y (the ground truth), is defined as follows:

$$J(X,Y) = \frac{X \cap Y}{X \cup Y} = \frac{TP}{TP + FP + FN} \tag{5.1}$$

where TP represents the number of true positives, FP the number of false positives and FN the number of false negatives (missed cones).

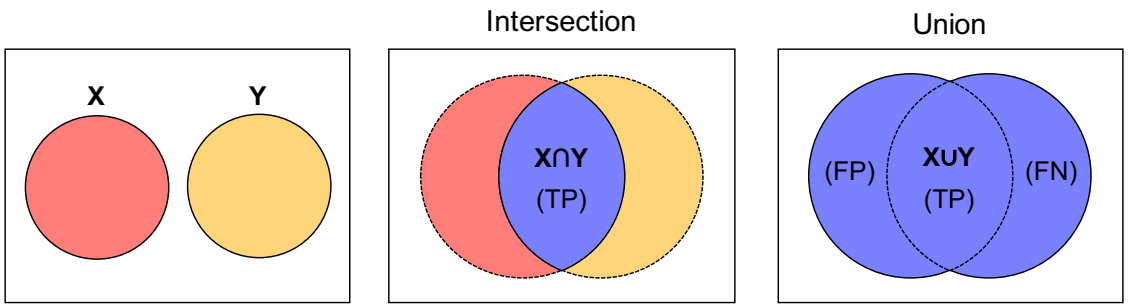


Figure 5.5: The Jaccard similarity coefficient of two sets (X and Y), where X represents the prediction and Y the ground truth. TP - true positives, FP - false positives, FN - false negatives.

5.3.2 Data

Three different datasets (one composed of synthetic cone mosaic images and two composed of real photoreceptor mosaic images from two different AOSLO systems) were used for training and testing the CNN models.

5.3.2.1 Dataset 3 - synthetic data

Synthetic images providing a realistic cone photoreceptor mosaic (described in Section 4.3.5) were generated to represent 19 retinal eccentricities (1° to 10° in 0.5° steps). Images did not simulate rod photoreceptors or retinal blood vessels. A total of 1,900 images, 480 × 300 pixels in size, were used. One pixel represented one micron on the retinal surface and each image contained a different pattern of photoreceptor reflectivity. To accurately represent real AOSLO images, noise and/or residual aberrations were added to all synthetic images (as specified for each experiment).

Noise introduced at any point in the system was considered as a constant in its characteristics over time. Gaussian distributed white noise (randomly generated for each image) was added with the same characteristics as the noise measured through the Oxford AOSLO. By considering noise as a constant, differences in signal to noise ratio between images will be related to the signal present in each image. To vary the signal to noise ratio of images, the signal, measured in nanowatts (nW) at the pinhole, was varied in the simulation. Information about several uncontrolled

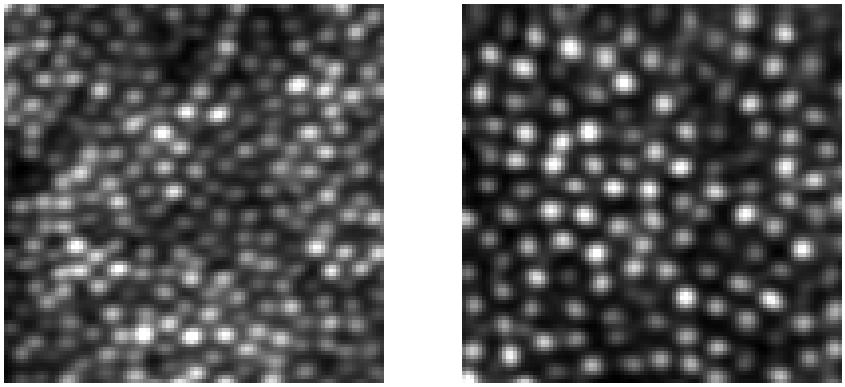
variables (such as laser power at the cornea, detector gain and temperature, although the detector temperature is controlled) is required in order to accurately derive power at the detector (signal) in nanowatts. In the simulation, these variables were specified based on measurements from the Oxford AOSLO. In order to facilitate the interpretation of this work for readers using a different AOSLO system design, the signal of synthetic images has been specified in both absolute (nanowatts) and relative (signal to noise ratio) terms.

Wavefront aberrations were obtained from a database consisting of a continuous 3.4 second time sequence (without blinks) of 50 healthy eyes, measured with a Shack-Hartmann aberrometer at a temporal frequency of 236 Hz [135]. These measurements were performed in a dark room with natural pupil dilation and no cycloplegia. The mean Zernike coefficient was calculated across the time series for each eye and for all 65 of the Zernike modes that had been measured (excluding piston). To simulate residual aberration a wavefront composed of a proportion of the mean Zernike coefficients was generated and its point spread function was used in the rendering of the synthetic images.

Datasets could be produced with (a) noise alone, (b) residual aberration alone or (c) noise and residual aberration.

5.3.2.2 Dataset 4 - real data (Oxford)

This dataset consisted of a total of 17 images 100 × 100 pixels in size from 7 normal participants (Figure 5.6). Images were obtained from the previously described Oxford AOSLO [26] with a 1.0° x 1.6° field of view, and were collected in collaboration with Sarah Regan and Anna-Katharina Hauperich from the Oxford Perception Lab, University of Oxford. Two eccentricities (3° and 6°) in the temporal and superior meridians from fixation were imaged and sufficient data were collected to produce one averaged image per location. Images at 3° in the temporal meridian were obtained in 4 participants, images at 3° in the superior meridian were obtained in 6 participants, images at 6° in the temporal meridian were obtained in 4 participants and images at 6° in the superior meridian were obtained in 3 participants. A 100



(a) 3° temporal to the fovea **(b)** 6° temporal to the fovea

Figure 5.6: Example confocal AOSLO images of one participant from Dataset 4, obtained using the Oxford AOSLO.

× 100 pixel region of interest was manually selected from each image. Cones were manually identified in all images by an experienced grader. The University of Oxford Central University Research Ethics Committee approval was obtained for AOSLO imaging of normal participants (reference R57315) with written participant consent.

5.3.2.3 Dataset 5 - real data (Milwaukee)

This publicly available dataset of 840 confocal AOSLO images 150 × 150 pixels in size from 21 participants (20 normal, one deutranopic) has been previously used for the development and validation of cone photoreceptor segmentation and identification algorithms [67, 142]. In this dataset, four retinal locations 0.65° from fixation were imaged for each participant (supero-temporal, supero-nasal, infero-nasal, infero-temporal) and 10 averaged images were acquired per location. The size of each 150 × 150 pixel image on the retina ranged from 58 µm^2 to 71 µm^2 (Figure 5.7). Axial length measurements were used to calculate the lateral resolution of the images. All images were accompanied by semi-automatically marked annotations of the cone photoreceptors performed by an expert grader.

5.3.2.4 Input images

The CNN inputs are images, which were selected from one of the three aforementioned datasets (as specified for each experiment). All input images were cropped to 96 × 96 pixels in size. They were divided between training, validation

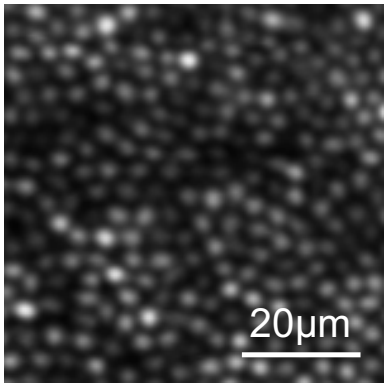


Figure 5.7: Example confocal AOSLO image from Dataset 5 at 0.65° from the fovea.

and testing sets in the ratio 60:20:20. Where noise and/or residual aberration was added or where images were of different retinal eccentricities, all data were stratified to ensure even representation between the training, validation and testing data sets. All synthetic image inputs were non-overlapping and real image inputs contained overlap of between 0% and 50%. The pixel intensities in each input image were normalised by subtracting the mean pixel intensity and then dividing by the standard deviation of pixel intensities.

5.3.2.5 Training labels

Ground truth data provides a list of image coordinates associated with cone centres from the corresponding input images. To convert this list to training labels that can be used with the CNN, each map of the pixel locations with the same pixel sampling as the image data was convolved with a Gaussian kernel to produce ground truth Gaussian annotated training labels (Figure 5.8). The Gaussian annotations with a standard deviation equal to the cone sigma were truncated to a diameter of one cone size. The cone sigma was determined from the corresponding input image as described in Section 5.3.1.1. Maximum intensity projection was used to eliminate the problem of overlapping Gaussians (Figure 5.8C). Ground truth Gaussian annotated training labels were normalised so that intensity values stretched between 0 and 255. Without performing this operation, most of the pixels within the Gaussian annotated ground truth belong to the background (the space between cone photoreceptors) and are labelled as zero by the CNN.

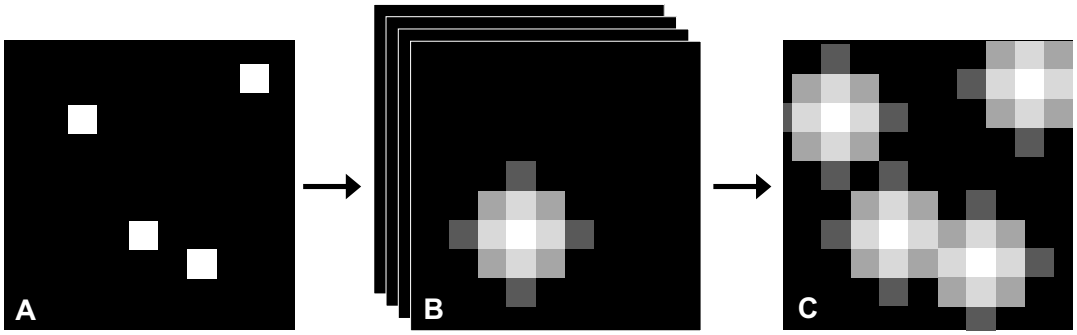


Figure 5.8: Ground truth Gaussian annotated training labels. **A** - map of the ground truth cone centres where each cone centre is represented by a single pixel. **B** - each pixel from the map of cone centre locations is separated (while maintaining its spatial position in the XY plane) to create a Z stack and these were subsequently convolved with a Gaussian kernel. **C** - Maximum intensity projection (projecting the pixel with the maximum intensity) is used through the Z stack onto a 2D image to eliminate the problem of overlapping Gaussians.

5.4 Experiment 1

The aim of Experiment 1 was to investigate two CNN architectures (FCRN and U-Net) and to assess their suitability for automated cone photoreceptor identification from synthetic confocal AOSLO images.

5.4.1 Input images

Synthetic images from Dataset 3 (described in Section 5.3.2.1), which included the effects of noise, were used. Gaussian distributed white noise along with the effects of diffraction (with a 5 mm pupil diameter) was added to the synthetic data and the image signal was varied from 0.04 nW to 0.16 nW in steps of 0.04 nW to create data with four signal to noise ratios (Figure 5.9). A total of 91,200 input images 96 $\times$ 96 pixels in size were split 60:20:20 between training, validation and testing.

5.4.2 FCRN and U-Net architectures

In general, a CNN accepts an input image and transforms it through a series of mathematical operations in order to extract features and produce an output. During the training process, the network does not have control over the input data, which is given and fixed, but the optimal trainable parameters (weights and biases) can

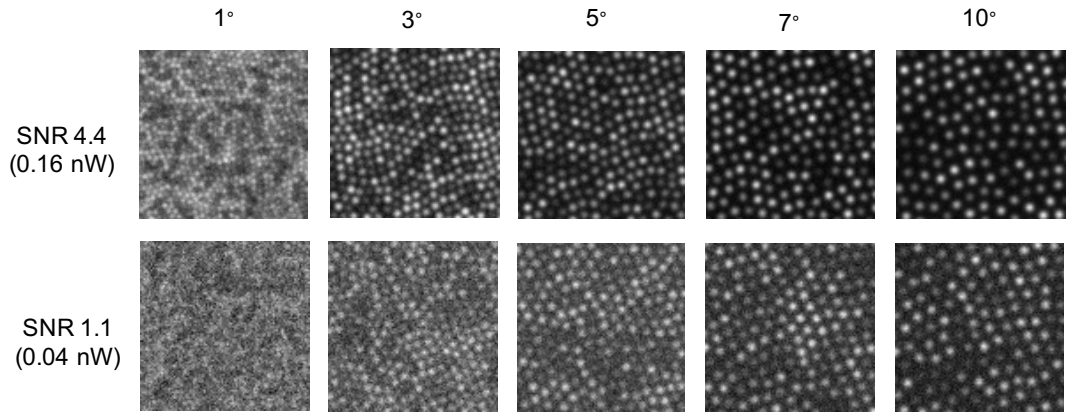


Figure 5.9: An example of the synthetic cone photoreceptor mosaic images from 1° to 10° used during Experiment 1. The image signal to noise ratio (SNR) is provided with image signal in parentheses, measured in nanowatts (nW) at the pinhole. The synthetic images assume perfect adaptive optics correction of wavefront aberrations but include the effects of diffraction (with a 5 mm pupil diameter) and noise.

be learned such that they will minimise the error between the computed output and the ground truth labels in the training data. The error between the computed outputs and the ground truth labels is measured using a loss function, which is used to estimate the model's performance at each stage of the training process. Examples of loss functions include mean squared error, mean absolute error and Huber loss (a loss function that is less sensitive to outliers than mean squared error). The loss function can be thought of as a high-dimensional non-convex (many local minima in the search space) landscape, whose minimum will correspond to a set of parameters (weights and biases) that minimise the error between a model's outputs and the ground truth labels. The model weights are randomly initialised at the beginning of the training process. Deep learning can be thought of as an optimisation problem and an artificial neural network as a mathematical function that presents a solution to the problem. During training the artificial neural network is attempting to minimise the loss function, which measures its performance on a given dataset. The goal of the training process is to find a set of parameters which minimises the loss function. This is achieved by computing the gradient of the loss function (using an optimisation algorithm) which informs adjustment of the model's parameters. *Adam* [157], derived from adaptive moment estimation,

is an adaptive learning rate optimisation algorithm that computes an individual learning rate for each parameter and is used to update model parameters from training data. Unlike model parameters, hyperparameters are external to the model and are set before the learning process begins i.e. they cannot be estimated from the data. The learning rate is a hyperparameter of the optimisation algorithm that determines the size of the step change in the model's parameters at each iteration and thus controls how fast the model learns.

During training, input images from the training dataset are propagated forward through the network and transformed through a series of mathematical operations (convolutions, non-linear operations and pooling) to produce an output. Next, the error between the model's outputs and the ground truth labels is calculated using the loss function. The error gradient is propagated backwards through the network to update the model's parameters and minimise the error. The batch size is a hyperparameter that represents the number of images from the training dataset used to determine the error gradient before updating the model's parameters.

The training and validation datasets are used during the training process. The training dataset is used to fit the model's parameters, while the validation dataset, which is independent of the training dataset, is used for hyperparameter selection and to avoid overfitting of the model. Training with only the training dataset is likely to lead to model overfitting and thus poor performance when evaluating the model with the test dataset, which is crucial for testing the model's generalisability and determining its accuracy. Cross-validation is a widely used data resampling method to assess a model's generalisability and to prevent overfitting. K-fold cross-validation is a cross-validation technique that results in a less biased estimate of the model's performance. The data to be used during the training process is randomly split into k folds (groups). The model is trained using $k-1$ folds while the remaining k^{th} fold is used for validation. The process is repeated until every fold has served as the validation dataset. The model's performance is averaged over the k folds. For example, with four-fold cross-validation the data is split into four folds and each fold is used once as the validation dataset while the other folds

are used as the training dataset. The training process is repeated four times and the model is evaluated using the average performance.

To reduce training time and to help prevent overfitting of the network, batch normalisation [158] can be used. Despite its popularity, the exact mechanisms that underlie the effectiveness of batch normalisation are not fully understood [158, 159]. It has been proposed to reduce internal covariate shift [158], a phenomenon that occurs during the training stage of an artificial neural network - as the parameters of the preceding layers are updated this changes the distribution of inputs to the current layer. This change leads to a continual shift of the underlying training problem that is thought to have a deleterious effect on the training process [159]. In deep neural networks this problem will be amplified as small changes are propagated through the network, leading to a substantial change in the input distribution of deeper layers. Batch normalisation aims to stabilise the distribution of inputs to a given layer by setting their mean and standard deviation to zero and one, respectively, and typically adds two trainable parameters to scale and shift the normalised inputs in order to preserve model expressivity [159]. A model's expressivity characterises the complexity of functions that it can compute.

Pooling layers are commonly inserted between successive convolutional layers in a CNN to progressively reduce the spatial dimensions of the feature map and the number of parameters in the network, preventing overfitting. A pooling layer with a filter size of 2×2 pixels and a stride of 2 will reduce the size of the input by a factor of two. Stride, which is usually symmetrical in height and width dimensions, is a parameter that controls how far a filter moves from one position to the next as it is convolved with an input. For example, a stride of 2 will move the filter 2 pixels down and right for each vertical and horizontal movement of the filter, respectively. Max pooling uses the max operation (returns the maximum value of a set of elements), which, with a filter size of 2×2 pixels, would take the maximum of 4 values.

The FCRN and U-Net model architectures tested in this experiment are illustrated in Figure 5.10 and Figure 5.11, respectively. The FCRN model architecture was based on the FCRN-A network described by Xie *et al.* [152] and contained

approximately 1.4 million trainable parameters with a kernel size of 5 × 5 pixels. Two different loss functions (mean absolute error and mean squared error) were tested. The U-Net model architecture was based on the network described by Ronneberger *et al.* [153] and contained approximately 2.7 million trainable parameters with a kernel size of 5 × 5 pixels. Three different loss functions (mean absolute error, mean squared error and Huber loss) were tested. For both CNNs, weights were initialised with a random orthogonal matrix and *Adam* was used for optimisation with a learning rate of 0.001, which was shown to be a good default setting by Kingma and Ba [157]. Using a random orthogonal matrix to initialise weights has been shown to improve the speed and stability of model convergence (minimisation of the loss function) and the backward propagation of error gradients [160, 161]. In order to improve the stability and performance of the networks and ensure they did not overfit the training data, batch normalisation was applied prior to the activation functions. Padding with zeros was used such that the output has the same size as the original input and the CNNs were trained with a batch size of 32 in 100 epochs and with four-fold cross-validation. The number of epochs is a hyperparameter that specifies the number of times the dataset will pass through the network to train the model. One epoch means the entire training dataset has passed through the network (and has had an opportunity to update the model's parameters) once.

5.4.3 Results

5.4.3.1 Fully Convolutional Regression Network (FCRN)

Training required approximately 75 epochs to minimise validation loss with mean absolute error as the loss function (Figure 5.12a), and approximately 50 epochs to minimise validation loss with mean squared error as the loss function (Figure 5.12b). Mean squared error (mean Jaccard similarity coefficient 0.977, standard deviation 0.004) outperformed mean absolute error (mean Jaccard similarity coefficient 0.957, standard deviation 0.003) as a loss function. Overall, accuracy was higher (greater true positives and fewer false positives) at higher retinal eccentricities (lower cone number) (Figure 5.13a and 5.13c). At lower retinal eccentricities (higher cone

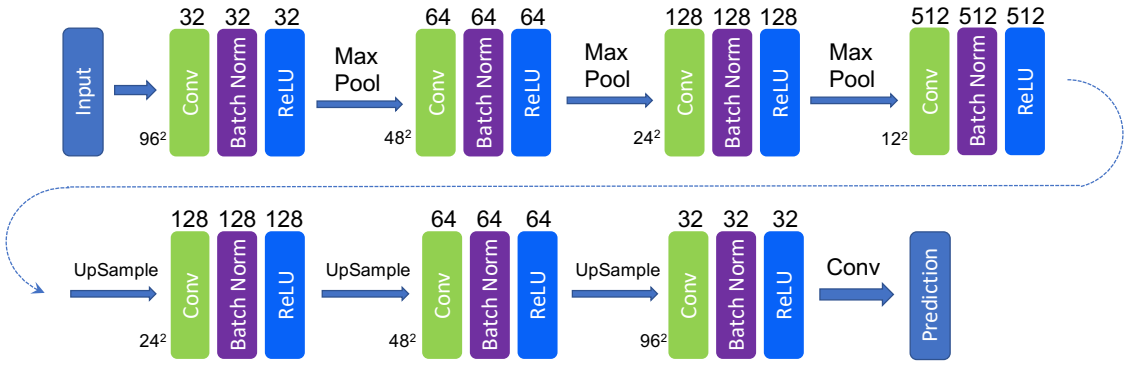


Figure 5.10: Fully Convolutional Regression Network (FCRN) model architecture. Inputs and predictions are 96 × 96 pixels in size. Each group of adjacent boxes corresponds to a single layer and the layer input size in pixels is provided at the lower left edge. The number of filters is provided on top of each box. Each box denotes a different operation. The kernel size for convolutional layers is set to 5 × 5. The max pooling window is set to 2 × 2 with a stride of 2. Batch Norm - batch normalisation, Conv - convolutional layer, Max Pool - max pooling, ReLU - Rectified Linear Unit.

number) the model performed better on input images with a higher signal to noise ratio (Figure 5.13b and 5.13d).

5.4.3.2 U-Net

Training required approximately 75 epochs to minimise validation loss with mean absolute error as the loss function (Figure 5.14a), approximately 50 epochs to minimise validation loss with mean squared error as the loss function (Figure 5.14b), and approximately 85 epochs to minimise validation loss with Huber loss as the loss function (Figure 5.14c). Mean squared error (mean Jaccard similarity coefficient 0.994, standard deviation 0.001) outperformed mean absolute error (mean Jaccard similarity coefficient 0.991, standard deviation 0.001) and marginally outperformed Huber loss (mean Jaccard similarity coefficient 0.993, standard deviation 0.001) as a loss function. Overall, accuracy was higher (greater true positives and fewer false positives) at higher retinal eccentricities (lower cone number) (Figure 5.15(a), (c) and (e)). At lower retinal eccentricities (higher cone number) the model performed better on input images with a higher signal to noise ratio (Figure 5.15(b), (d) and (f)).

In Experiment 1 the validation dataset was used to avoid overfitting of the models during training with the training dataset. The U-Net (mean Jaccard

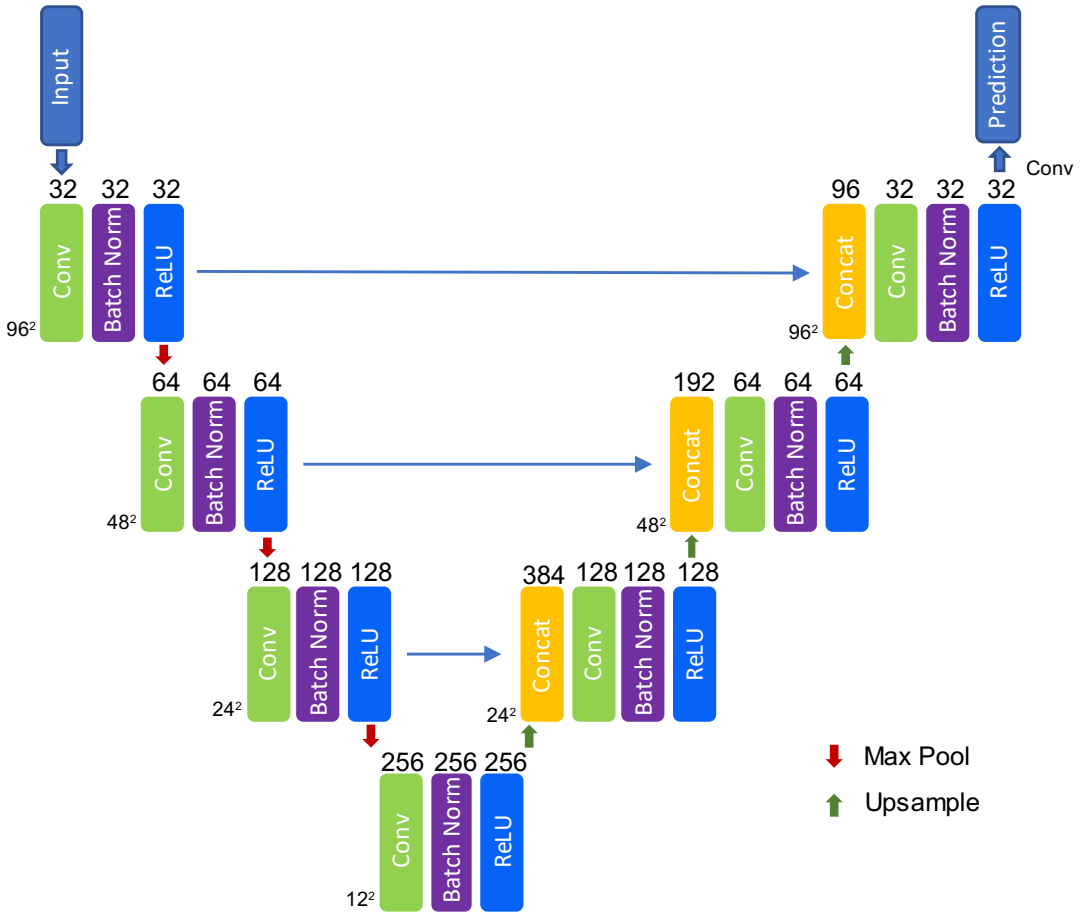
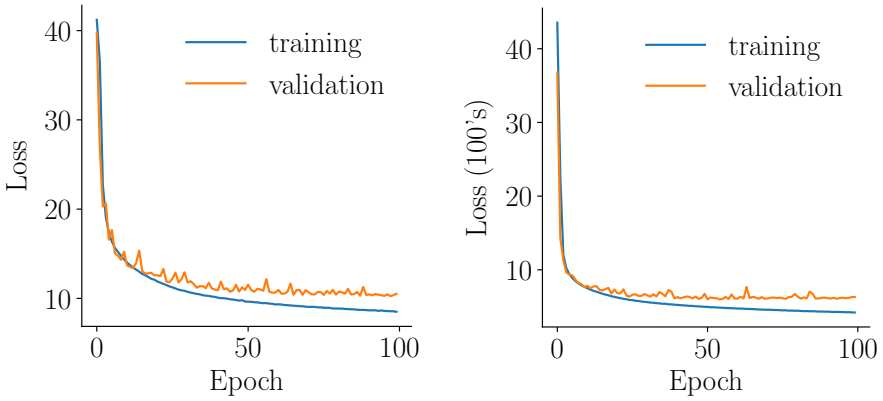


Figure 5.11: U-Net model architecture. Inputs and predictions are 96 × 96 pixels in size. Each group of adjacent boxes corresponds to a single layer and the layer input size in pixels is provided at the lower left edge. The number of filters is provided on top of each box. The horizontal blue arrows represent higher resolution feature maps from the contracting path that are concatenated with upsampled features from the expansive path. Each box denotes a different operation. The kernel size for convolutional layers is set to 5 × 5. The max pooling window is set to 2 × 2 with a stride of 2. Batch Norm - batch normalisation, Concat - concatenation, Conv - convolutional layer, Max Pool - max pooling, ReLU - Rectified Linear Unit.

similarity coefficient 0.994) consistently outperformed the FCRN (mean Jaccard similarity coefficient 0.977) for automated cone photoreceptor identification from synthetic confocal AOSLO images. Additionally, mean squared error outperformed Huber loss and mean absolute error as a loss function with the synthetic data used. Following Experiment 1 and on the basis of results from the test set from Dataset 3, further investigation of the U-Net model architecture with mean squared error as a loss function was undertaken.

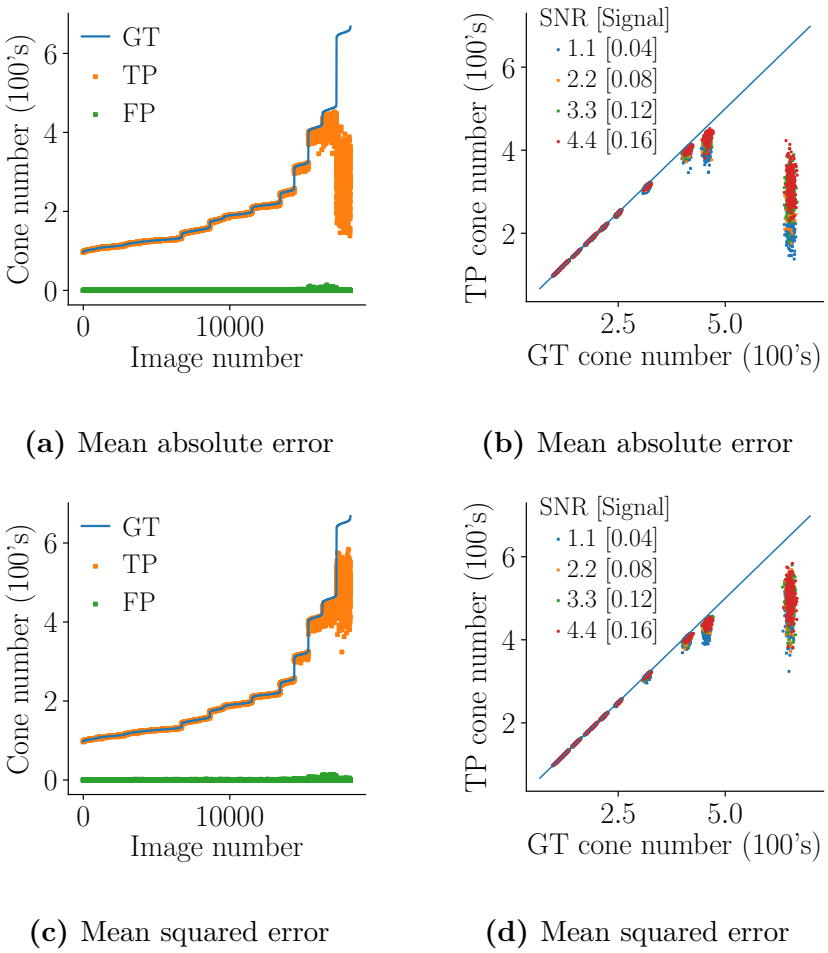


(a) Mean absolute error **(b)** Mean squared error

Figure 5.12: Fully Convolutional Regression Network (FCRN) training losses from one cross-validation fold are shown. Each figure illustrates the model's training and validation loss with a different loss function. Both loss functions show a smooth loss profile with no evidence of overfitting.

5.5 Experiment 2

The aim of Experiment 2 was to investigate the optimal kernel size for the U-Net model architecture described in Section 5.4.2. The dataset described in Section 5.4.1 was also used for Experiment 2. Weights were initialised with a random orthogonal matrix and *Adam* was used for optimisation with an initial learning rate of 0.001. The learning rate was reduced by a factor of 5 every 10 epochs. This learning rate schedule was selected based on experimental results aiming to investigate optimal values for this hyperparameter. Experiments investigated an initial learning rate of 0.001 with three learning rate schedules: learning rate reduced by a factor of 5 every 5 epochs, learning rate reduced by a factor of 5 every 10 epochs and learning rate reduced by a factor of 10 every 10 epochs. The validation dataset was used to avoid overfitting of the models during training with the training dataset and the optimal learning rate schedule was selected on the basis of results from the test set from Dataset 3. The models were trained with a batch size of 32 in 100 epochs and with four-fold cross-validation. Five kernel sizes were tested (3, 5, 7, 9 and 11 pixels).



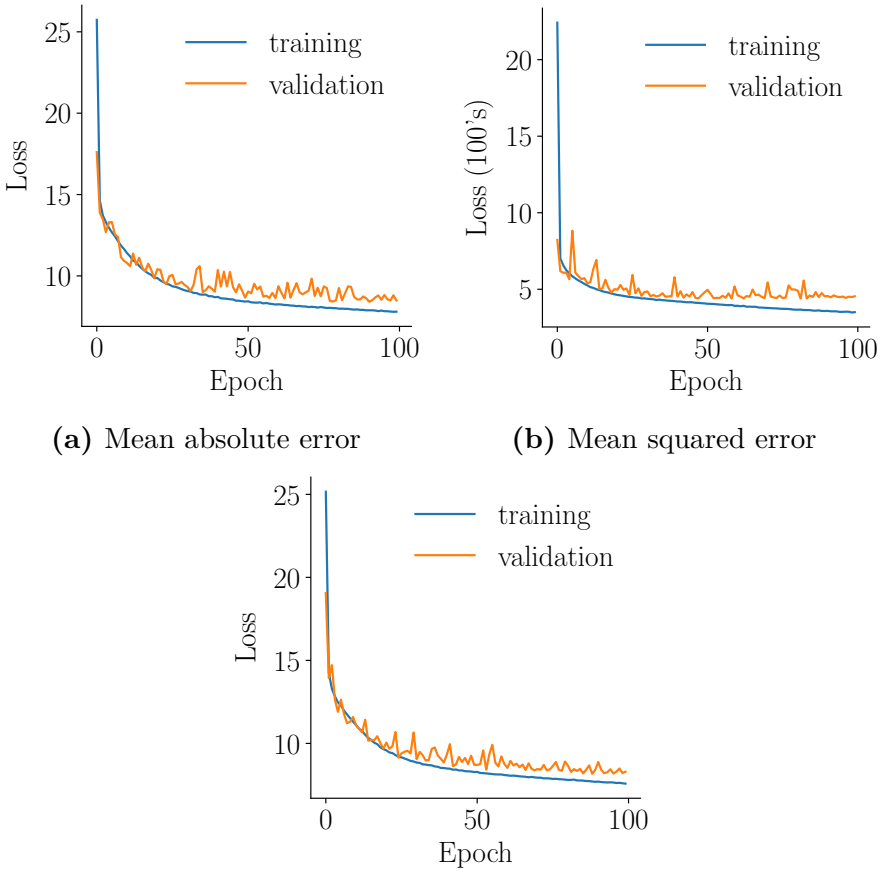
(a) Mean absolute error **(b)** Mean absolute error

(c) Mean squared error **(d)** Mean squared error

Figure 5.13: Fully Convolutional Regression Network (FCRN) accuracy. Results for each loss function tested from one cross-validation fold are shown. **(a)** and **(c)** illustrate the number of true positive (TP) and false positive (FP) cones identified relative to the ground truth (GT). **(b)** and **(d)** illustrate model accuracy (TP) stratified by input image signal to noise ratio (SNR) with the image signal in parentheses, measured in nanowatts (nW) at the pinhole.

5.5.1 Results

Training required between 50 - 60 epochs to minimise validation loss. Table 5.1 shows the results of the trained models when run on test data. The best results were achieved with a kernel size of 7 pixels (see Table 5.1). Accuracy remained higher (greater true positives and fewer false positives) at higher retinal eccentricities (lower cone number) (Figure 5.16) and at lower retinal eccentricities (higher cone number) the model performed better on input images with a higher signal to noise ratio (Figure 5.17). In Experiment 2 the validation dataset was used to avoid overfitting



(a) Mean absolute error

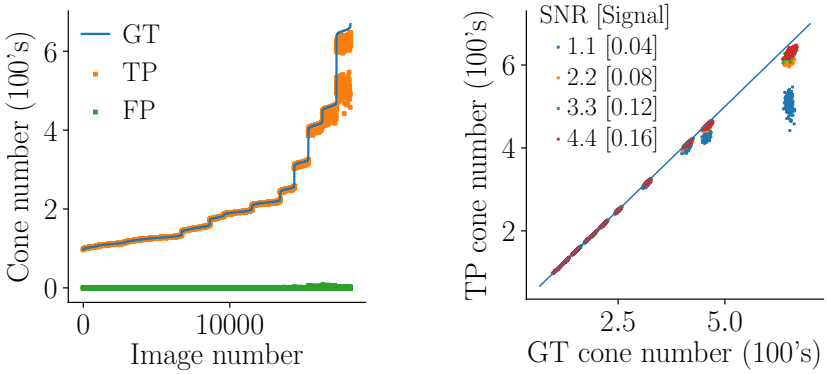
(b) Mean squared error

(c) Huber loss

Figure 5.14: U-Net training losses from one cross-validation fold are shown. Each figure illustrates the model's training and validation loss with a different loss function. All loss functions show a smooth loss profile with no evidence of overfitting.

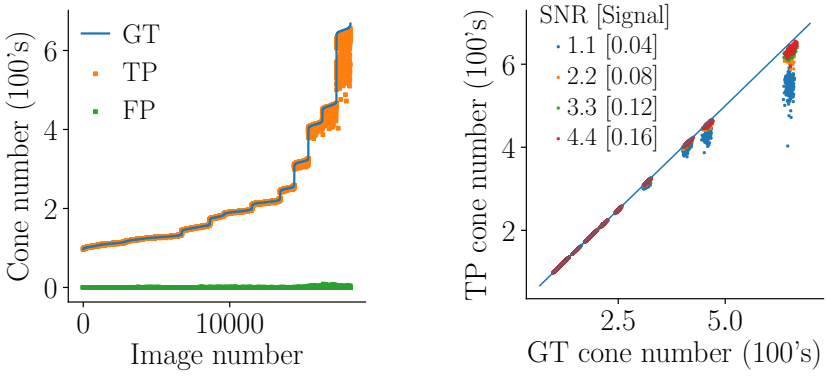
of the models during training with the training dataset and kernel size (see below) was selected on the basis of results from the test set from Dataset 3.

The results of Experiment 2 were further evaluated by comparing kernel size with cone size from the synthetic dataset used to train and evaluate the U-Net model. The size of an average synthetic cone ranged between 4.6 pixels and 7.2 pixels (Figure 5.18), where cone size was estimated by calculating 6 × the cone sigma (see Section 5.3.1.1). The best results from Experiment 2 were achieved with a kernel size of 7 pixels and the worst results (with a marginal reduction in accuracy) were achieved with a kernel size of 3 or 11 pixels. Based on the results of Experiment 2 and the average synthetic cone sizes, a kernel size of between 5 and 9 pixels is appropriate for the U-Net model. The results of Experiment 2



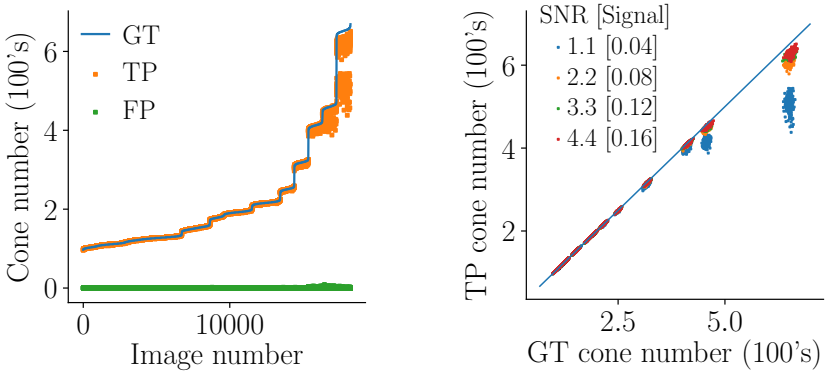
(a) Mean absolute error

(b) Mean absolute error



(c) Mean squared error

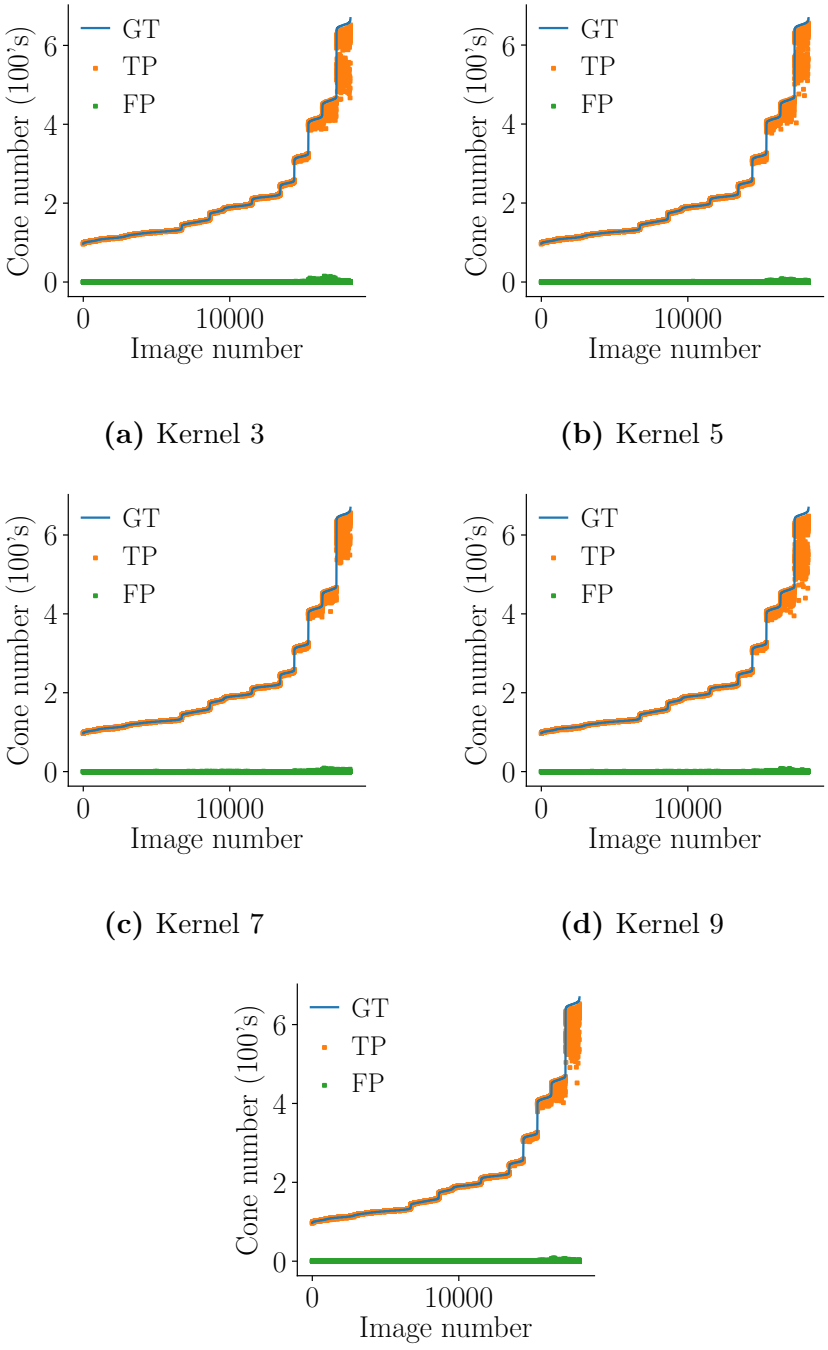
(d) Mean squared error



(e) Huber loss

(f) Huber loss

Figure 5.15: U-Net accuracy. Results for each loss function tested from one cross-validation fold are shown. **(a)**, **(c)** and **(e)** illustrate the number of true positive (TP) and false positive (FP) cones identified relative to the ground truth (GT). **(b)**, **(d)** and **(f)** illustrate model accuracy (TP) stratified by input image signal to noise ratio (SNR) with the image signal in parentheses, measured in nanowatts (nW) at the pinhole.



(a) Kernel 3

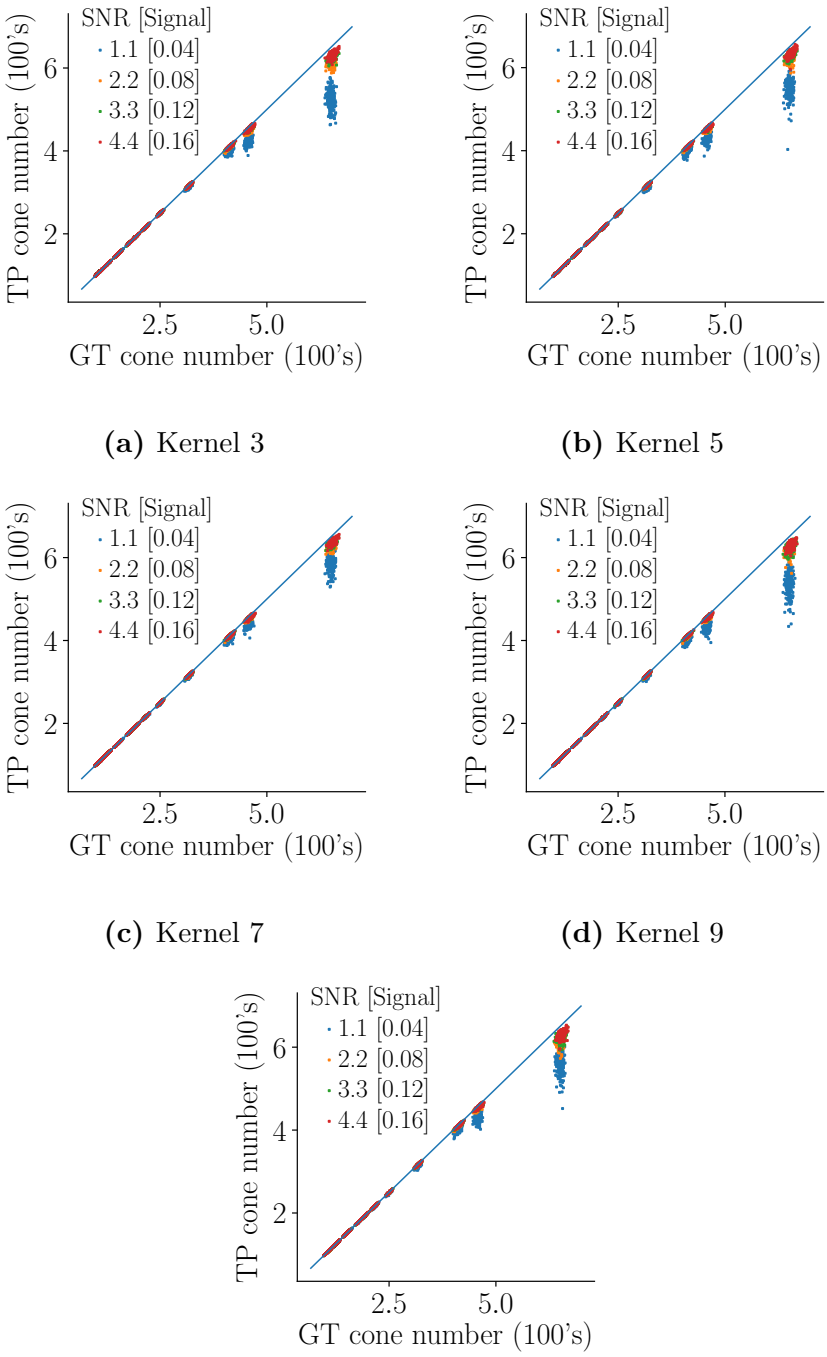
(b) Kernel 5

(c) Kernel 7

(d) Kernel 9

(e) Kernel 11

Figure 5.16: U-Net accuracy. Results for each kernel size tested from one cross-validation fold are shown. All figures illustrate the number of true positive (TP) and false positive (FP) cones identified relative to the ground truth (GT).



(a) Kernel 3

(b) Kernel 5

(c) Kernel 7

(d) Kernel 9

(e) Kernel 11

Figure 5.17: U-Net accuracy stratified by input image signal to noise ratio. Results for each kernel size tested from one cross-validation fold are shown. All figures illustrate model accuracy (TP - true positive) stratified by input image signal to noise ratio (SNR) with the image signal in parentheses, measured in nanowatts (nW) at the pinhole. GT - ground truth.

Table 5.1: Mean Jaccard similarity coefficients of all cross-validation folds for each kernel size tested with the U-Net.

	Kernel				
	3	**5**	**7**	**9**	**11**
μ	0.9926	0.9948	0.9952	0.9946	0.9944
σ	0.0003	0.0003	0.0004	0.0007	0.0012

μ - mean, σ - standard deviation.

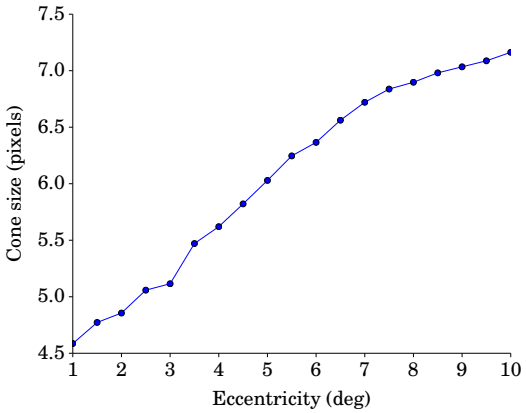


Figure 5.18: Average cone size from synthetic input images (Dataset 3) for all retinal eccentricities (1° to 10°). Cone size was estimated to further evaluate experimental results investigating the optimal kernel size (Experiment 2).

were also evaluated by comparing kernel size with retinal eccentricity from the synthetic dataset used to test the model. There was only a marginal variation in the Jaccard similarity coefficient between different kernel sizes for images from low retinal eccentricities (between 1° and 2°) and no difference at higher retinal eccentricities. Although one would expect a relationship between kernel size and retinal eccentricity (as retinal eccentricity increases, cone size and spacing between cones increases) it is possible that cone size across the range of retinal eccentricities used in this study did not vary enough to benefit from increasing kernel sizes.

5.6 Experiment 3

Artefacts, which are inherent to AOSLO imaging, include noise and residual aberration. Experiments 1 and 2 were conducted with synthetic data where only

the signal to noise ratio of AOSLO images was considered. The aim of Experiment 3 was to investigate the U-Net architecture with synthetic confocal AOSLO images that include noise and/or residual aberration, and with two real image datasets.

5.6.1 Input images

5.6.1.1 Synthetic images

Three synthetic datasets of averaged images (derived from Dataset 3, described in Section 5.3.2.1) that included noise, residual aberration and the interaction between noise and residual aberration were used.

Dataset 3a included noise along with the effects of diffraction. Gaussian distributed white noise along with the effects of diffraction (with a 5 mm pupil diameter) was added to the synthetic data and the image signal was varied from 0.2 nW to 0.35 nW in steps of 0.05 nW. Averaged images produced by averaging 10, 20 and 30 single frames were simulated. By combining 1,900 images from Dataset 3 with the variations in image signal (n = 4) and image averaging (n = 3), 22,800 images, 480 × 300 pixels in size, were generated. Each image was then split into 12 non-overlapping sub-images to create a total of 273,600 input images, 96 × 96 pixels in size, for Dataset 3a.

Dataset 3b included residual aberration. As described in Section 5.3.2.1 wavefront aberrations were obtained from a database consisting of measured aberrations from 50 healthy eyes [135]. Four eyes with the lowest wavefront aberrations were selected. To simulate residual aberration a wavefront composed of a proportion (30%, 20% and 10%) of the mean Zernike coefficients was generated and its point spread function was used to render 1,900 synthetic images from Dataset 3. A total of 22,800 images, 480 × 300 pixels in size, were generated and each image was then split into 12 non-overlapping sub-images to create a total of 273,600 input images, 96 × 96 pixels in size, for Dataset 3b.

A total of 547,200 input images, 96 × 96 pixels in size, containing noise (Dataset 3a) or residual aberration (Dataset 3b) were generated and split 60:20:20 between training, validation and testing (Figure 5.19).

Dataset 3c included the interaction between noise and residual aberration. Ten eyes with the lowest wavefront aberrations were selected. To simulate residual aberration a wavefront composed of 10% of the mean Zernike coefficients was generated and its point spread function convolved with the synthetic images. Following this, Gaussian distributed white noise was added to the synthetic data with an image signal of 0.3 nW. Averaged images produced by averaging 20 single frames were simulated. Residual aberration from 10 eyes and noise were added to 1,900 images from Dataset 3 (Figure 5.19) to generate 19,000 images, 480 × 300 pixels in size. Each image was then split into 12 non-overlapping sub-images to create a total of 228,000 input images, 96 × 96 pixels in size, that included the interaction between noise and residual aberration. These input images were split 60:20:20 between training, validation and testing.

5.6.1.2 Real images

The U-Net model was also evaluated on Dataset 4 (described in Section 5.3.2.2, collected using the Oxford AOSLO), and test images from Datasets 5a and 5b (described in Section 5.7.1). Test Datasets 5a and 5b each comprised 160 images from Dataset 5 (described in Section 5.3.2.3), a publicly available dataset of 840 confocal AOSLO images. In Dataset 5a images were resized to represent a micron to pixel ratio of 1:1 and zero padding was used to ensure input images remained 96 × 96 pixels in size (Figure 5.20A). In Dataset 5b images were cropped to a size of 96 × 96 pixels and represented a micron to pixel ratio between 0.4:1 and 0.5:1 (Figure 5.20B).

5.6.2 New U-Net architecture

Cone size (in microns) increases with retinal eccentricity, and so the measured cone size (in pixels) also increases as shown in Figure 5.18. However, the cone size in pixels will also increase for images with a smaller pixel scale (in microns). In order to address this challenge a new U-Net architecture was designed, based on the previous U-Net architecture, to incorporate information from features of different sizes at each level (Figure 5.21). The new architecture was inspired by

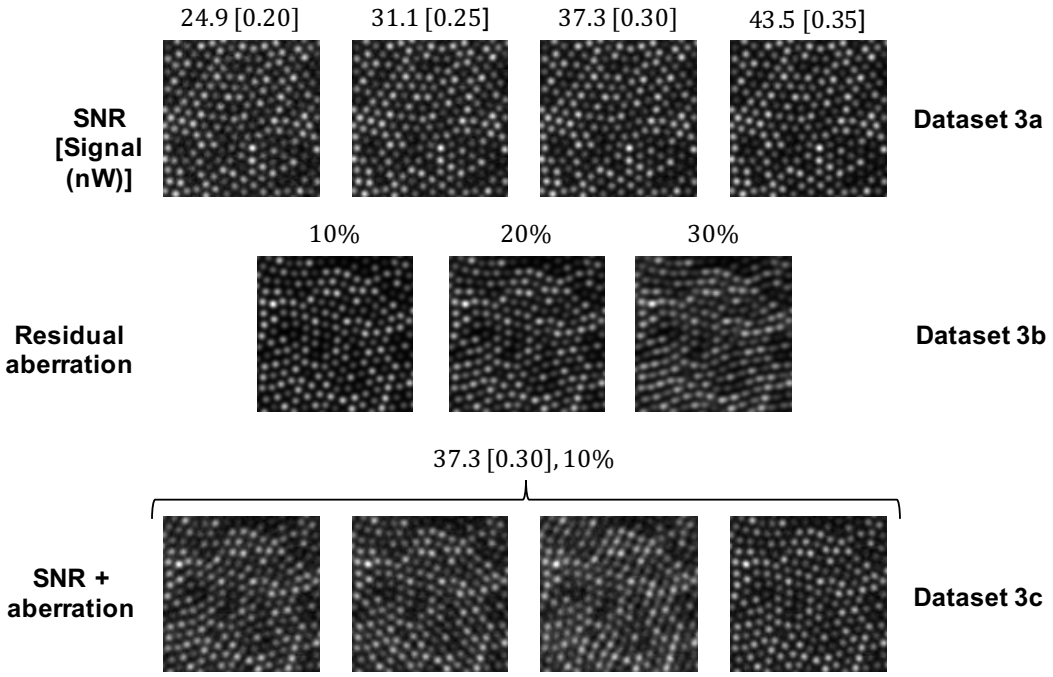


Figure 5.19: Examples of the averaged synthetic input images (from 5°) used for training and evaluating the new U-Net model. **Top row** - synthetic input images that included noise and the effects of diffraction (Dataset 3a). Each averaged image (simulated by averaging 20 single frames) is labelled with its signal to noise ratio (SNR) and the image signal in parentheses, measured in nanowatts (nW) at the pinhole. **Middle row** - synthetic input images that included residual aberration (Dataset 3b). Each image is labelled with the percentage of the mean Zernike coefficients that were applied to simulate residual aberration. **Bottom row** - synthetic input images that included the interaction between noise and residual aberration (Dataset 3c), where each averaged image has a SNR of 37.3 (image signal of 0.3 nW) and 10% of the mean Zernike coefficients were applied to simulate residual aberration.

the Inception networks [146] with multiple sized kernels operating on the same level. Deep neural networks are difficult to train because it is hard to pass gradient updates throughout the entire network. As the gradient is back-propagated to earlier layers, each of the network's weights receives an update. However, repeated multiplication may make the gradient extremely small (effectively preventing the weights from changing value) or large (impairing the neural network's performance). Shortcut connections, inspired by the residual networks [145] were implemented in the new U-Net architecture (Figure 5.21B) to overcome the problem of gradient back-propagation. These connections provide a highway for gradient back-propagation along with the added benefit of ensuring a deeper network is at least as effective as

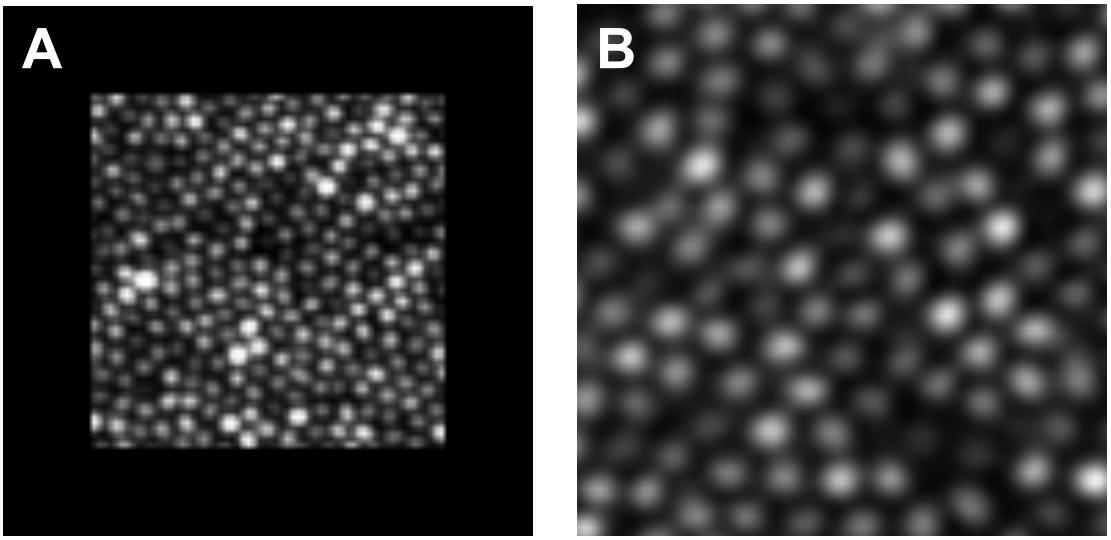
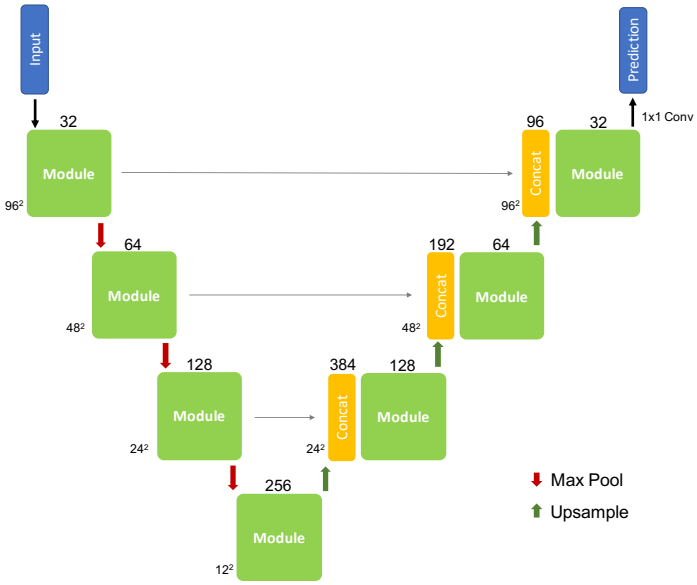


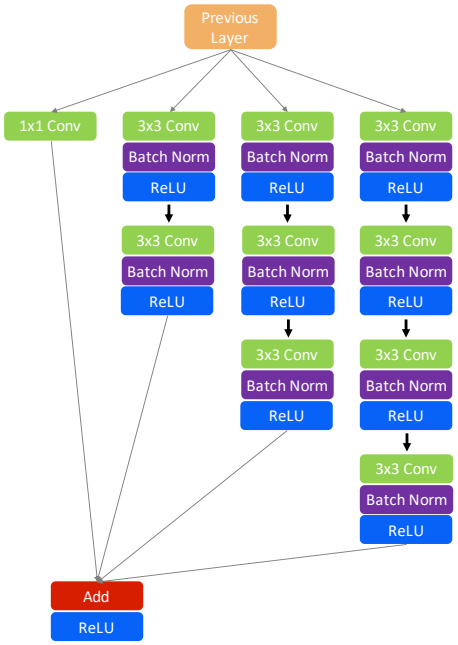
Figure 5.20: Example of real input images used for transfer learning and evaluation of the new U-Net model. **A** - example image from Dataset 5a. The images from this dataset have been resized to represent a micron to pixel ratio of 1:1. Zero padding was used to maintain an input image size of 96 × 96 pixels. **B** - example image from Dataset 5b. Images from this dataset were cropped to a size of 96 × 96 pixels and represented a micron to pixel ratio of between 0.4:1 and 0.5:1. The presented image represents a micron to pixel ratio of 0.4:1.

its shallower counterpart by allowing every additional layer to contain the identity function (where input equals output) as one of its elements. A convolutional layer with a kernel size of 1 × 1 pixels was added to the shortcut connection such that its output is the same shape as the output from the convolutional blocks, allowing matrix addition. Changes to the size of the new U-Net architecture were not investigated as a hyperparameter due to the favourable results achieved with the current architecture and due to the additional computational cost (training with Datasets 3a and 3b took 7 days).

The new U-Net architecture contained approximately 8.9 million trainable parameters. Mean squared error was used as a loss function and weights were initialised with a random orthogonal matrix. *Adam* was used for optimisation with an initial learning rate of 0.001, which was reduced by a factor of two if no reduction in validation loss greater than or equal to 1 occurred after five epochs, to a minimum learning rate of 10^{-5}. The learning rate schedule was selected following qualitative analysis of the validation losses from the trained models, with the aim of



(a) Overview of the new U-Net model architecture



(b) New U-Net module

Figure 5.21: The new U-Net model architecture. Inputs and predictions are 96 × 96 pixels in size. **(a)** - Overview of the new U-Net model architecture. The green boxes represent the module architecture shown in (b). The input size in pixels and the number of filters are provided at the lower left edge an on top of each box, respectively. The horizontal grey arrows represent higher resolution feature maps from the contracting path that are concatenated with upsampled features from the expansive path. The max pooling window is set to 2 × 2 with a stride of 2. **(b)** - The module architecture illustrated by the green boxes in (a). Add - matrix addition, Batch Norm - batch normalisation, Concat - concatenation, Conv - convolutional layer (with kernel size specified), Max Pool - max pooling, ReLU - Rectified Linear Unit.

minimising the validation loss. To maintain computational efficiency and increase feature representation of the input by stacking convolutional layers containing non-linearities (making features more expressive), a kernel size of 3 × 3 pixels was used with effective receptive fields of 5 × 5, 7 × 7 and 9 × 9 pixels, that were determined following Experiment 2. The receptive field in a CNN is the region of an image that is selected as an input to the kernel of a convolutional layer and increases linearly as convolutional layers are stacked together. For example, stacking two convolutional layers, each with a kernel size of 3 × 3 pixels, would result in an effective receptive field of 5 × 5 pixels and stacking an additional layer would result in an effective receptive field of 7 × 7 pixels. In order to improve the stability and performance of the network and ensure it did not overfit the training data, batch normalisation was applied prior to the activation functions. Padding with zeros was used such that the output has the same size as the original input and the networks were trained with a batch size of 32 with four-fold cross-validation. Early stopping was implemented to stop training of the model if no reduction in validation loss greater than or equal to 1 occurred after seven epochs.

5.6.3 Results

Training was conducted in two stages (using Datasets 3a and 3b to fit the model's parameters and parameter fine-tuning with Dataset 3c) to assess whether fine-tuning parameters with Dataset 3c improves automated cone photoreceptor identification. The U-Net architecture was evaluated on synthetic (Datasets 3a, 3b, 3c) and real data (Datasets 4, 5a and 5b). In Experiment 3 the validation dataset was used to avoid overfitting of the models during training with the training dataset.

5.6.3.1 Stage one

Training with Datasets 3a and 3b took 7 days and required approximately 60 epochs to minimise validation loss (Figure 5.22). Table 5.2 and Figure 5.23 show the results of the trained model when run on test data. Accuracy was high when evaluating the trained model on synthetic data with mean Jaccard similarity coefficients of

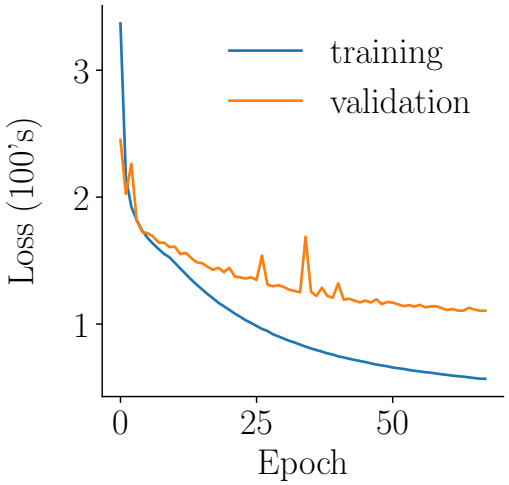


Figure 5.22: New U-Net model training losses during training with Datasets 3a and 3b from one cross-validation fold are shown.

Table 5.2: Jaccard similarity coefficients for all datasets used to test the new U-Net architecture after training with Datasets 3a and 3b.

	Dataset 3a + 3b	**Dataset 3c**	**Dataset 4**	**Dataset 5a**	**Dataset 5b**
μ	0.999	0.999	0.781	0.921	0.755
σ	<0.001	<0.001	0.050	0.011	0.131

μ - mean, σ - standard deviation.

0.999 and 0.999 for Datasets 3a and 3b, and 3c, respectively (Figure 5.24A and B). Accuracy with real data was lower with a mean Jaccard similarity coefficient of 0.781, 0.921 and 0.755 for Datasets 4, 5a and 5b, respectively, predominantly due to a greater number of false positives (Figure 5.24C, D and E). When evaluating the trained model with real data, accuracy was highest with Dataset 5a. This may be due to the higher sampling rate of the original images from Dataset 5 (micron to pixel ratio of between 0.4:1 and 0.5:1), which were subsequently resized to match the scale of the synthetic images (micron to pixel ratio of 1:1).

5.6.3.2 Stage two

Parameter fine-tuning with Dataset 3c took less than 12 hours and required less than 10 epochs to minimise validation loss (Figure 5.25). Table 5.3 and Figure 5.26 show the results of the trained model when run on test data. Accuracy

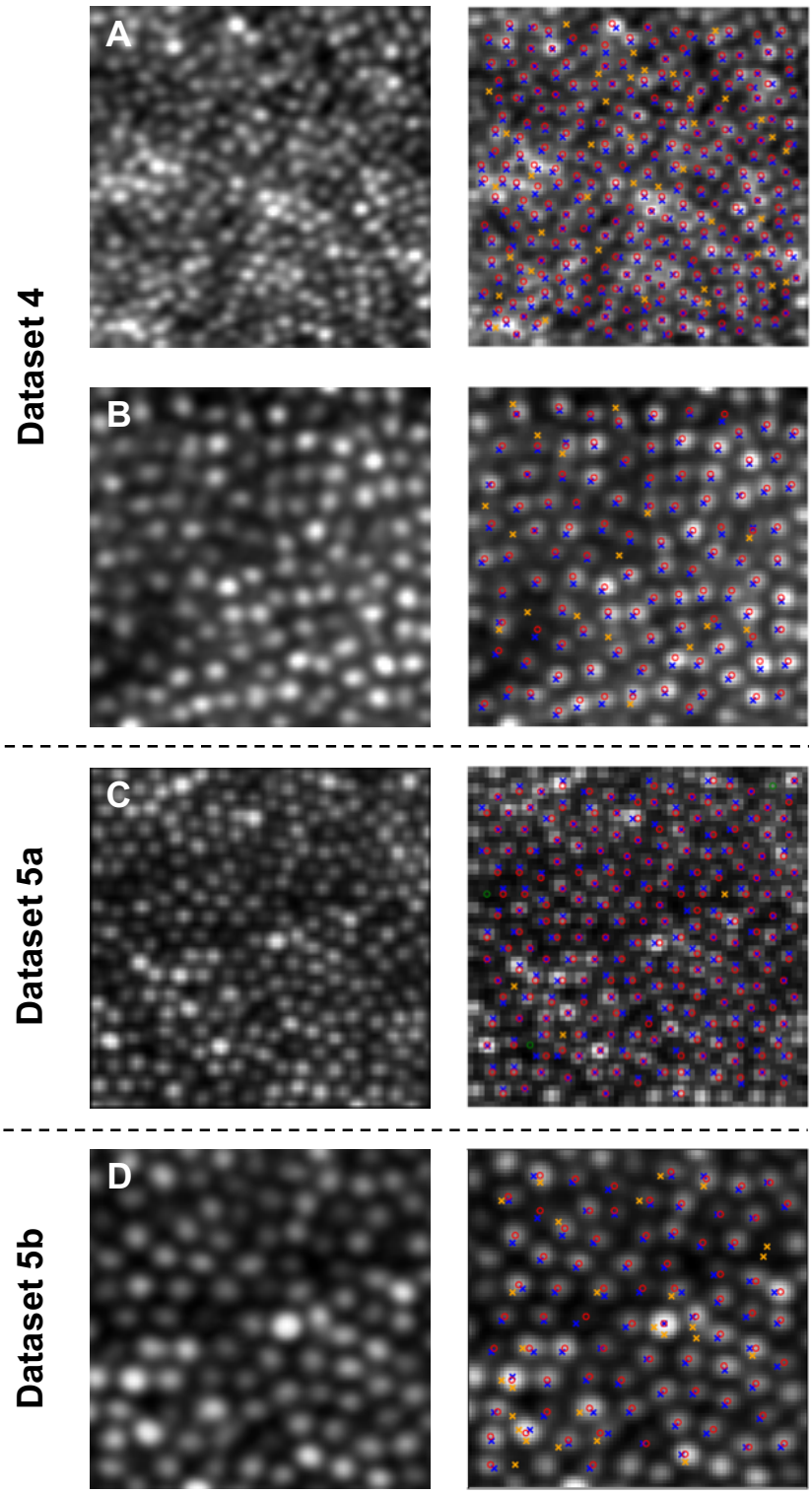
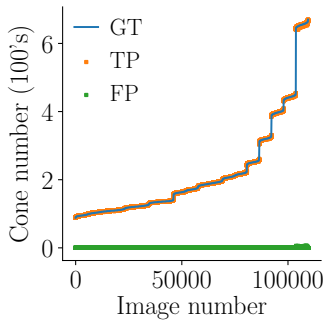
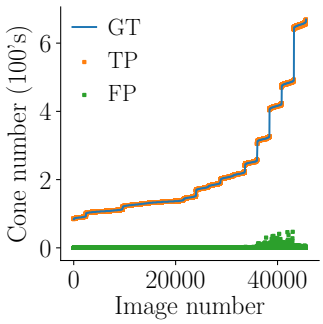


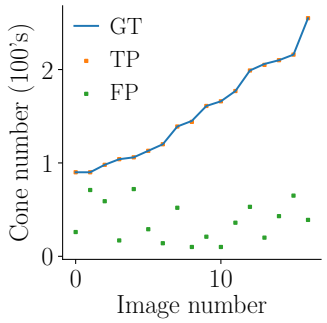
Figure 5.23: Performance of the new U-Net model on three real test datasets after training with Dataset 3a and Dataset 3b. Original images are shown in the left column. Cone detection results are shown in the right column. Crosses denote predicted cone centre locations with blue crosses denoting true positives and orange crosses denoting false positives. Circles denote ground truth cone centre locations with green circles denoting misses (false negatives) and red circles denoting hits (true positives). Jaccard similarity coefficients for each image are: **A** = 0.867, **B** = 0.860, **C** = 0.972 and **D** = 0.745.



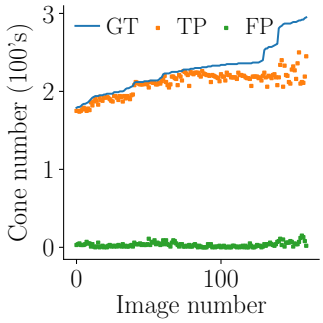
(a) Synthetic Datasets 3a and 3b



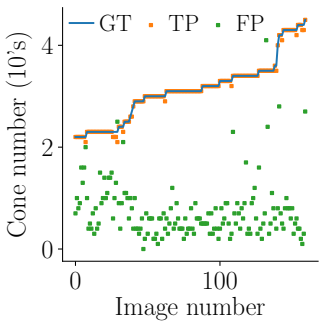
(b) Synthetic Dataset 3c



(c) Real Dataset 4



(d) Real Dataset 5a



(e) Real Dataset 5b

Figure 5.24: New U-Net model accuracy after training with synthetic datasets that included noise (Dataset 3a) and residual aberration (Dataset 3b). The trained U-Net model was tested with 5 different datasets (synthetic and real). Results for each dataset tested from one cross-validation fold are shown. All figures illustrate the number of true positive (TP) and false positive (FP) cones identified relative to the ground truth (GT).

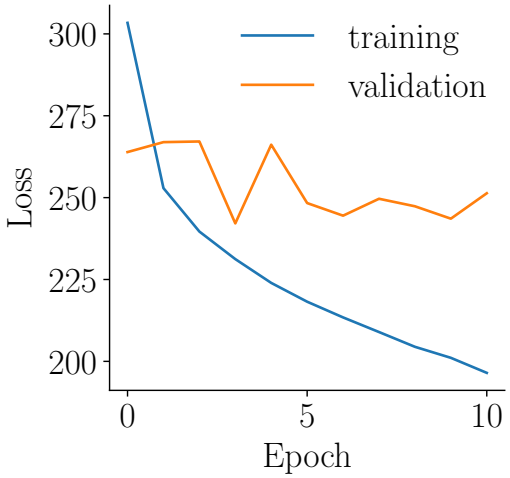


Figure 5.25: New U-Net model training losses during parameter fine tuning with Dataset 3c from one cross-validation fold are shown.

Table 5.3: Jaccard similarity coefficients for all datasets used to test the new U-Net architecture after parameter fine-tuning with Dataset 3c.

	Dataset 3c	**Dataset 4**	**Dataset 5a**	**Dataset 5b**
μ	0.999	0.847	0.951	0.818
σ	<0.001	0.053	0.007	0.0975

μ - mean, σ - standard deviation.

remained high when evaluating the model with synthetic data (mean Jaccard similarity coefficient 0.999) and improved with real data (mean Jaccard similarity coefficients 0.847, 0.951 and 0.818 for Datasets 4, 5a and 5b, respectively) due to a lower false positive rate for Datasets 4 and 5b and a higher true positive prediction rate for Dataset 5a (Figure 5.27).

5.7 Experiment 4

Despite the synthetic data described in Section 5.3.2.1 providing a realistic cone photoreceptor mosaic with added noise and residual aberration, there is still a gap between the synthetic and real datasets (illustrated by Experiment 3) that a CNN model will be unable to overcome without additional training data.

The aim of Experiment 4 is to use transfer learning to leverage the learned

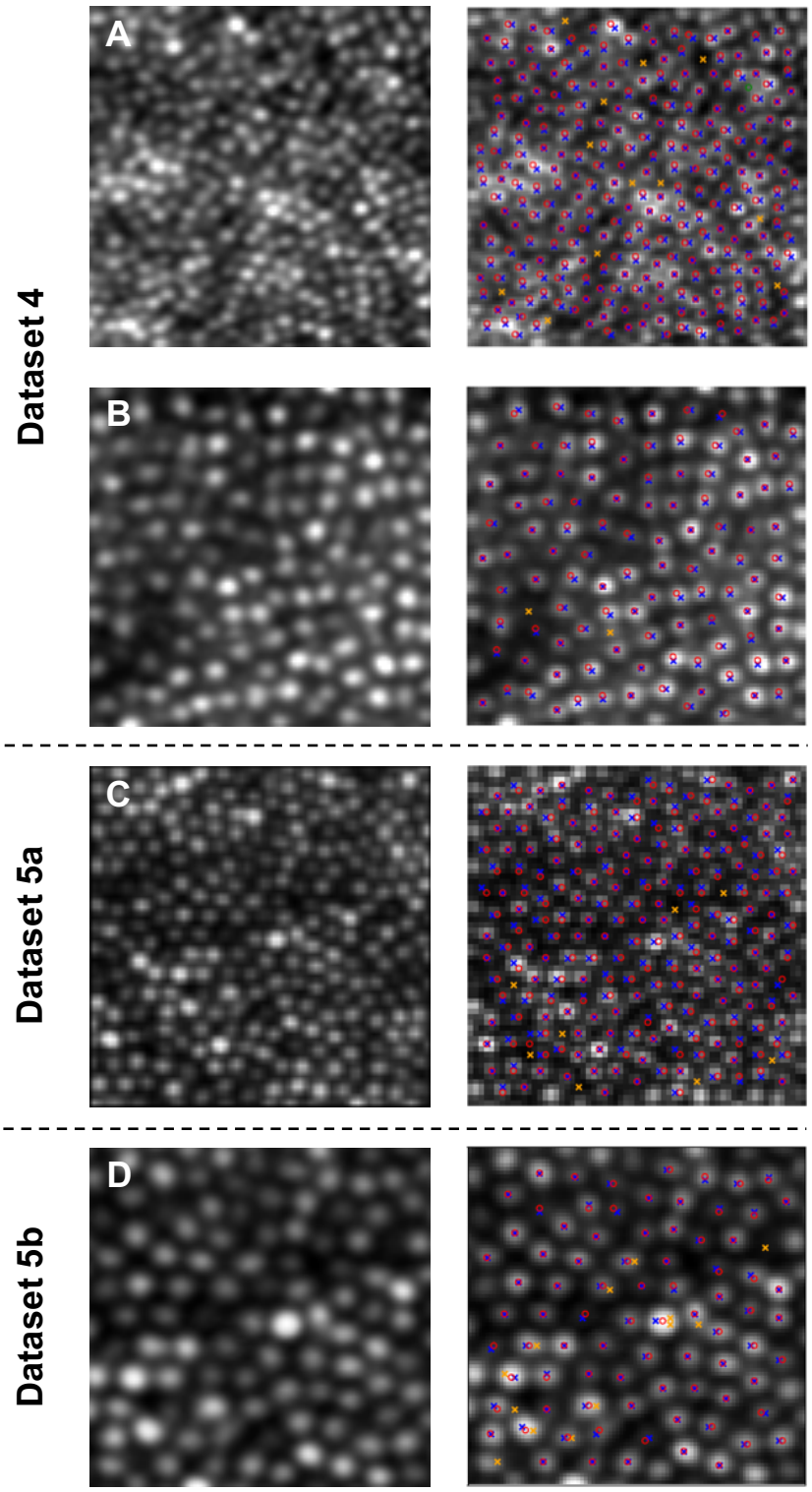
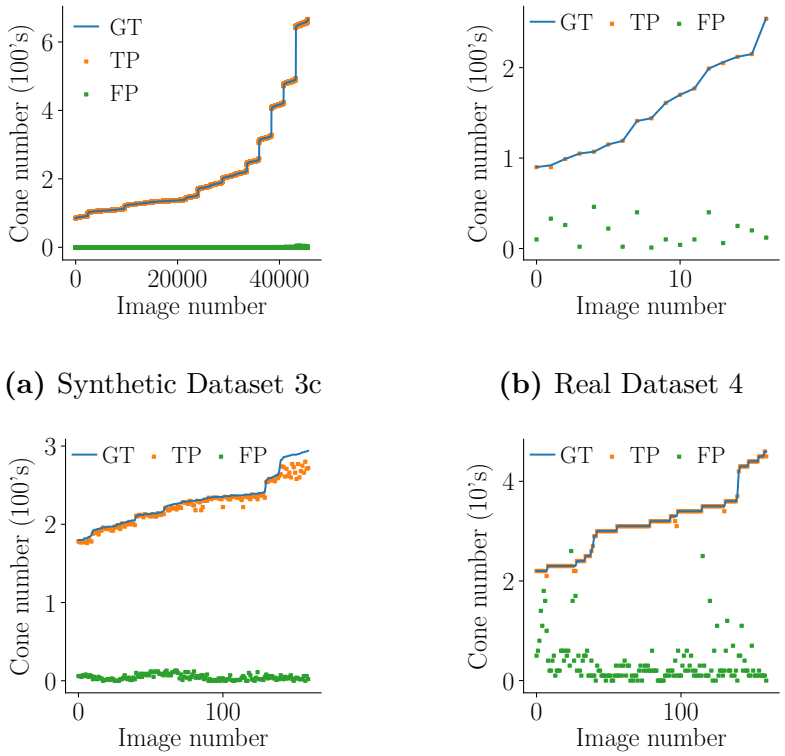


Figure 5.26: Performance of the new U-Net model on three real test datasets after parameter fine-tuning with Dataset 3c. Original images are shown in the left column. Cone detection results are shown in the right column. Crosses denote predicted cone centre locations with blue crosses denoting true positives and orange crosses denoting false positives. Circles denote ground truth cone centre locations with green circles denoting misses (false negatives) and red circles denoting hits (true positives). Jaccard similarity coefficients for each image are: **A** = 0.951, **B** = 0.981, **C** = 0.964 and **D** = 0.863.



(a) Synthetic Dataset 3c

(b) Real Dataset 4

(c) Real Dataset 5a

(d) Real Dataset 5b

Figure 5.27: New U-Net model accuracy after parameter fine-tuning with a synthetic dataset that included the interaction between noise and residual aberration (Dataset 3c). The trained U-Net model was tested with 4 different datasets (synthetic and real). Results for each dataset tested from one cross-validation fold are shown. All figures illustrate the number of true positive (TP) and false positive (FP) cones identified relative to the ground truth (GT).

feature maps from Experiment 3 (trained only with synthetic data) with real training data and evaluate the generalisability and accuracy of the trained U-Net for automated cone photoreceptor identification from real confocal AOSLO images. Transfer learning is a machine learning method that allows learned feature maps from a previously trained model to be leveraged for training a new model.

5.7.1 Input images

Evaluation of the U-Net model from Experiment 3 with test images from Datasets 5a and 5b suggested that calibrated pixel size may have a substantial effect on model performance. Real training data were augmented to reflect different calibrated pixel

sizes for transfer learning to investigate their effect on model performance.

The following formula defines the size of one pixel (P) in microns at the retinal surface for an eye with axial length A:

$$P = A \cdot c \cdot \gamma \tag{5.2}$$

where c is the calibrated pixel size and γ is the retinal scale. Given that the standard deviation of normal axial length measurements is approximately 1.5 mm (mean 23.4 mm) [162–164], variations in AOSLO image scale between participants are likely to be minimal. However, variations in the calibrated pixel size (c) between imaging systems are larger [26, 32, 165, 166] and likely to lead to significant changes in AOSLO image scale.

In order to manipulate image scale (micron to pixel ratio), two real datasets from Dataset 5 (described in Section 5.3.2.3) with different image scales were used. Images from Dataset 5a were resized to represent a micron to pixel ratio of 1:1 while image scale remained unchanged in Dataset 5b. In order to generate Datasets 5a and 5b, images from Dataset 5 were split 80:20 between training (680 images) and testing (160 images). As Dataset 5 consisted of 40 averaged images per participant taken at four retinal locations, data were split to ensure there was no overlap of participants between training and testing. Training data were augmented by 200 fold and this was achieved through a combination of image rotation (between 0° and 360° in 100 steps) and reflection along the x-axis. Datasets 5a and 5b each consisted of 136,160 images (136,000 for training and 160 for testing) with a micron to pixel ratio of 1:1 (Figure 5.20A) and between 0.4:1 and 0.5:1 (Figure 5.20B), respectively. The training data were split 80:20 between training and validation. Zero padding was used to ensure input images remained 96 × 96 pixels in size. The U-Net models were also evaluated using Dataset 4 described in Section 5.3.2.2.

5.7.2 U-Net transfer learning

The pre-trained U-Net model described in Section 5.6.2 was used with transfer learning to overcome the problem of having limited real data. Here, the key

consideration was to determine at what layer depth (L) along the expansive path the pre-trained U-Net model's filters still extracted relevant features from the AOSLO images, and hence at what layer depth the pre-trained model's weights should be frozen (Figure 5.28). Freezing a layer means that its learnable parameters cannot be modified any further. During transfer learning this improves model performance while limiting the time taken for training. The U-Net architecture contained approximately 2.9 million, 0.7 million, 0.14 million and 32 trainable parameters when model weights were frozen at L1, L2, L3 and L4, respectively. An initial learning rate of 0.0001 was used, which was reduced by a factor of two if no reduction in validation loss ≥ 1 occurred after five epochs, to a minimum learning rate of 10^{-5}. The network was trained with a batch size of 32 with four-fold cross-validation. Early stopping was implemented to stop training of the model if no reduction in validation loss ≥ 1 occurred after 7 epochs.

5.7.3 Results

5.7.3.1 Experiment 4a

The aim of this experiment was to use transfer learning with Dataset 5a and evaluate the generalisability and accuracy of the trained U-Net for automated cone photoreceptor identification from real confocal AOSLO images. Transfer learning with Dataset 5a required between approximately 10 (L1 and L4) and 60 (L3) epochs to minimise validation loss (Figure 5.29). In Experiment 4a the validation dataset was used to avoid overfitting of the model during training with the training dataset. Figure 5.30 shows the results of the trained models when run on test data. Evaluating the U-Net on Datasets 4, 5a and 5b, retraining only the final layer along the expansive path of the U-Net architecture (L = 4) provided the most accurate results (mean Jaccard similarity coefficients 0.906, 0.958 and 0.909, respectively, Figure 5.31). Transfer learning improved model performance with real data when compared to training on synthetic data alone (see Section 5.6.3.2).

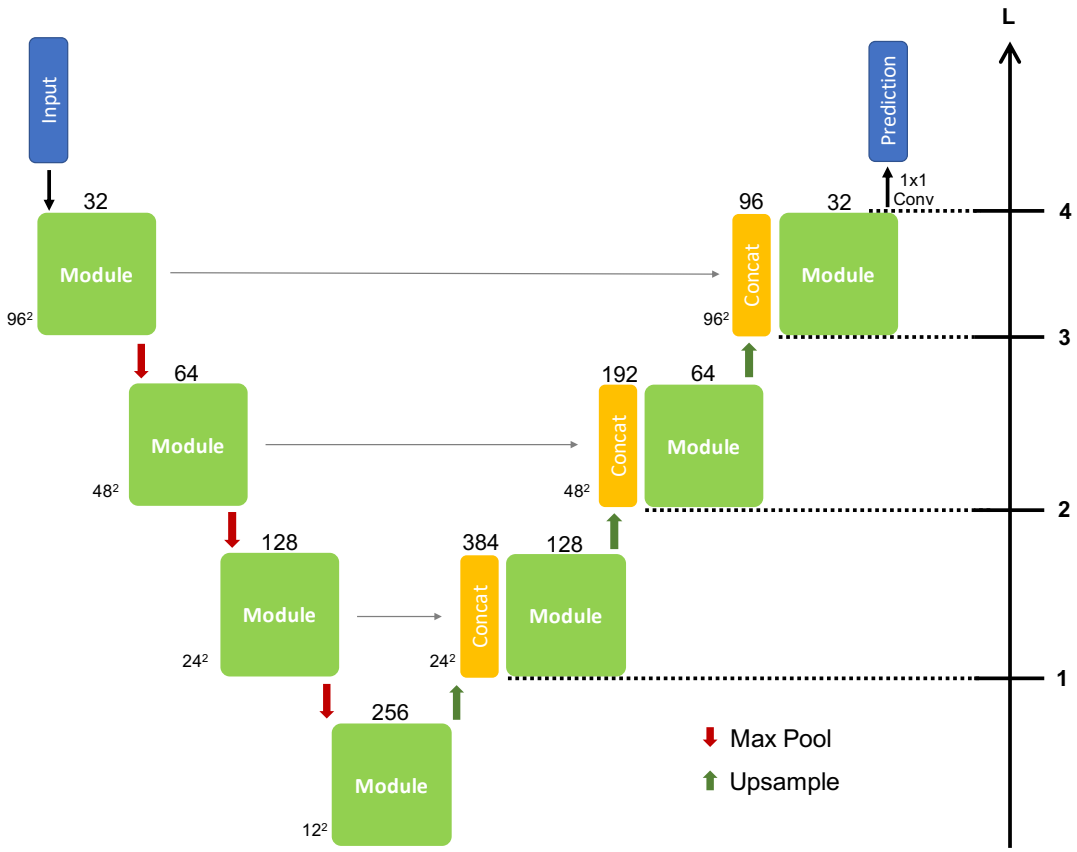


Figure 5.28: The different layer depth (L) considered for transfer learning with a pre-trained new U-Net model.

5.7.3.2 Experiment 4b

The aim of this experiment was to use transfer learning with Dataset 5b and evaluate the generalisability and accuracy of the trained U-Net for automated cone photoreceptor identification from real confocal AOSLO images. Training with Dataset 5b required approximately 10 epochs to minimise validation loss (Figure 5.32). In Experiment 4b the validation dataset was used to avoid overfitting of the model during training with the training dataset. Figure 5.33 shows results of the trained models when run on test data. Retraining the last convolutional block along the expansive path (L = 3) provided the most accurate results when evaluating the model with Datasets 4 and 5b (mean Jaccard similarity coefficient 0.984 and 0.891, respectively. Figure 5.34). Transfer learning improved model performance with real data when compared to training on synthetic data alone (see Section 5.6.3.2).

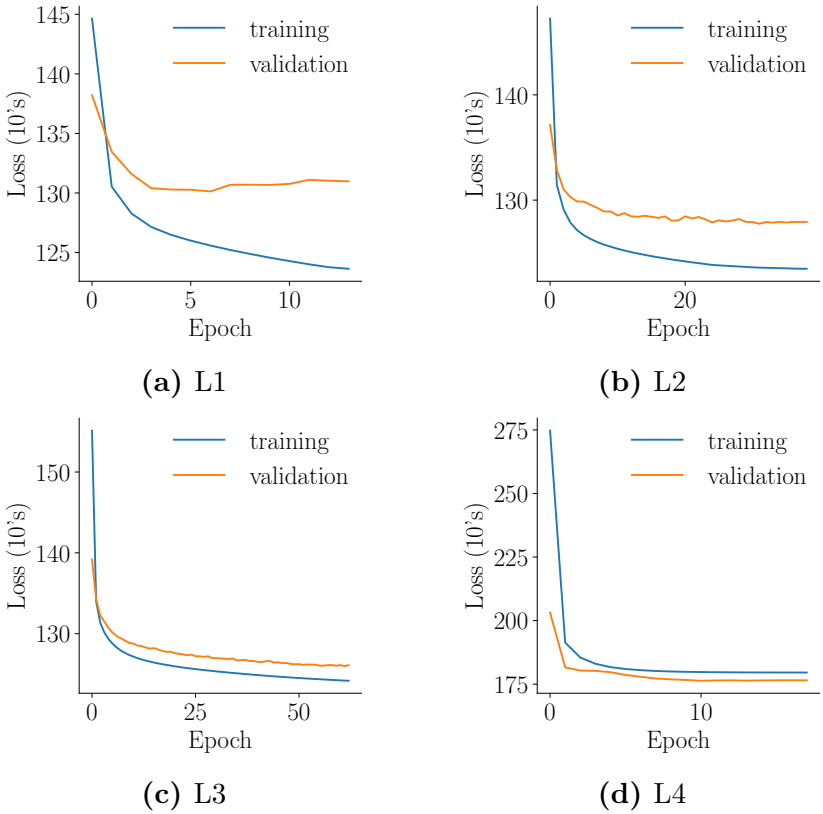


Figure 5.29: New U-Net model training losses for each layer depth (L) considered during transfer learning with Dataset 5a from one cross-validation fold are shown.

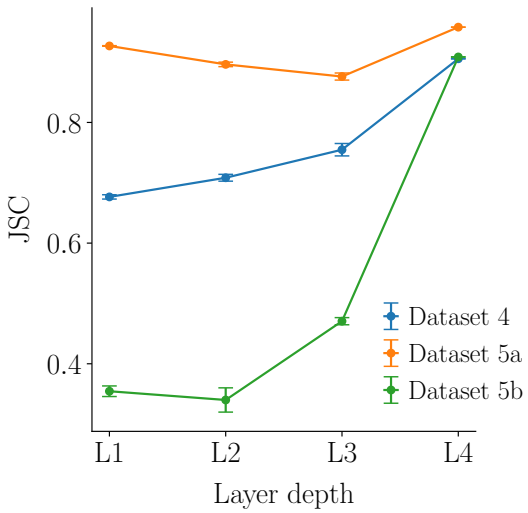


Figure 5.30: Jaccard similarity coefficients for each layer depth (L) considered for transfer learning with Dataset 5a and each real dataset tested (Datasets 4, 5a and 5b). Mean values with error bars (standard deviation over 4 cross-validation folds) are shown. JSC - Jaccard similarity coefficient.

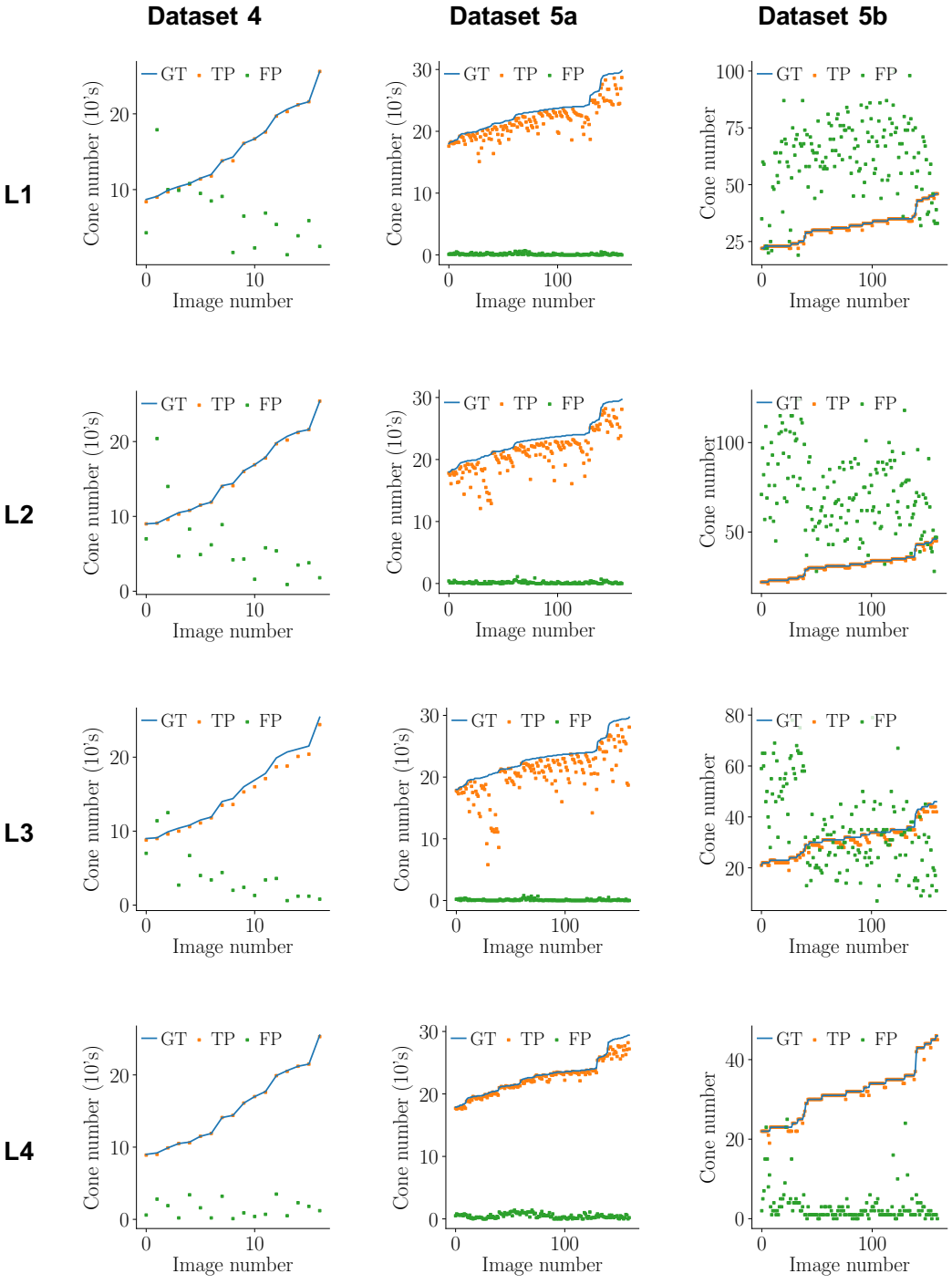


Figure 5.31: Evaluation of the new U-Net model with three real datasets (4, 5a, 5b) following transfer learning with Dataset 5a. Each layer depth (L) considered for transfer learning is represented on a different row. Each column represents the model's results for each real dataset tested. All figures illustrate the number of true positive (TP) and false positive (FP) cones identified in all tested images relative to the ground truth (GT).

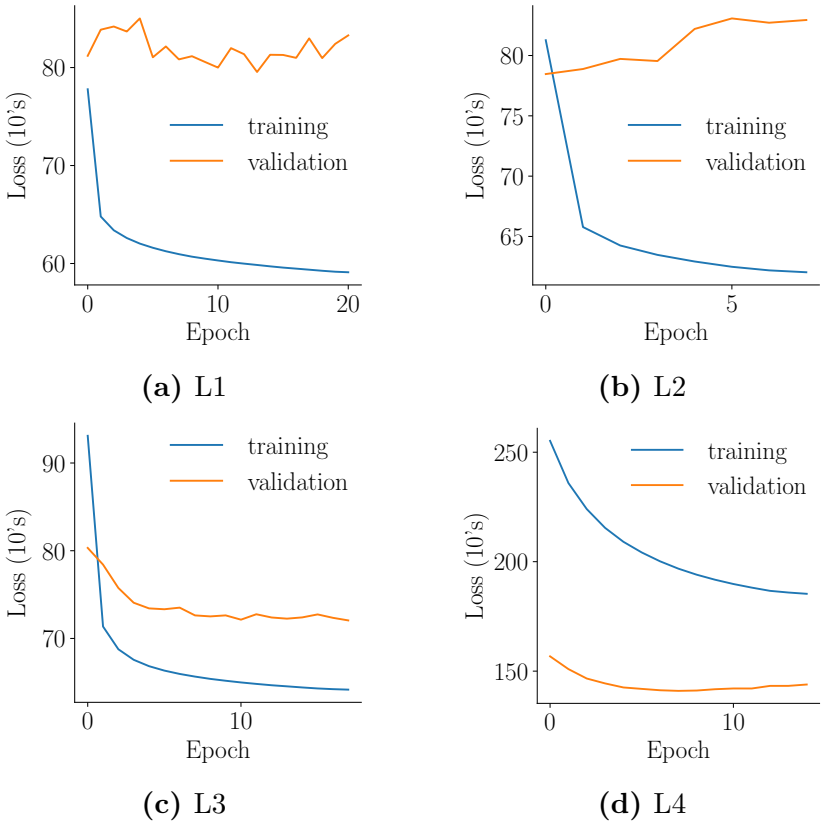


Figure 5.32: New U-Net model training losses for each layer depth (L) considered during transfer learning with Dataset 5b from one cross-validation fold are shown.

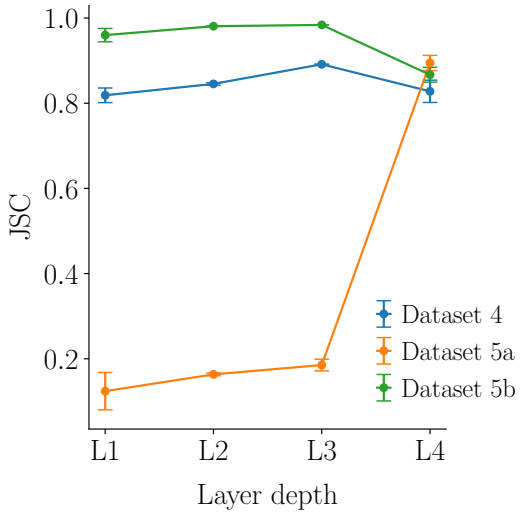


Figure 5.33: Jaccard similarity coefficients for each layer depth (L) considered for transfer learning with Dataset 5b and each real dataset tested (4, 5a and 5b). Mean values with error bars (standard deviation over 4 cross-validation folds) are shown. JSC - Jaccard similarity coefficient.

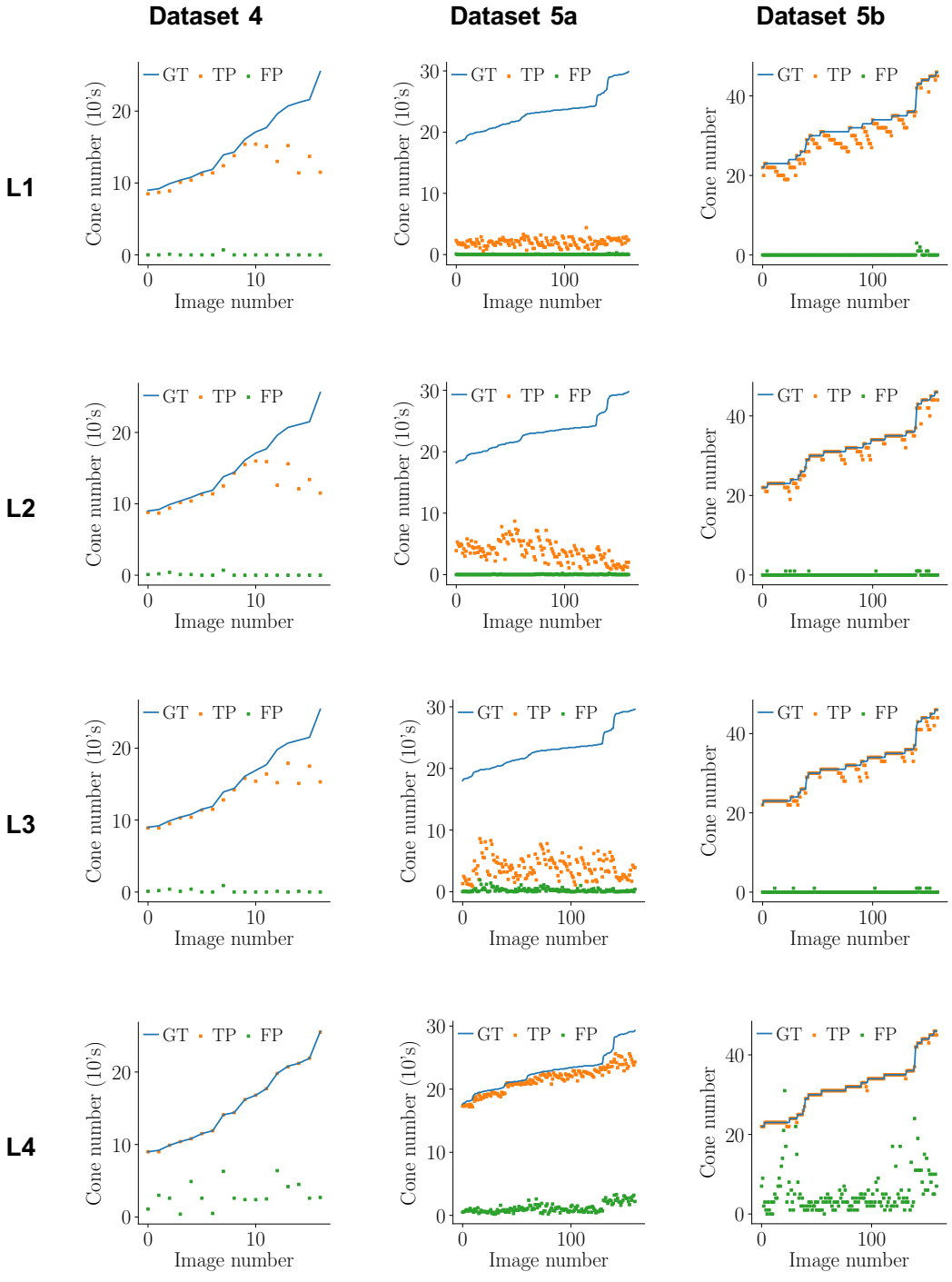


Figure 5.34: Evaluation of the new U-Net model with three real datasets (4, 5a, 5b) following transfer learning with Dataset 5b. Each layer depth (L) considered for transfer learning is represented on a different row. Each column represents the model's results for each real dataset tested. All figures illustrate the number of true positive (TP) and false positive (FP) cones identified in all tested images relative to the ground truth (GT).

5.7.4 Comparison with state-of-the-art

In order to facilitate comparison of the U-Net model, trained and evaluated with images from Dataset 5, with current state-of-the-art cone detection methods [67, 167], the true positive rate (TPR), false discovery rate (FDR) and Dice's coefficient (DC) were calculated for each test image.

$$TPR = \frac{TP}{TP + FN} \tag{5.3}$$

$$FDR = \frac{FP}{TP + FP} \tag{5.4}$$

$$DC = \frac{2TP}{2TP + FP + FN} = \frac{2J}{1 + J} \tag{5.5}$$

where TP represents the number of true positives, FP the number of false positives and FN the number of false negatives.

The true positive rate and false discovery rate both range between 0 and 1. A true positive rate of 1 signifies perfect detection of cones between the prediction and ground truth with no missed cones, while a false discovery rate of 0 signifies perfect detection of cones between the prediction and ground truth with no false positives. Dice's coefficient, similar to the Jaccard similarity coefficient (J), is a commonly used metric in image segmentation. Like the Jaccard similarity coefficient it ranges between 0 and 1, with 0 signifying no similarity and 1 signifying perfect similarity between the prediction and ground truth.

The average performance of the current state-of-the-art confocal CNN (C-CNN) [167] and graph-theory and dynamic programming (GTDP) [67] cone detection methods, implemented by Cunefare *et al.* [167] using Dataset 5, were compared with the average performance (from four-fold cross-validation) of the new U-Net model from Experiment 4b after retraining the last convolutional block along the expansive path (L = 3, see Section 5.7.3.2). This model was selected on the basis of results from the test set used in Experiment 4 from Datasets 4, 5a and 5b.

Briefly, the C-CNN classifies overlapping image patches in order to identify cone locations within an image. The C-CNN, trained on 33 × 33 pixel image patches centred around cone and non-cone locations, classifies whether or not each image patch contains a cone to generate a probability map, which is then processed to identify individual cone locations. Of the 840 confocal AOSLO images comprising Dataset 5, 200 images from five participants were used by Cunefare *et al.* [167] for training the C-CNN and 640 images from 16 participants were used for model evaluation. This produced over 34,000 cone and 88,000 non-cone patches for training the C-CNN model.

The GTDP cone detection method segments cone photoreceptors in AOSLO images [67]. Briefly, it uses a maxima operator to attain preliminary estimates of cones. A quasi-polar transform [168] is then used to transform the preliminary cone estimate consisting of connected pixels from the Cartesian domain into layers in the quasi-polar domain. Segmentation of the layered structure is then performed using graph-theory and dynamic programming. The cone segmentation is that transformed back into the Cartesian domain by applying the inverse quasi-polar transform.

The new U-Net model described in this study was trained using 680 images from 17 participants and evaluated using 160 images from four participants (see Section 5.7.1). Data were augmented (by 200 fold) to produce 136,000 images for training the new U-Net model. A greater proportion of Dataset 5 was used for training the new U-Net model described in this study when compared to the C-CNN [167] due to the differences in the cone detection methods of both models (the C-CNN classifies overlapping image patches to generate a probability map of cone locations while the U-Net generates an output where intensity is related to the probability of a cone being present). However, similarly sized datasets of input images were used to train both CNN models (new U-Net model: 136,000 input images, C-CNN: over 122,000 input image patches). Due to the different proportion of images from Dataset 5 used for training and evaluation of both CNN models, it is possible that although they are from the same dataset, the images used for evaluating the new U-Net model may be different to those used by Cunefare *et al.* for evaluation.

Table 5.4: Comparing results of the new U-Net model with current state-of-the-art cone detection methods.

	TPR (SD)	FDR (SD)	DC (SD)
U-Net	0.986 (<0.001)	0.002 (<0.001)	0.992 (<0.001)
C-CNN*	0.989 (0.012)	0.008 (0.014)	0.990 (0.010)
GTDP*	0.990 (0.011)	0.015 (0.018)	0.988 (0.012)

C-CNN - confocal convolutional neural network [167] for cone detection, DC - Dice's coefficient, FDR - false discovery rate, GTDP - graph-theory and dynamic programming [67] cone detection method, SD - standard deviation, TPR - true positive rate, *results of the cone detection methods as implemented by Cunefare *et al.* [167].

The average performance of the new U-Net model, C-CNN and GTDP cone detection methods are summarised in Table 5.4. Figure 5.35 displays the results of all three cone detection methods compared to manual grading. The mean Dice's coefficient was 0.992 for the new U-Net model, 0.990 for the C-CNN and 0.988 for the GTDP methods. While the true positive rate was highest for the GTDP method (0.990), the new U-Net model had the lowest false discovery rate (0.002).

5.8 Discussion

This study demonstrates a fully automated deep learning based method for identifying cone photoreceptors in confocal AOSLO images that has a good agreement with the current gold standard of manual marking. Using synthetic images of the cone photoreceptor mosaic over a wide range of retinal eccentricities provided sufficient data to train a newly developed CNN and leverage the learned feature maps with manually marked real images to identify cone photoreceptors. The success of this approach demonstrates the suitability of using synthetic data to overcome the limited availability of annotated real datasets. The use of synthetic data over a wide range of retinal eccentricities also increases the generalisability of this model to confocal AOSLO images taken over a similarly wide range of retinal eccentricities.

AOSLO imaging systems provide highly sensitive measures of retinal structure and function, and can be used to enhance our understanding of underlying disease

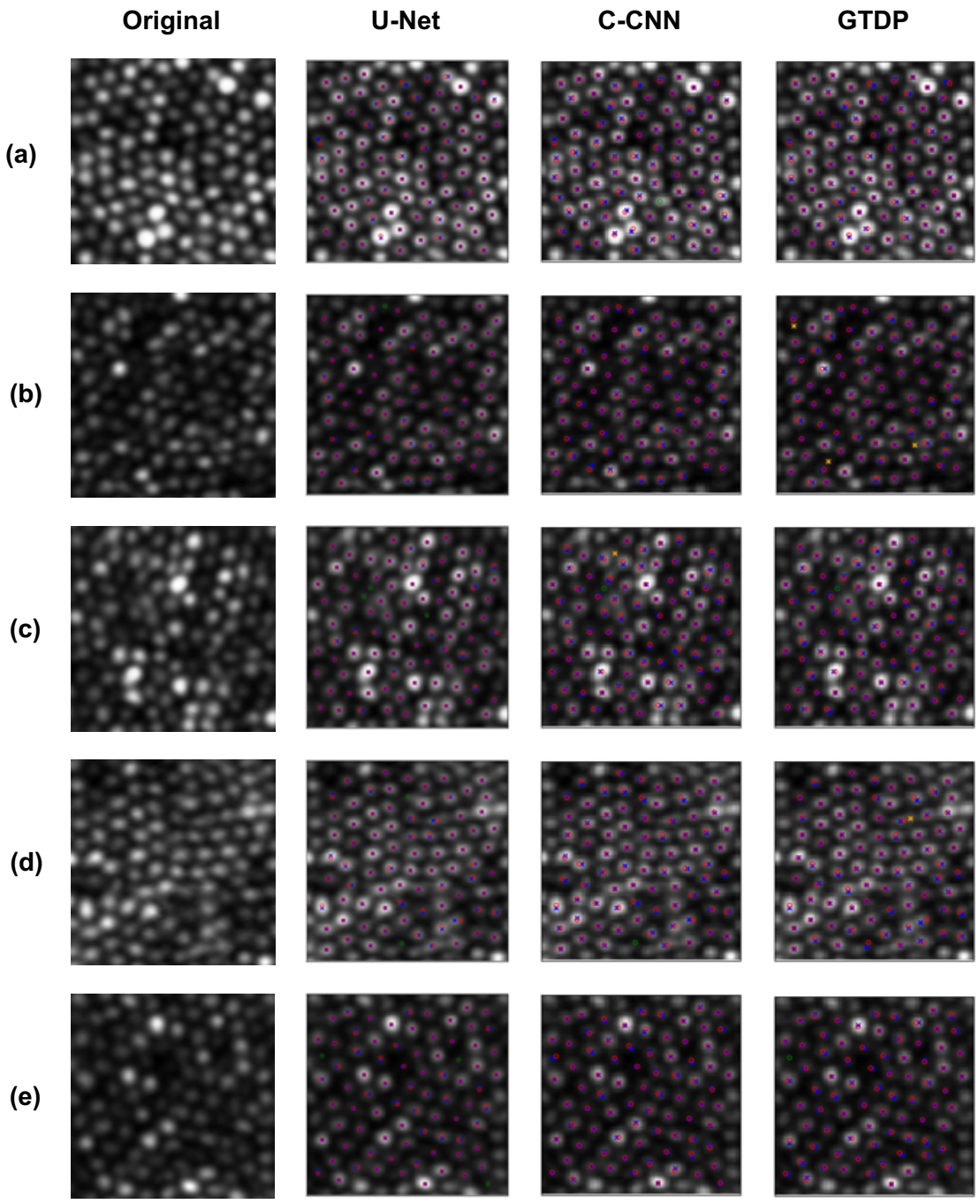


Figure 5.35: Comparing performance of the new U-Net model with current state-of-the-art cone detection methods for five images. Crosses denote predicted cone centre locations with blue crosses denoting true positives and orange crosses denoting false positives. Circles denote ground truth cone centre locations with green circles denoting misses (false negatives) and red circles denoting hits (true positives). Jaccard similarity coefficients for each image with the new U-Net method are: **(a)** = 1.0, **(b)** = 0.988, **(c)** = 0.967, **(d)** = 0.975 and **(e)** = 0.963. Jaccard similarity coefficients for each image with the C-CNN method are: **(a)** = 0.988, **(b)** = 1.0, **(c)** = 0.978, **(d)** = 0.988 and **(e)** = 1.0. Jaccard similarity coefficients for each image with the GTDP method are: **(a)** = 1.0, **(b)** = 0.966, **(c)** = 0.989, **(d)** = 0.989 and **(e)** = 0.987. C-CNN - confocal convolutional neural network [167] for cone detection, GTDP - graph-theory and dynamic programming [67] cone detection method.

mechanisms, allow earlier disease detection, and provide cell-specific biomarkers for monitoring treatment outcomes. Identification of individual photoreceptors is a prerequisite for extracting quantitative information from AOSLO images of the photoreceptor mosaic (such as photoreceptor density or packing regularity). Manual marking of individual cone photoreceptors is subjective and time consuming (taking several hours for a single participant). This has created a bottleneck in the image processing pipeline, limiting the clinical use of AO retinal imaging technologies. Automated cone detection methods represent an important step towards the widespread clinical use of AO retinal imaging.

Evaluating the proposed U-Net model on the dataset used by Cunefare *et al.* [167] (Dataset 5) yielded a Dice's coefficient of 0.992 (Table 5.4). Not only do our results compare favourably with manual marking, considered to be the current gold standard, but they marginally outperform the results of current state-of-the-art automated cone detection algorithms [67, 167] that have been evaluated using the same dataset [167] (Table 5.4). Although our results are promising, it is important to note that we evaluated our algorithm using 160 images from four participants in Dataset 5, while Cunefare *et al.* evaluated the automated algorithms using 640 images from 16 participants in the same dataset (see Section 5.7.4). Furthermore, selection of the new U-Net model used for comparison with state-of-the-art cone detection methods was on the basis of results from the test set in Dataset 5b. It is therefore possible that this may have biased the results in favour of the new U-Net architecture.

Evaluating the model on the Oxford dataset (Dataset 4, see Section 5.3.2.2) demonstrates the generalisability of this approach. Evaluation with Dataset 4 following transfer learning with Dataset 5a and 5b yielded mean Jaccard similarity coefficients of 0.906 (Dice's coefficient of 0.950) and 0.891 (Dice's coefficient of 0.943), respectively. Results from transfer learning with Dataset 5a, which more closely resembles the micron to pixel ratio of images in Dataset 4, were slightly higher. Future work will include expanding Dataset 4 to include more annotated data followed by transfer learning, which is likely to improve the model's accuracy with

this dataset. Although the synthetic data provided a realistic cone photoreceptor mosaic that was modelled on histologic data, the images did not simulate rod photoreceptors (only a problem at higher retinal eccentricities) or blood vessels and therefore were not truly representative of real data. Updating the simulation used to produce synthetic data to improve agreement between synthetic and real datasets will enable future work to improve model accuracy and enable automated rod photoreceptor identification. Supervised deep learning models are trained using labelled data. Poor repeatability and reliability of cone identification [60, 169] from manual marking of training labels will adversely affect a model's performance. Utilising datasets that have been manually marked by multiple graders to improve agreement may help to improve future performance.

Split detector AOSLO imaging [46] allows visualisation of cone photoreceptor inner segments and has been shown to reveal inner segment structures in the absence of reflectivity with confocal AOSLO [46, 170, 171]. A number of algorithms using split detector and multimodal AOSLO images for automated photoreceptor identification have been proposed [142, 143, 167, 172–174]. Deep learning provides an effective way of learning features directly from the image data, enabling extension of the U-Net model to incorporate additional non-confocal imaging modalities, images containing different features (such as rod photoreceptors), and images from patients with retinal disease. Future work will include expansion of the model's capabilities with the addition of non-confocal imaging modalities, detection of rod photoreceptors, and photoreceptor identification from diseased retinae. With further work, similar approaches utilising synthetic data could be used to develop fully automated algorithms to provide reliable and accurate photoreceptor statistics to assist clinicians with the diagnosis and treatment of patients with macular diseases.

6

AOSLO imaging in participants with Stargardt disease

Contents

6.1 Contributions 157
6.2 Background 157
6.3 Methods 159
6.3.1 AOSLO imaging 159
6.3.2 AOSLO image scaling and processing 164
6.4 Results 165
6.4.1 AOSLO images in Stargardt disease 165
6.5 Automated reference frame selection in participants with Stargardt disease 173
6.6 Automated cone photoreceptor identification in participants with Stargardt disease 174
6.7 Discussion 176
6.7.1 Discrete hyper-reflective foci 177
6.7.2 ABCA4-associated optical gap phenotype 178
6.7.3 Retinal pigment epithelium mosaic 179
6.7.4 Automated reference frame selection in participants with Stargardt disease 181
6.7.5 Automated cone photoreceptor identification in participants with Stargardt disease 181
6.8 Conclusions 182

6.1 Contributions

My contributions to this work include: obtaining ethical approvals; study design; participant recruitment; acquisition, analysis and interpretation of data. The default image processing pipeline for the Oxford AOSLO, developed by Laura Young, was used to produce the averaged AOSLO images presented in this chapter. I had full access to all of the data presented in this chapter and take responsibility for the integrity of the data and the accuracy of the data analysis.

6.2 Background

There are a number of ongoing treatment trials in patients with Stargardt disease involving gene replacement therapy, cell-based therapy and pharmacotherapy [51]. With the development of these new therapies cell-specific biomarkers are becoming increasingly important for monitoring treatment outcomes. The cellular resolution achieved with *in vivo* AO retinal imaging systems makes the use of cell-specific imaging biomarkers possible. Understanding the significance of structural changes observed with high-resolution *in vivo* retinal imaging is a prerequisite to using image based biomarkers. The interpretation of image features from high-resolution images of patients with retinal disease is challenging and *in vivo* imaging precludes histopathological correlation of image features. Studies reporting the results of high-resolution *in vivo* imaging in patients with retinal disease are crucial to improve understanding and interpretation of observed structural features.

There are a limited number of publications reporting the results of AOSLO imaging in patients with Stargardt disease [112, 113, 175, 176]. Investigations of early Stargardt disease with *in vivo* AOSLO imaging by Song *et al.* [112] have demonstrated evidence of centrifugal disease progression from the fovea. They observed areas of increased rod and cone photoreceptor spacing in areas of normal conventional clinical imaging, suggesting that photoreceptor loss precedes clinically detectable retinal pigment epithelial (RPE) disease [112]. Longitudinal studies of AOSLO imaging in patients with Stargardt disease have observed that the

earliest cone photoreceptor spacing abnormalities occur in areas of homogeneous short-wavelength fundus autofluorescence (SW-FAF, indicating an absence of RPE disease), normal vision and normal outer retinal structure on spectral-domain optical coherence tomography (SD-OCT) [175]. These early abnormalities may support the hypothesis that photoreceptor degeneration is a primary event and does not occur secondarily to RPE cell death. These changes are followed by a heterogeneous increase in SW-FAF signal with cone loss, and then by a reduction of SW-FAF signal with cone and RPE cell death [175]. *In vivo* autofluorescence AOSLO imaging in patients with Stargardt disease has revealed patterns of autofluorescence that appear to colocalise with photoreceptors [113]. These observations support histologic reports of photoreceptor bisretinoid accumulation [95] and may be useful as an early biomarker in patients with Stargardt disease. New methods to facilitate the interpretation of complex multimodal imaging datasets from patients with Stargardt disease with comparison of structural and functional information [176], as well as methods for the identification of cone photoreceptors [60, 143, 174] have also been described.

In addition to the aforementioned publications, there are a limited number of published abstracts reporting the results of AOSLO imaging in patients with Stargardt disease [68, 177–187]. Thus far, only a few of these abstracts have been published as papers [179, 182, 183, 186, 187]. In addition to the results of the previously mentioned publications, these abstracts report the usefulness of AOSLO imaging as an end point in interventional therapeutic clinical trials for patients with Stargardt disease [181] and the importance of alternative (non-confocal) high-resolution imaging modalities in interpreting image features [178, 184].

Participant numbers are limited in published papers and abstracts reporting the results of AOSLO imaging in Stargardt disease. Further studies are required to report image features identified in patients with Stargardt disease to facilitate AOSLO image interpretation in this cohort. The main objectives of this chapter are: to present the results of AOSLO imaging in a cohort of participants with Stargardt disease described in Chapter 2 that represent heterogeneous retinal

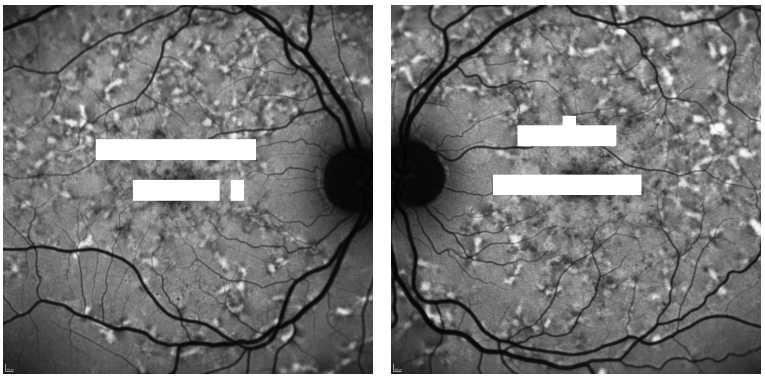
phenotypes; and to correlate structural information from AOSLO imaging with multimodal conventional retinal imaging to assist in the interpretation of AOSLO images in these patients.

6.3 Methods

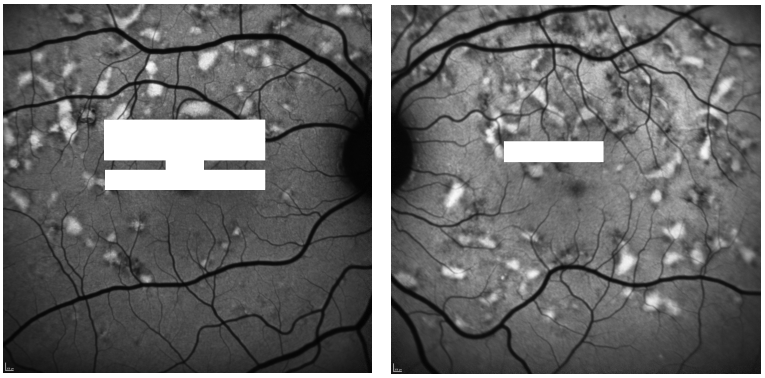
Sixteen participants with Stargardt disease and seven controls underwent clinical assessment (described in Section 2.3.2) and confocal AOSLO imaging (Figures 6.1, 6.2, 6.3, 6.4 and 6.5). All participants provided written informed consent prior to participation in this study and National Health Service Research Ethics Committee approval was obtained from the South Central – Oxford B Research Ethics Committee (reference - 15/SC/0144) and the Yorkshire & The Humber - Sheffield Research Ethics Committee (reference - 16/YH/0265).

6.3.1 AOSLO imaging

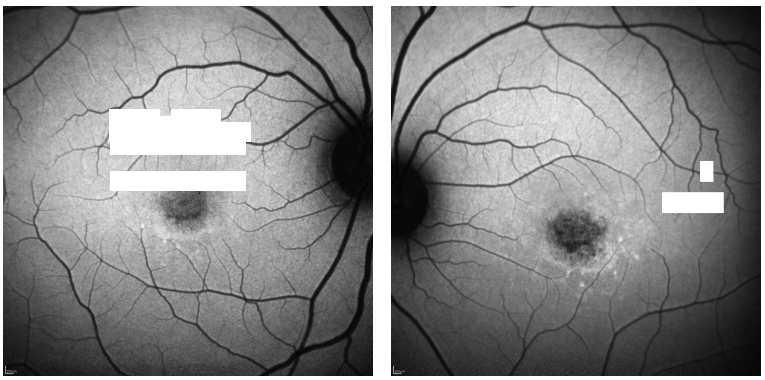
Confocal near-infrared reflectance imaging (wavelength = 850 nm, full width at half maximum = 50 nm) was carried out using a previously described setup [26]. A participant's head is stabilised using a wax dental impression on an adjustable bite-bar. Alignment of the eye to the system was performed using the wavefront-sensor image as a guide and the use of refractive correction (prescription glasses or trial lenses) was based on retinal image quality. Images were acquired at a rate of 30 Hz with a 1.6° × 1.0° field of view, and retinal locations were targeted by directing the participant's gaze using a fixation target on a display located 1.75 m from the eye. In participants with extra-foveal fixation, an adjustment was made to account for the position of the preferred retinal locus (PRL) relative to the foveal centre and retinal features from fundus images were used to check accuracy. The centre of all fixation points recorded after the initial 10 seconds of a macular microperimetry examination was used to determine the PRL position on the retinal surface (see Macular microperimetry in Section 2.3.3). Pharmacologic pupil dilation and cycloplegia was not performed routinely, which allowed participants to focus naturally on the display. The selection of a long imaging wavelength reduced



(a) Participant S02

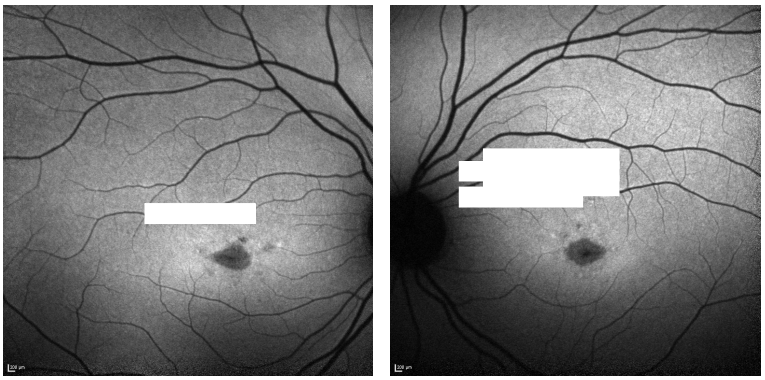


(b) Participant S04

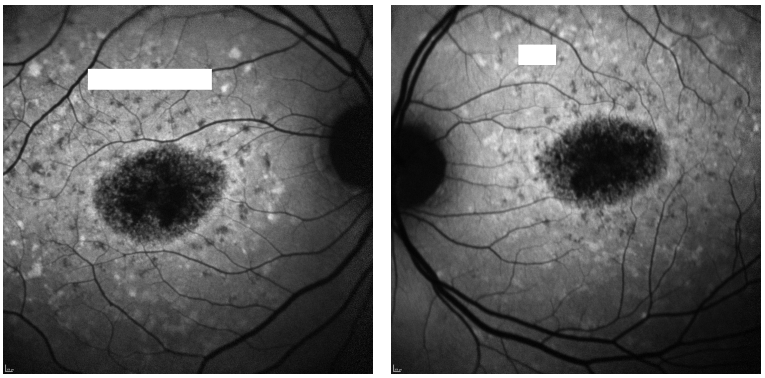


(c) Participant S06

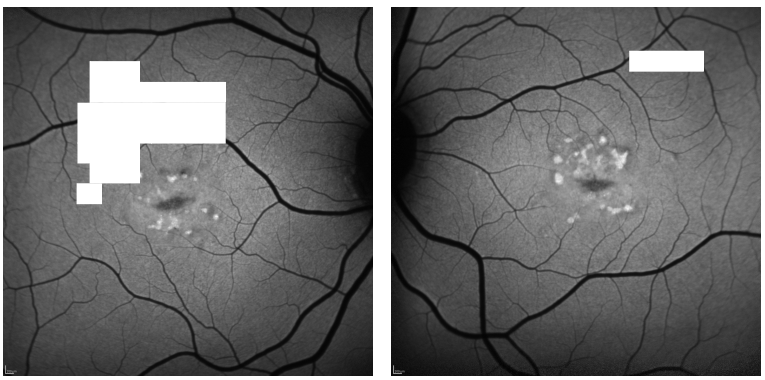
Figure 6.1: Retinal locations imaged with the AOSLO in participants with Stargardt disease. White boxes representing a 1.6° × 1.0° field of view (the size of the AOSLO scan patch) have been placed on a 30° short-wavelength fundus autofluorescence image to illustrate the retinal locations imaged with the Oxford AOSLO. Right eye images are in the left column and left eye images are in the right column.



(a) Participant S07

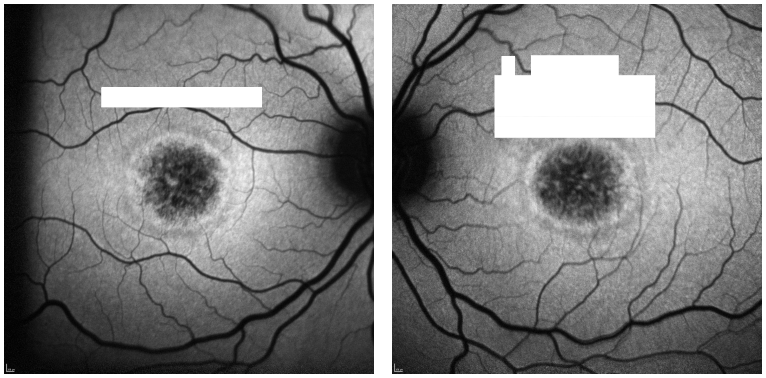


(b) Participant S08

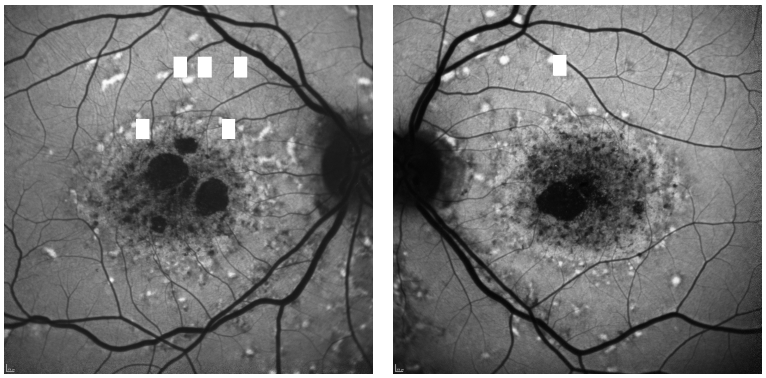


(c) Participant S10

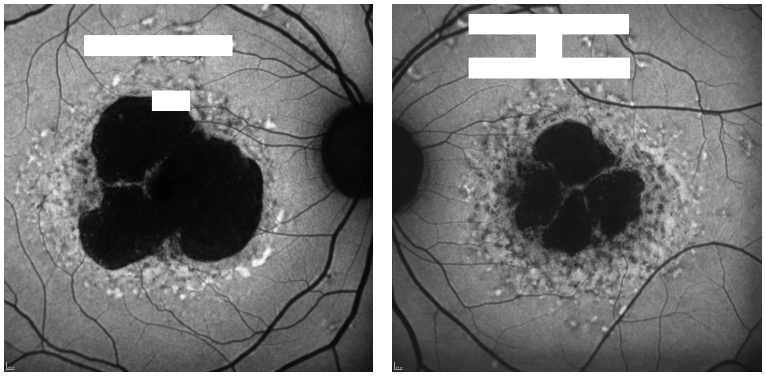
Figure 6.2: Retinal locations imaged with the AOSLO in participants with Stargardt disease. White boxes representing a 1.6° × 1.0° field of view (the size of the AOSLO scan patch) have been placed on a 30° short-wavelength fundus autofluorescence image to illustrate the retinal locations imaged with the Oxford AOSLO. Right eye images are in the left column and left eye images are in the right column.



(a) Participant S12

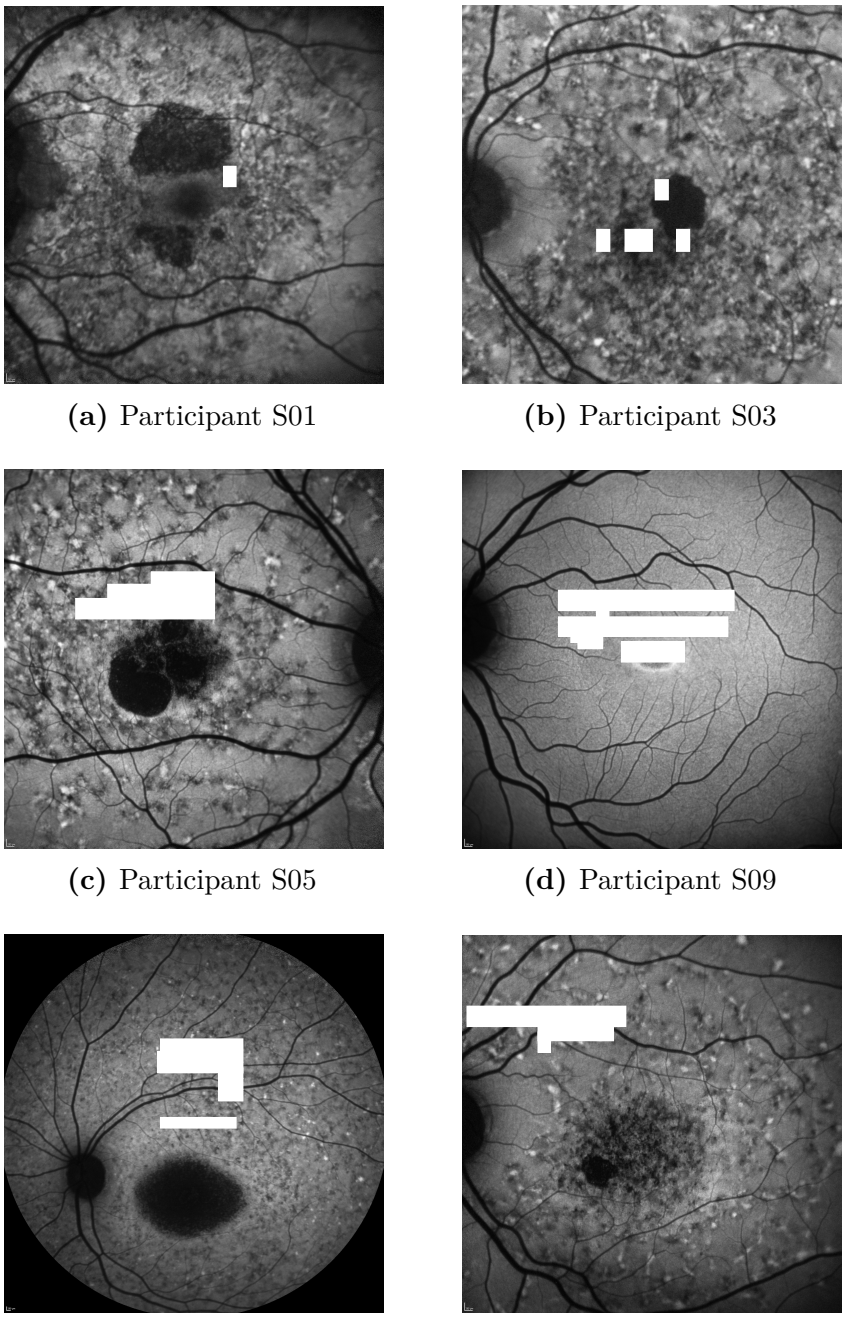


(b) Participant S13



(c) Participant S15

Figure 6.3: Retinal locations imaged with the AOSLO in participants with Stargardt disease. White boxes representing a 1.6° × 1.0° field of view (the size of the AOSLO scan patch) have been placed on a 30° short-wavelength fundus autofluorescence image to illustrate the retinal locations imaged with the Oxford AOSLO. Right eye images are in the left column and left eye images are in the right column.

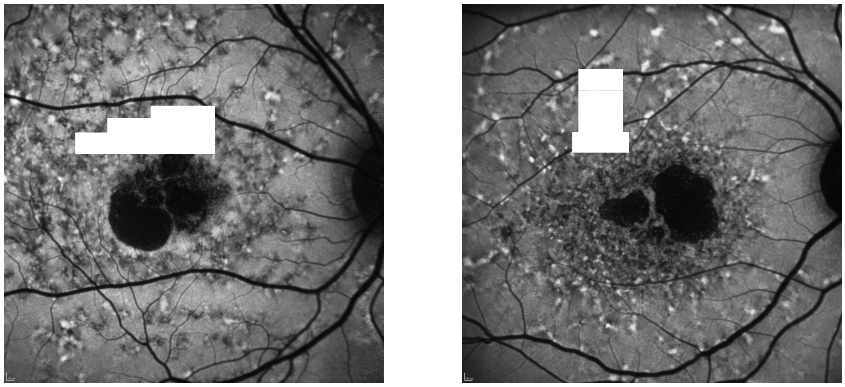


(a) Participant S01 **(b)** Participant S03

(c) Participant S05 **(d)** Participant S09

(e) Participant S11 **(f)** Participant S16

Figure 6.4: Retinal locations imaged with the AOSLO in the left eye of participants with Stargardt disease. White boxes representing a 1.6° × 1.0° field of view (the size of the AOSLO scan patch) have been placed on a 30° or 55° short-wavelength fundus autofluorescence image to illustrate the retinal locations imaged with the Oxford AOSLO.



(a) Participant S05 **(b)** Participant S14

Figure 6.5: AOSLO imaging locations in the right eye of participants with Stargardt disease. White boxes representing a 1.6° × 1.0° field of view (the size of the AOSLO scan patch) have been placed on a 30° short-wavelength fundus autofluorescence image to illustrate the retinal locations imaged with the Oxford AOSLO.

illumination-related pupil constriction and, to maximise natural pupil dilation, all imaging was performed in a darkened room.

Regions of interest were selected prior to imaging using multimodal retinal imaging collected during the clinical assessment, and during AOSLO image capture fundus images were used with a graphical user interface to target these regions. The regions of interest in participants with Stargardt disease were selected to include the leading disease edge, the transition from an area of relatively normal to degenerated retina and the site of active degeneration. Due to the variability in retinal phenotype and disease severity between participants, a standard set of retinal locations for imaging were not used.

6.3.2 AOSLO image scaling and processing

The Gullstrand schematic human eye with an axial length of 24 mm yields a conversion factor of 291 µm per degree of visual angle [188], which was scaled by the participant's axial length (measured for each participant using the Zeiss IOL Master, Section 2.3.2) to determine the scale of each retinal image in microns per pixel (see equation 5.2). Axial length was assumed to be constant with eccentricity as it varies by less than 2% across the range of retinal eccentricities imaged in this

study (0° - 20°) [189]. Raw images were corrected for the distortions produced by fixational eye movements before being averaged (described in Section 1.4.3).

6.4 Results

Sixteen unrelated participants with Stargardt disease and 7 controls underwent confocal AOSLO imaging and were included in this study. Of these, 4 participants with Stargardt disease and 3 controls were followed up after their initial visit every 6 months for between 6 and 18 months. The demographics and clinical phenotype of participants are described in Section 2.4.1 and 2.4.2, respectively. AOSLO images near the foveal centre have been previously published for 5 control participants (C01, C02, C04, C05, C06) [26].

6.4.1 AOSLO images in Stargardt disease

Images were acquired over a range of retinal eccentricities (0° to 20°) and were of variable quality. A large imaging area was attempted in most participants, however, the area of usable images was low. Imaging areas with more severe retinal degeneration yielded fewer usable images compared to imaging areas of near normal retina. Additional factors that impacted the area of usable images were related to the AOSLO system. Our understanding of how to improve the quality of captured AOSLO images developed over the course of this study. Consequently, the area of usable images may have been affected by the participant recruitment date. The biggest impact on AOSLO image quality was related to participant factors. These are discussed in detail in Chapter 4 and have a substantial effect on image quality.

Image features of retinal structures identified from AOSLO images include the photoreceptor mosaic in areas of near normal and degenerated retina, discrete hyper-reflective foci and the RPE mosaic. Cone photoreceptors were readily identifiable from the photoreceptor mosaic in areas of near normal retinal structure. However, in areas of retinal degeneration the photoreceptor mosaic was disorganised and lacked unambiguous cones. Hyper-reflective spots, termed discrete hyper-reflective

foci, were identified in images from participants with Stargardt disease due to their relative hyper-reflectance compared to other image features.

The identification and interpretation of image features from participants with retinal disease is challenging. *In vivo* retinal imaging precludes histopathological correlation of image features, which is still considered to be the gold standard for accurate interpretation. The addition of multimodal high-resolution retinal imaging using alternative imaging modalities such as split-detector, dark-field and autofluorescence, coupled with functional cellular information may aid in the identification and interpretation of relevant image features from these participants.

6.4.1.1 Photoreceptor mosaic

AOSLO images showed a regular photoreceptor mosaic in areas of normal or near normal retinal structure (Figure 6.6). The retinal area corresponding to the photoreceptor mosaic illustrated in Figure 6.6 demonstrates a relatively uniform SW-FAF signal with normal appearance of outer retinal layers on SD-OCT despite the evidence of adjacent retinal flecks.

In areas of retinal degeneration, the photoreceptor mosaic was disorganised and lacking unambiguous cones (Figure 6.7). The retinal area corresponding to the photoreceptor mosaic in Figure 6.7 exhibits reduced SW-FAF signal and loss of the outer retinal layers on SD-OCT.

6.4.1.2 Discrete hyper-reflective foci

To aid with identification and interpretation of discrete hyper-reflective foci, their size (full width at half maximum) in microns was quantified. The following formula was used to determine the full width at half maximum of the discrete hyper-reflective foci (W_{DHF}):

$$W_{DHF} = \sqrt{2 \cdot \log_2 \sigma} \tag{6.1}$$

where σ was determined by fitting a 2D Gaussian to an average hyper-reflective foci image. As hyper-reflective foci are represented as areas of increased intensity

Figure 6.6: Photoreceptor mosaic with a regular packing arrangement in an area of near normal retina 3° superior to the fovea in participant S04. (**A**) 30° short-wavelength fundus autofluorescence image, white box represents the retinal area corresponding to the photoreceptor mosaic in **C** and retinal cross section in **B**. (**B**) Spectral-domain optical coherence tomography of the retinal region represented by the white box in **A**. (**C**) Photoreceptor mosaic of participant S04. (**D**) Photoreceptor mosaic of an age-matched control (C05) to participant S04 from the same retinal eccentricity.

Figure 6.7: Photoreceptor mosaic of participant S06 in an area of degenerated retina showing a lack of unambiguous cones and loss of the regular photoreceptor packing arrangement. (**A**) 30° short-wavelength fundus autofluorescence image, white box represents the retinal area corresponding to the photoreceptor mosaic in C. (**B**) Spectral-domain optical coherence tomography. The horizontal white arrow indicates the retinal area that corresponds to the white box in **A**. (**C**) Photoreceptor mosaic of participant S06. (**D**) Photoreceptor mosaic of an age-matched control (C02) to participant S06 from the same retinal eccentricity. Asterisk represents the foveal centre.

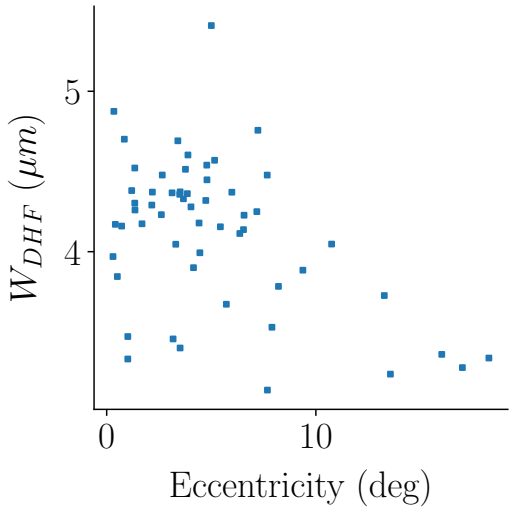


Figure 6.8: The relationship between retinal eccentricity and size (full width at half maximum) of the discrete hyper-reflective foci (W_{DHF}). Each datum represents the W_{DHF} from a single frame. The mean W_{DHF} for all participants and all retinal eccentricities is 4.1 µm (SD = 0.5 µm). The retinal eccentricity dependent variations in size that would be expected if the discrete hyper-reflective foci were cone photoreceptors is not observed.

within images of the photoreceptor mosaic, an average image of hyper-reflective foci was created by stacking the brightest hyper-reflective foci and averaging through the stack. Discrete hyper-reflective foci were observed in nine participants with Stargardt disease, their size and distribution is illustrated in Figure 6.8. They were identified at retinal eccentricities ranging from 0.3° to 18.4° and lacked a regular packing arrangement (Figure 6.9). The discrete hyper-reflective foci were only identified in areas with evidence of retinal degeneration on SD-OCT, autofluorescence and near-infrared reflectance imaging (Table 6.1). The relative hyper-reflectivity of discrete hyper-reflective foci compared to other hyper-reflective features, such as cone photoreceptors, in participants with Stargardt disease or controls was not quantified.

6.4.1.3 Retinal pigment epithelium mosaic

Typically, dark-field imaging (which optically suppresses signals from the photoreceptor layer) is used to enable direct visualisation of individual RPE cells using an AOSLO. The characteristic appearance of a continuous RPE cell mosaic (hexagonal cells with a dark centre surrounded by a brighter ring made up of discrete spots) was clearly observed in one participant (S09) with an optical gap phenotype

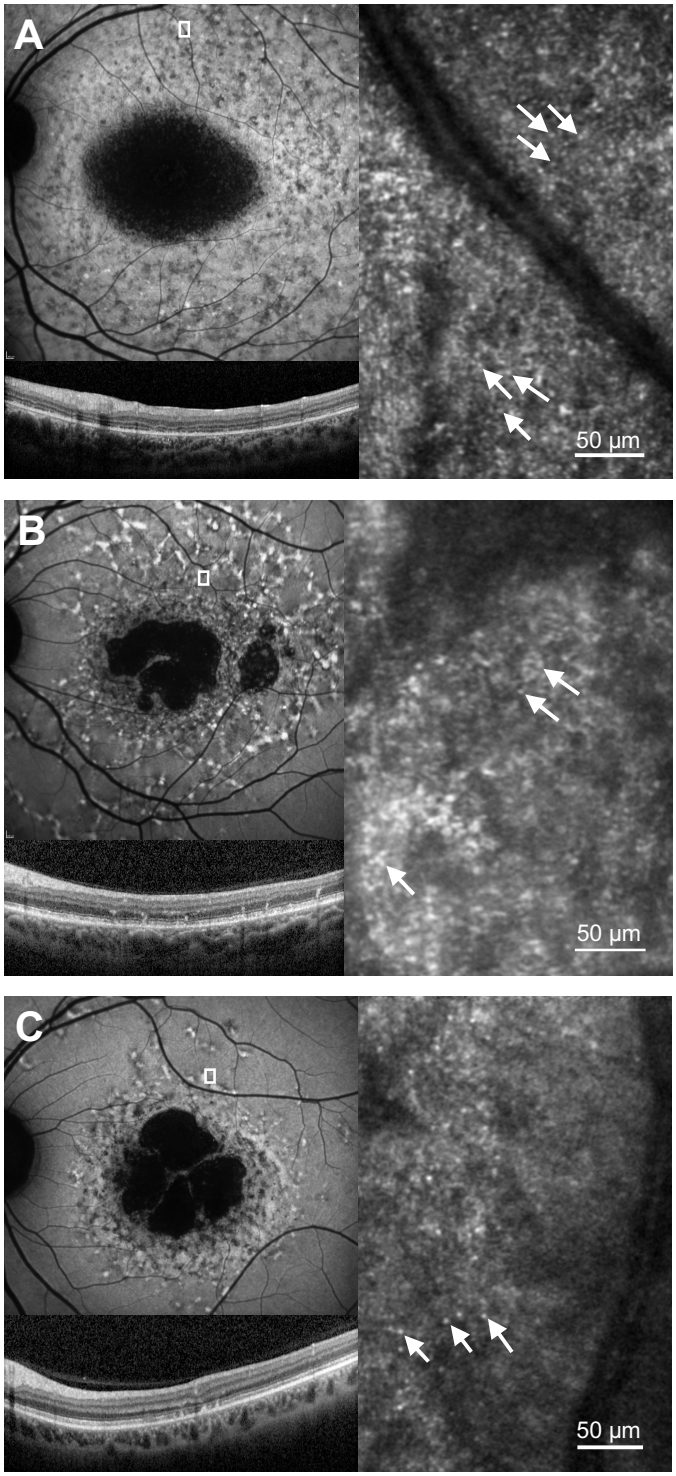


Figure 6.9: Examples of the discrete hyper-reflective foci. 30° short-wavelength fundus autofluorescence (upper left), spectral-domain optical coherence tomography (lower left) and confocal AOSLO (right) images are shown. The white box in the short-wavelength fundus autofluorescence image represents the retinal area corresponding to the photoreceptor mosaic. White arrows indicate examples of the discrete hyper-reflective foci that can be seen throughout each image. (**A**) participant S11, (**B**) participant S14, (**C**) participant S15.

Table 6.1: Discrete hyper-reflective foci in participants with Stargardt disease.

Participant	Eye	Eccentricity (degrees)		SW-FAF	NIR/NI-FAF	SD-OCT	Mean FWHM [SD] (µm)
		Min	Max				
S02	OD	0.3	6.0	Heterogeneous and hypo-AF	Heterogeneous NI reflectance	Disruption and/or loss of ELM and EZ	4.4 [0.2]
S03	OS	4.0	4.0	Heterogeneous AF	Heterogeneous NI reflectance	Loss of ELM and EZ	4.3 [0]
S05	OD	3.4	7.0	Heterogeneous AF	Reduced NI reflectance	Disruption and/or loss of ELM and EZ	4.4 [0.4]
S06	OD	0.5	1.0	Hypo-AF	Heterogeneous NI reflectance	Loss of ELM and EZ	3.6 [0.2]
S09	OS	0.3	6.2	Normal and hypo-AF	Heterogeneous NI reflectance in areas of normal SW-FAF	ONL thinning in areas of normal SW-FAF or hyper-reflective ELM and EZ loss	4.3 [0.1]
S11	OS	13.4	18.4	Heterogeneous AF	Heterogeneous AF	Loss of ELM and EZ	3.4 [0.2]
S12	OS	5.9	8.0	Hypo-AF	Hypo-AF	Hyper-reflective foci within ONL	3.5 [0.2]
S14	OS	7.6	7.6	Heterogeneous AF	Heterogeneous AF	Disruption of ELM and EZ	4.3 [0]
S15	OS	10.5	10.9	Heterogeneous AF	Hypo-AF	Disruption of ELM and EZ	4.0 [0.1]

NIR - near infrared reflectance imaging. NI-FAF - near-infrared fundus autofluorescence, OD - right eye, OS - left eye, SD-OCT - spectral-domain optical coherence tomography, SW-FAF - short-wavelength fundus autofluorescence.

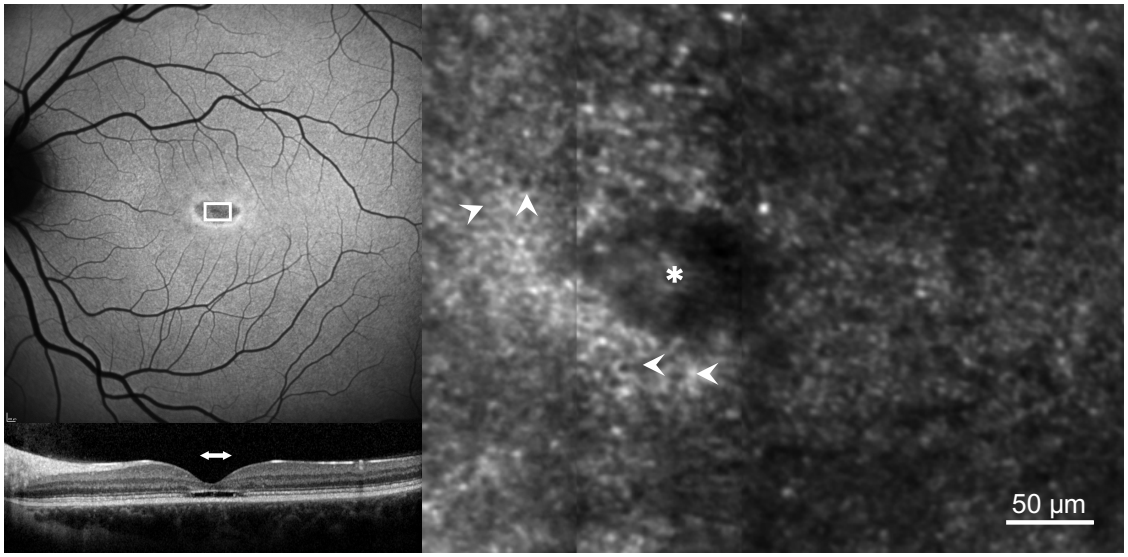


Figure 6.10: Retinal pigment epithelial (RPE) mosaic visible in participant S09. 30° short-wavelength fundus autofluorescence (SW-FAF, upper left), spectral-domain optical coherence tomography (SD-OCT, lower left) and confocal AOSLO (right) images are shown. The white box in the SW-FAF image represents the retinal area corresponding to the photoreceptor mosaic. The horizontal white arrow in the SD-OCT image indicates the retinal area that corresponds to the white box in the SW-FAF image. The white arrowheads in the confocal AOSLO image indicate examples of the RPE cells visible throughout the image. Asterisk represents the foveal centre.

(Figure 6.10). The optical gap phenotype is associated with Stargardt disease and characterised by foveal cavitation that appears to represent a focal loss of ellipsoid zone reflectance on optical coherence tomography [190] (Figure 6.10). Alternative explanations of the observed image features that are consistent with a continuous RPE mosaic include spatial aliasing of optical signals below the resolution limit, or spurious structures introduced by motion blur. Spatial aliasing artefacts seem unlikely because these image features are not visible in AOSLO images of control participants or other participants with Stargardt disease, even though there are the same resolution limits. Artefacts introduced by motion blur also seem an unlikely explanation as these image features are consistently present in all imaging frames rather than just those frames exhibiting specific motion.

No clearly discernible photoreceptors were identifiable within the area of focal ellipsoid zone loss. The area of foveal cavitation was characterised by reduced SW-FAF signal with focal loss of the outer nuclear layer and ellipsoid zone on SD-OCT.

Despite the evidence of retinal degeneration and absence of clearly discernable photoreceptors, best corrected visual acuity was 0.08 LogMAR with a foveal PRL (0.03°) and mean PRL sensitivity of 16.0 dB (SD 4.2 dB) (Table 2.2 and Table 2.6).

6.5 Automated reference frame selection in participants with Stargardt disease

To evaluate the automated reference frame selection algorithm described in chapter 4 (Method 1) with AOSLO images from patients with retinal disease, images from a single patient with Stargardt disease were used. Image sequences from participant S10 collected during one imaging session (at eccentricities between 5° and 13° from the fovea) were selected as they were considered to be representative of a typical patient image sequence. Quantification of the residual motion contained within a chosen reference frame requires a perfect, distortion free reference image. However, obtaining a perfect reference image is not possible with *in vivo* imaging. Therefore, to quantify the difference between Method 1 (the automated reference frame selection algorithm described in Chapter 4) and Method 2 (the default method of reference frame selection using normalised cross correlation described in Section 4.4.1), a power spectrum analysis of the output images (reference frame or averaged image) was performed [123]. A power ratio was calculated as a function of frequency such that a ratio less than one indicates a lower spectral power for Method 1 compared to Method 2 at a given frequency. If Method 1 is better than Method 2 at selecting reference frames and subsequently producing an averaged image, a power ratio of greater than one would be expected.

Averaged images were generated using the reference frame output from each method. Single frames contributed to the averaged image if the variance of the intraframe eye motion (determined using the reference frame) was less than the mean variance of all frames within the image sequence or cluster. Calculating a power ratio requires the output images to be spatially aligned before being transformed using the discrete Fourier Transform. Averaged images and reference frames from the two reference frame selection methods were cropped to a common

area and registered to within 1 pixel using linear translation. Each image was then normalised by subtracting the mean pixel intensity and dividing by the standard deviation of pixel intensities. A power ratio (R_p) was calculated from the radial profile of the magnitude of the discrete Fourier Transform as a function of frequency using the following formula:

$$R_p = \frac{rp_1}{rp_2} \tag{6.2}$$

where rp_1 and rp_2 represent the radial profile of Method 1 and Method 2, respectively. The radial profile of the power spectrum was analysed up to an optical cutoff frequency of approximately 0.37 cycles/pixel based on a system resolution of 2.9 µm with a maximum pupil size of 6 mm. The optical cutoff frequency represents the resolution limit of the system according to the Rayleigh criterion, thus frequencies greater than the cutoff are mainly related to noise. A one sample t-test was used to compare the mean power ratio from Method 1 with a hypothetical value of one (where the power ratio of Method 1 is equivalent to Method 2).

A total of 49 reference frames and 32 averaged images were evaluated from each reference frame selection method (Figure 6.11). The mean power ratios from Method 1 were 1.015 (SD 0.057) for reference frames and 1.042 (SD 0.120) for averaged images, and did not differ significantly from one (p = 0.0817, 95% CI [-0.002, 0.031] and p = 0.0567, 95% CI [-0.001, 0.085], respectively).

6.6 Automated cone photoreceptor identification in participants with Stargardt disease

To evaluate performance of the algorithm developed for automated cone photoreceptor identification (described in Section 5.7) in patients with retinal disease, images from a single patient with Stargardt disease were used.

The test dataset consisted of eight confocal AOSLO images 96 × 96 pixels in size from participant S04. Participant S04 was selected as they exhibited retinal areas with the least degeneration of all participants with Stargardt disease. The

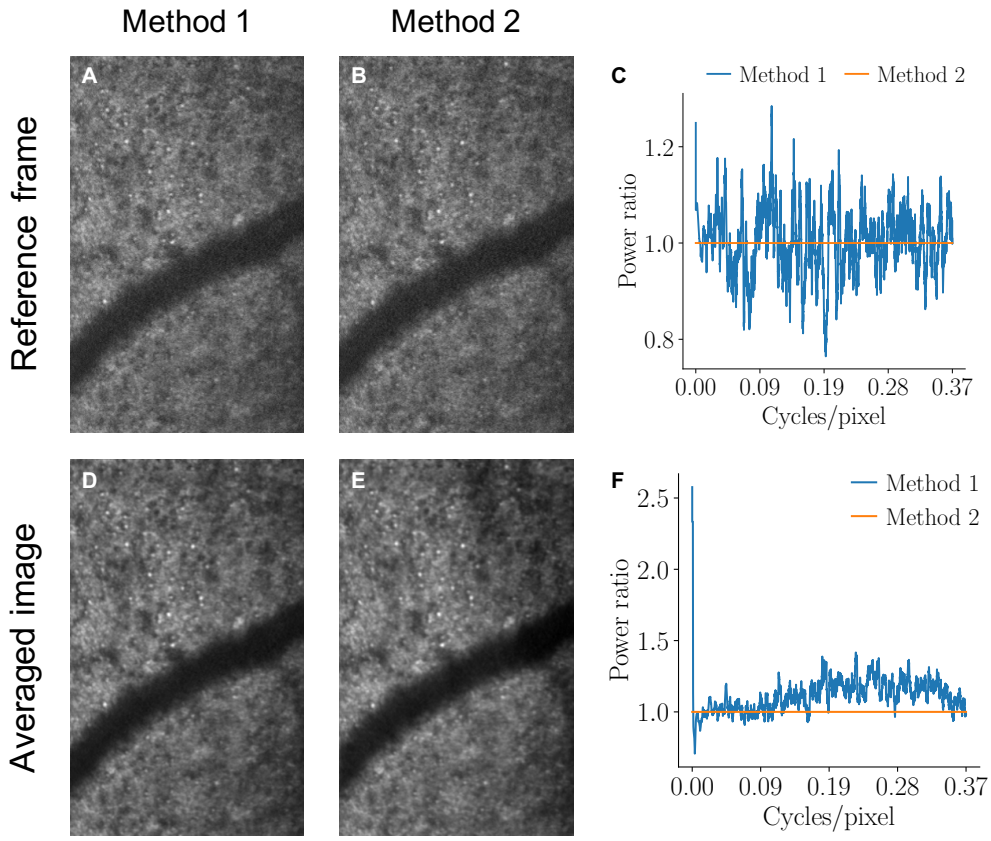


Figure 6.11: An example of the power spectrum analysis used to evaluate automated reference frame selection in a participant with Stargardt disease. A power ratio (**C** and **F**) was used to quantify the difference between Method 1 (the automated reference frame selection algorithm described in Chapter 4) and Method 2 (the default method of reference frame selection using normalised cross correlation described in Section 4.4.1). A power ratio greater than one indicates a higher spectral power at a given frequency. **Top row**: power ratio to quantify differences in the reference frame output by Method 1 (**A**) and Method 2 (**B**). **Bottom row**: power ratio to quantify differences in the averaged image from Method 1 (**D**) and Method 2 (**E**).

convolutional neural network (CNN) developed for automated cone photoreceptor identification was trained and tested on real and synthetic images of normal control participants. Testing the CNN with previously unseen images of the photoreceptor mosaic from areas of substantial retinal degeneration is unlikely to provide a fair evaluation of its performance on images from patients with retinal disease without further training. Furthermore, areas of retinal degeneration without unambiguous cones exacerbate the challenging task of manual cone identification, which is required to provide a ground truth. Images from participant S04 were obtained at eccentricities between 1.6° and 5.5° from the fovea. Cones were manually

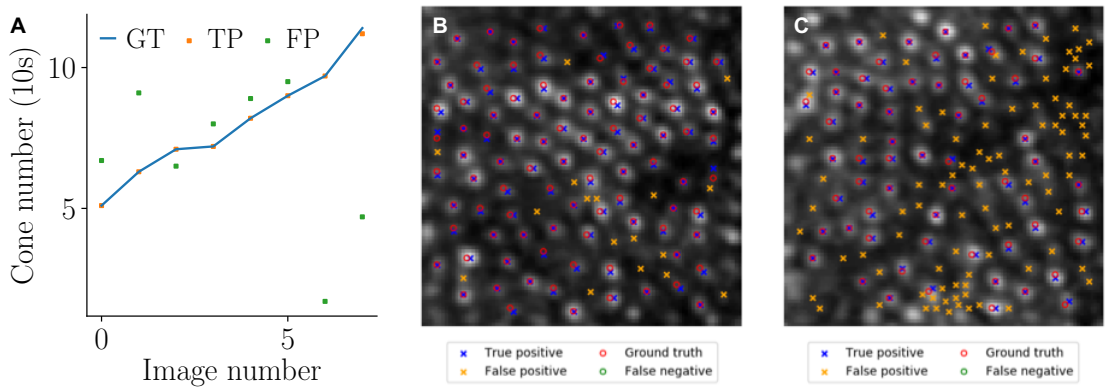


Figure 6.12: Automated cone photoreceptor identification in a participant with Stargardt disease. **A** - Accuracy of automated cone photoreceptor identification. The graph illustrates the number of true positive (TP) and false positive (FP) cones identified in all tested images relative to the ground truth (GT). **B** and **C** - An example of the cone identification results overlaid on the input images.

identified in all images by an experienced grader. The model described in Section 5.7 with the highest accuracy when tested with images from Dataset 2 (described in Section 5.3.2.2) was used.

Mean performance with data from participants with Stargardt disease (Jaccard similarity coefficient 0.544, SD 0.142) was worse when compared to control data (Dataset 2, Jaccard similarity coefficient 0.906, SD 0.0777) due to a higher number of false positives (mean false discovery rate 0.455, SD 0.143, mean true positive rate 0.998, SD 0.006; Figure 6.12), which appear to cluster in areas of reduced intensity in patient images.

6.7 Discussion

This study shows that the clinical heterogeneity observed in Stargardt disease correlates with the findings on confocal AOSLO imaging. The main outcomes of this chapter include the presentation of AOSLO images from a range of retinal phenotypes, visualisation of the RPE mosaic in one participant using confocal AOSLO imaging and the observation of discrete hyper-reflective foci in areas of retinal degeneration that have not previously been described. While a regular photoreceptor mosaic was observed in areas of near normal retina, in areas of

retinal degeneration the mosaic was disorganised with a lack of unambiguous cones. Using confocal AOSLO imaging we also observed structural changes that are not detectable with conventional clinical imaging modalities. The RPE mosaic (visible in one participant with an optical gap phenotype) appeared qualitatively regular in its packing arrangement and (in multiple participants) discrete hyper-reflective foci were observed throughout areas of retinal degeneration. This cohort also represents the first group of participants with retinal disease to successfully undergo retinal imaging using the Oxford AOSLO.

6.7.1 Discrete hyper-reflective foci

Discrete hyper-reflective foci were predominantly associated with disruption or loss of the external limiting membrane and ellipsoid zone on SD-OCT and abnormal autofluorescence and near-infrared reflectance signals (Table 6.1). In one participant (S09) they were observed in an area of normal SW-FAF signal but this was associated with heterogeneous near-infrared reflectance signal and outer nuclear layer thinning on SD-OCT. Due to their small size (mean full width at half maximum 4.1 µm) the discrete hyper-reflective foci were not visible on SD-OCT (lateral resolution 6-11 µm).

The aetiology and significance of the observed discrete hyper-reflective foci are currently unknown. The observation of hyper-reflective structures has been described in AOSLO images from patients with diabetes [191, 192] and central serous chorioretinopathy [193]. Hyper-reflective foci observed on SD-OCT have also been reported in association with age related macular degeneration [88], diabetic macular oedema [89, 90], retinal vein occlusion [91] and type 2 idiopathic macular telangiectasia [92]. Suggested hypotheses regarding the aetiology of these hyper-reflective foci include aggregates of activated microglial cells [88], degenerated photoreceptors [93], and migrating RPE cells [94]. The small size of the hyper-reflective foci observed in this study makes it unlikely that they are migrating RPE cells or activated macrophages. RPE cells demonstrate an eccentricity dependent increase in size [194] and are smallest at the macula with a diameter of between 10-14 µm [195, 196], while macrophages have a diameter of 21 µm [197]. Observation

of the discrete hyper-reflective foci at a retinal eccentricity of 0.3° makes it less likely that they represent degenerated rod photoreceptors as the horizontal diameter of the central rod-free zone extends up to 1.25° from the fovea [25]. It is possible that they could represent degenerated cone photoreceptors, however, the eccentricity dependent variation in size that would be expected of cones [25] was not observed. During light induced photoreceptor death in ABCA4$^{-/-}$ mice, microglia (the resident immune cells of the retina) have been shown to migrate towards the outer retina [101]. Hyper-reflective deposits that disappeared following treatment were observed on SD-OCT in patients with neovascular age related macular degeneration and led Coscas *et al.* to suggest that these features could represent aggregates of activated inflammatory cells [88]. Based on their uniform size, distribution and presence in areas of retinal degeneration, we hypothesize that the discrete hyper-reflective foci are cellular in nature and could represent cellular debris related to the ongoing retinal degeneration. However, further investigation is required in order to determine their precise aetiology.

6.7.2 ABCA4-associated optical gap phenotype

In one participant (S09) with an optical gap phenotype there were no clearly discernible cone photoreceptors observed within the fovea. Despite the lack of cone photoreceptors, there was evidence of visual function. An incongruous relationship between the extent of retinal degeneration and visual acuity has previously been reported in patients with an ABCA4-associated optical gap phenotype, which has been attributed to preservation of photoreceptors at the PRL [190]. Evidence of maintained central visual function (best corrected visual acuity of 0.08 LogMAR and a foveal PRL) in participant S09 supports the hypothesis that photoreceptors are preserved in patients with an ABCA4-associated optical gap phenotype. However, the presence of foveal cavitation on SD-OCT (loss of ellipsoid zone) and the observation of a continuous RPE mosaic within the fovea on confocal AOSLO imaging (which is easier to image without photoreceptors) suggest a loss of photoreceptors. Although not readily identifiable from the confocal AOSLO images in participant

S09, it is possible that some degenerated photoreceptors may remain intact. In such cases, the addition of complementary non-confocal imaging modalities may be useful in identifying preserved photoreceptors.

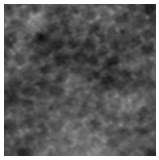
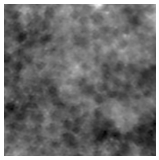
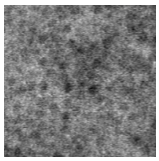
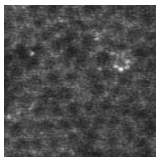
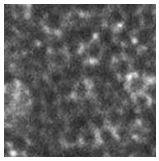
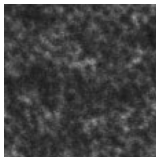
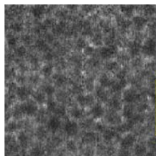
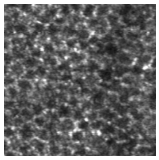
6.7.3 Retinal pigment epithelium mosaic

Although dark-field imaging is typically required to directly visualise individual RPE cells, in one participant (S09) structures consistent with a continuous RPE mosaic were observed within the fovea. To date, there has been one report describing the RPE cell mosaic in patients with Stargardt disease using AO *in vivo* autofluorescence imaging [113]. The RPE mosaic described by Song *et al.* [113] was visualised in the retinal periphery in areas with normal photoreceptors. However, in more central areas with disrupted cone and rod photoreceptors abnormal autofluorescent structures appeared to be more consistent with photoreceptor reflectance rather than RPE cells.

It has been suggested that due to the strong scattering from photoreceptors and the limited axial resolution of the AOSLO, the RPE cell mosaic is not usually apparent [198]. When the photoreceptor layer is compromised or absent, such as in participant S09 who presented with foveal cavitation, the RPE cell mosaic is revealed. The characteristic appearance of RPE cells on high-resolution *in vivo* retinal imaging has been attributed to the cell nucleus (dark centre) and lipofuscin granules located within the cell cytoplasm (brighter surrounding ring made up of discrete spots) [199]. As summarised in Table 6.2, this typical appearance in normal retinas has been observed with short-wavelength [199, 200] and near-infrared [200, 201] autofluorescence, dark-field [47] (non-confocal reflectance imaging), and AO-OCT [202, 203] imaging modalities. The characteristic signals obtained in these different imaging modalities raise a number of puzzling questions.

The signal from autofluorescence imaging with short wavelength light (488 nm excitation) is purported to be derived primarily from lipofuscin [14–16], a by-product of the visual cycle, while with near-infrared light (787 nm excitation) it appears to originate predominantly from melanin within in the RPE and choroid

Table 6.2: Adaptive optics images of the retinal pigment epithelium mosaic.

AO imaging modality	Example images	
Dark-field	 [47]	 [201]
NI-AF	 [201]	 [200]
SW-AF	 [200]	 [199]
OCT	 [202]	 [203]

Numbers in parentheses are citations and indicate the source of the images. AF - autofluorescence, AO - adaptive optics, NI - near-infrared, OCT - optical coherence tomography, SW - short-wavelength.

[16–18]. The ultrastructure of RPE cells reveals melanin granules located within or at the base of the apical processes, a basal nucleus, and lipofuscin granules distributed throughout the cytoplasm [204].

Why do RPE cells reveal this characteristic appearance with multimodal *in vivo* high-resolution retinal imaging? The dark centre has been suggested to be caused by the presence of the cell nucleus. However, due to its basal position within the cell relative to the melanin and lipofuscin granules, one would expect differences between RPE cell images from dark field (reflectance) and autofluorescence images. It is possible that the dark centre could be caused by light absorption from melanin granules [205], but this does not explain the aetiology of the brighter surrounding ring, nor does it account for the similar appearance of RPE cells on *in vivo*

multispectral autofluorescence imaging [200]. Burns *et al.* tested whether light generated by fluorescence within the RPE is guided through cone photoreceptors and found that it is not [206], suggesting that light from deeper retinal layers is unlikely to be recaptured by cones. The discrete spots found within the bright ring surrounding the RPE cells are therefore unlikely to represent cone photoreceptors. The aetiology and significance of the characteristic appearance of RPE cells observed with multimodal high-resolution *in vivo* retinal imaging remains unclear.

6.7.4 Automated reference frame selection in participants with Stargardt disease

Evaluation of the automated reference frame selection algorithm described in Chapter 4 with images from a participant with Stargardt disease did not show a statistically significant improvement in image quality of reference frames or averaged images. Without high quality template-independent image quality metrics, objectively and automatically assessing image quality to select a reference frame is challenging. Further work to objectively evaluate additional template-independent image quality metrics would be required to determine the efficacy of this approach for selecting superior reference frames in participants with retinal disease compared to the default method of reference frame selection used with the Oxford AOSLO.

6.7.5 Automated cone photoreceptor identification in participants with Stargardt disease

Performance of the algorithm developed for automated cone photoreceptor identification (see Section 5.7) with images from a patient with Stargardt disease was worse when compared to control data. Although the true positive rate (sensitivity) was high (0.998, SD 0.006), performance was poorer with patient images due to a high false discovery rate. Areas of reduced intensity in the patient images seem to be associated with false positives and may be a key factor of the poor performance (Figure 6.12). Dark spaces within the photoreceptor mosaic of control retinas have been shown to contain poorly reflective cones with normal sensitivity to light [207]. Accurate identification of these persistently dark cones will require alternative criteria to be

developed, which may include new imaging protocols (e.g. non-confocal imaging, longitudinal imaging) and functional measures (e.g. cone-based microperimetry). Dark cones, which have also been identified in images from patients with retinal disease [208], are just one of the reasons why any cone counting, including manual marking of cone photoreceptors, from patient images is so challenging.

The deep learning classification model from this experiment was trained using synthetic confocal AOSLO images containing only cone photoreceptors and real images of the rod free zone at an eccentricity 0.65° from fixation. The model will therefore not have learned feature representations for non-cone structures that are present in confocal AOSLO images outside the rod free zone. Allowing the model to learn this additional feature information by transfer learning with real confocal AOSLO images from retinal eccentricities greater than 0.65° may improve accuracy.

In patients with retinal disease where cone loss has occurred, enlargement of the remaining photoreceptors occurs [45]. This may affect model accuracy in images from patients with retinal disease. Additional transfer learning with patient images may improve model accuracy of cone identification in these images further. The use of images from non-confocal imaging geometries, such as split-detection, may be particularly useful in retinal diseases where outer segment involvement in the disease process is thought to occur prior to the rest of the photoreceptor [46, 209, 210]. Additionally it may help to determine whether reflectivity in confocal imaging is from cone or non-cone structures. Reducing the gap between synthetic and real datasets by updating the simulation used to produce synthetic data to include rod photoreceptors, retinal vasculature and diseased retinas may improve the model's performance with real data.

6.8 Conclusions

This study contributes to the number of patients with Stargardt disease imaged using the AOSLO, and specifically includes participants with heterogeneous retinal phenotypes. The clinical utility of AOSLO images from patients with Stargardt disease remains unclear at present, mainly because of difficulties in interpreting

image features. However, there are clear directions for improving clinical utility through research such as with multimodal high-resolution retinal imaging and the imaging of more patients. The specific issues related to the processing and analysis of AOSLO images tackled in earlier chapters in this thesis (reference frame selection and automated cone identification) have not yet shown clear improvements. However, in both cases, the approaches taken here have identified clear next steps. AOSLO imaging in general shows great promise in ophthalmology, especially for helping us to understand disease mechanisms and to provide the cell-specific biomarkers that will be required for new therapies currently under development. For example, high-resolution imaging of the photoreceptor and RPE mosaic may help to determine whether primary RPE or photoreceptor degeneration occurs in Stargardt disease.

7

Summary and conclusions

Contents

7.1 Chapter 2: The participant cohort 185
7.2 Chapter 3: Participant selection and image acquisition with the AOSLO 185
7.3 Chapter 4: Automated reference frame selection . . . 187
7.4 Chapter 5: Automated cone photoreceptor identification using deep learning 187
7.5 Chapter 6: AOSLO imaging in participants with Stargardt disease 189
7.6 Conclusions 189

The overall aim of this thesis was to investigate the adaptive optics scanning laser ophthalmoscope (AOSLO) as a clinical tool and to develop and validate fully automated image processing and analysis pipelines for AOSLO images from patients with retinal disease. The relationship between multimodal conventional clinical and AOSLO imaging at the leading disease edge in participants with Stargardt disease was investigated to determine if adaptive optics (AO) imaging provides additional clinically relevant information. Synthetic data were used to develop and validate automated algorithms for reference frame selection and cone identification. This chapter briefly summarises the main findings of each chapter and discusses the challenges that need to be addressed to promote the clinical

adoption of AO retinal imaging.

7.1 Chapter 2: The participant cohort

The improvement in resolution that is achieved by using an AOSLO has already provided evidence of disease progression before visible changes on conventional retinal imaging or a measurable reduction in visual function [5]. The purpose of the study described in this chapter was to generate high-resolution *in vivo* retinal images from patients with Stargardt disease (reported in Chapter 6). The aim of this chapter was to characterise the phenotypic heterogeneity in a cohort of participants with Stargardt disease using multimodal conventional retinal imaging.

The results showed that the retinal phenotype was variable in this cohort of participants with Stargardt disease. Findings from examination of the leading disease edge are consistent with previously reported studies and suggest that outer nuclear layer thickness, determined using spectral-domain optical coherence tomography (SD-OCT), may be the preferred biomarker for monitoring early patterns of degeneration. Analysis of SD-OCT and fundus autofluorescence data support the hypothesis that cone photoreceptor degeneration is a primary phenomenon and does not only occur secondary to RPE atrophy. Hyper-reflective foci were observed on SD-OCT and we hypothesise that they may represent photoreceptor mediated activation of immune cells as a consequence of neurosensory retinal degeneration. This chapter describes the retinal phenotype of participants with Stargardt disease who underwent high-resolution retinal imaging with the Oxford AOSLO (the results of which are presented in Chapter 6).

7.2 Chapter 3: Participant selection and image acquisition with the AOSLO

There are a number of challenges associated with *in vivo* AOSLO imaging, particularly in patients with retinal disease. The development of specific participant selection criteria and image acquisition protocols may help to mitigate these

challenges and optimise AOSLO image quality in patients with retinal disease. This chapter describes recommendations for future participant selection criteria and AOSLO image acquisition protocols that were learned and developed through the experience and expertise gained in AOSLO imaging of participants with retinal disease.

A summary of the recommended participant selection criteria for future studies with the Oxford AOSLO in its current configuration include:

1. Absence of significant refractive error.
2. Absence of ocular media opacity.
3. Fixation stability.
4. Preferred retinal locus within 10° of the foveal centre.
5. The presence of sufficient retinal structure within the retinal region of interest.

A summary of the recommended image acquisition procedures for future studies include:

1. Ensure that the participant is seated comfortably with a straight back and neck.
2. Correct defocus and astigmatism if necessary.
3. Use lubricating eye drops prophylactically.
4. Match the size of the fixation target to the participant's visual acuity.
5. Avoid dark adaptation.
6. Maintain active participation during the imaging session.
7. Use fundus-guided imaging.
8. Account for variations in the local retinal topography.

7.3 Chapter 4: Automated reference frame selection

Reference frames are required to measure and correct intraframe eye motion in order to produce averaged AOSLO images with a high signal to noise ratio. Reference frame selection has traditionally been performed manually by an experienced image grader. However, this is both time consuming and subjective. The aim of this chapter was to develop and validate an unsupervised automated reference frame selection algorithm using synthetic data that considers the unique challenges associated with processing AOSLO images from patients with retinal disease.

The automated reference frame selection algorithm was developed and validated using synthetic images of the cone mosaic. Although there was no statistically significant difference between selected reference frames using the automated reference frame selection algorithm developed in this project (Method 1) or the default method used for reference frame selection with the Oxford AOSLO (Method 2) when compared to a perfect reference image, averaged images generated using Method 1 had greater similarity to a perfect reference image when compared to those from Method 2. Evaluating the metric used for detection of intraframe motion revealed a greater accuracy when detecting large intraframe shear artefacts (greater than 10°, equivalent to a speed of 1.02°/s) and no correlation with normal simulated fixational eye movements. By developing and validating a fully automated reference frame selection algorithm, this work has contributed to the field by enabling increased automation of the AOSLO image processing pipeline.

7.4 Chapter 5: Automated cone photoreceptor identification using deep learning

Identification of individual photoreceptors from AOSLO images is often a prerequisite for quantifying changes in the photoreceptor mosaic. Photoreceptor identification is typically performed manually, which is both subjective and time consuming.

The significant effort required for the analysis of these images is a major rate-limiting step in the clinical application of *in vivo* AO retinal imaging. Although existing automated cone detection algorithms demonstrate good performance for the specific problem they were designed for, their generalisability remains limited. Deep learning provides an effective way of automatically extracting and learning features directly from the image data, potentially increasing generalisability. Traditional deep learning approaches require large volumes of high-quality training data with accompanying annotations. Not only are these large datasets difficult to acquire but they are not practical for retinal diseases with a low prevalence. The aim of this chapter was to investigate whether synthetic data could be used to develop a convolutional neural network (CNN) for the automated identification of cone photoreceptors.

This study demonstrated good agreement between the developed CNN, trained with synthetic and real data, and the current gold standard of manual marking. The developed CNN also marginally outperformed the current state-of-the-art automated cone detection algorithms. Using synthetic images of the cone photoreceptor mosaic provided sufficient data to train a newly developed CNN and leverage the learned feature maps with manually marked real images to identify cone photoreceptors. The success of this approach demonstrates the suitability of using synthetic data to overcome the limited availability of annotated real datasets. Synthetic data representing photoreceptor mosaics from diseased eyes may help to develop and validate tools for automated cone photoreceptor identification in patients with retinal disease. This work has contributed to the field through the development and validation of an automated cone detection algorithm that marginally outperformed the current state-of-the-art cone detection methods. Furthermore, it also demonstrates the suitability of synthetic data to overcome the limited availability of annotated real datasets for the development of cone identification algorithms.

7.5 Chapter 6: AOSLO imaging in participants with Stargardt disease

Stargardt disease is the commonest inherited macular dystrophy in adults and children [51], and although there are currently no approved treatments, there are a number of ongoing treatment trials investigating gene replacement therapy, cell-based therapy and pharmacotherapy [51]. With the development of these new therapies cell-specific biomarkers, which can be provided by AO retinal imaging systems, are becoming increasingly important for monitoring treatment outcomes. To date, there are a limited number of studies reporting the results of AOSLO imaging in patients with Stargardt disease. The aim of this chapter was to report the results of AOSLO imaging in the cohort of participants with Stargardt disease described in Chapter 2.

Cone photoreceptors were readily identifiable in areas of near normal retinal structure, however, in areas of retinal degeneration the photoreceptor mosaic was disorganised and lacked unambiguous cones. Hyper-reflective spots, termed discrete hyper-reflective foci, were identified in AOSLO images from participants with Stargardt disease. Based on their uniform size, distribution and presence in areas of retinal degeneration, we hypothesize that the discrete hyper-reflective foci are cellular in nature and could represent cellular debris related to ongoing retinal degeneration. Finally, a continuous RPE mosaic within the fovea (which typically requires dark-field imaging for visualisation) was observed in one participant with foveal cavitation on SD-OCT. This work has contributed to the field by reporting the results of AOSLO imaging in patients with Stargardt disease, contributing to the limited number of published studies to date.

7.6 Conclusions

Despite the advantages of AO retinal imaging, its widespread clinical adoption remains limited due to a number of challenges. Technical challenges include the development of AO imaging systems that provide a wider overall field of view, are

more robust to eye motion and ocular media opacity, are less sensitive to patient movement, and are quick and easy to use. Image processing challenges include the development of robust image processing pipelines and automated analysis algorithms suitable for images from patients with retinal disease. The interpretation of AO images in patients with retinal disease can be challenging. Further research using AO retinal imaging systems is required to improve our understanding of the clinical and functional consequences of observed structural changes in different retinal diseases. The barriers to entry for the clinical adoption of AO retinal imaging systems are high due to the limited access to existing systems, the high cost, and technical expertise required for research grade systems. Working with commercial partners to develop and promote AO retinal imaging systems and to reduce the barriers to entry for their clinical adoption is important. Similar to the development of optical coherence tomography technologies, a major driver of the clinical demand for AO retinal imaging will be the development of novel treatments that require cell-based biomarkers of disease. The widespread clinical adoption of AO retinal imaging will require these challenges to be addressed.

References

[1] Lynne Pezzullo et al. “The economic impact of sight loss and blindness in the UK adult population”. In: *BMC Health Services Research* 18.1 (2018).

[2] Gerald Liew, Michel Michaelides, and Catey Bunce. “A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010”. In: *BMJ Open* 4.2 (2014), pp. 1–6.

[3] Jugnoo S. Rahi and Noriko Cable. “Severe visual impairment and blindness in children in the UK”. In: *Lancet* 362.9393 (Oct. 2003), pp. 1359–1365.

[4] V. Manjunath et al. “Wide-field imaging and OCT vs clinical evaluation of patients referred from diabetic retinopathy screening”. In: *Eye (Basingstoke)* 29.3 (2015), pp. 416–423.

[5] Lynn W. Sun et al. “Assessing photoreceptor structure in retinitis pigmentosa and usher syndrome”. In: *Investigative Ophthalmology and Visual Science* 57.6 (2016), pp. 2428–2442.

[6] Katherine E. Talcott et al. “Longitudinal study of cone photoreceptors during retinal degeneration and in response to ciliary neurotrophic factor treatment”. In: *Investigative Ophthalmology and Visual Science* 52.5 (2011), pp. 2219–2226.

[7] Helga Kolb. *Photoreceptors*. 2013. URL: `https://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/photoreceptors/` (visited on 02/22/2020).

[8] Robert S. Molday and Orson L. Moritz. “Photoreceptors at a glance”. In: *Journal of Cell Science* 128.22 (2015), pp. 4039–4045.

[9] Rick Cote. *Photoreceptor Phosphodiesterase (PDE6): A G-Protein-Activated PDE Regulating Visual Excitation in Rod and Cone Photoreceptor Cells*. Ed. by J.A. Beavo, S.H. Francis, and M.D. Houslay. CRC Press, Dec. 2006. Chap. 8, pp. 165–193.

[10] Bruce M. Koeppen and Bruce A. Stanton. *Berne & Levy Physiology*. 7th ed. Philadelphia: Elsevier, 2017, p. 129. eprint: `arXiv:1011.1669v3`.

[11] Olaf Strauss. *The Retinal Pigment Epithelium*. URL: `https://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/the-retinal-pigment-epithelium/` (visited on 11/01/2019).

[12] D J Spalton, J Marshall, and V Chong. “The normal retina, retinal imaging and the interpretation of pathological changes”. In: *Atlas of clinical ophthalmology* (2005). Ed. by David Spalton, Roger Hunter, and Hitchings Paul, pp. 397–436.

[13] Ian Harvey. *Raster-scan.svg*. 2018.

[14] F. C. Delori et al. “In vivo fluorescence of the ocular fundus exhibits retinal pigment epithelium lipofuscin characteristics”. In: *Investigative Ophthalmology and Visual Science* 36.3 (1995), pp. 718–729.

[15] Janet R. Sparrow et al. “The bisretinoids of retinal pigment epithelium”. In: *Progress in Retinal and Eye Research* 31.2 (2012), pp. 121–135.

[16] Janet Sparrow and Tobias Duncker. “Fundus Autofluorescence and RPE Lipofuscin in Age-Related Macular Degeneration”. In: *Journal of Clinical Medicine* 3.4 (2014), pp. 1302–1321.

[17] Claudia N. Keilhauer and François C. Delori. “Near-infrared autofluorescence imaging of the fundus: Visualization of ocular melanin”. In: *Investigative Ophthalmology and Visual Science* 47.8 (2006), pp. 3556–3564.

[18] HaeWon Jung et al. “Longitudinal adaptive optics fluorescence microscopy reveals cellular mosaicism in patients”. In: *JCI Insight* 4.6 (2019).

[19] NHS Digital. *Hospital Episode Statistics for England. Outpatient statistics, 2018 - 2019*. 2019. URL: `https://digital.nhs.uk/data-and-information/publications/statistical/hospital-outpatient-activity/2018-19` (visited on 10/29/2019).

[20] D Huang et al. “Optical coherence tomography”. In: *Science* 254.5035 (Nov. 1991), pp. 1178–81.

[21] Tomas Ilginis, Jonathan Clarke, and Praveen J. Patel. “Ophthalmic imaging”. In: *British Medical Bulletin* 111.1 (2014), pp. 77–88.

[22] Marco Lombardo and Giuseppe Lombardo. “Wave aberration of human eyes and new descriptors of image optical quality and visual performance”. In: *Journal of Cataract and Refractive Surgery* 36.2 (2010), pp. 313–331.

[23] Chad Dockter and Chris Freed. “How Blur Light Affects The Eye and Body: A Close Look at Our Eyes”. In: *How to Save Your Eyes in the Digital Age: The Handbook for Eye Care and Electronics* ().

[24] Stephen A. Burns et al. “Adaptive optics imaging of the human retina”. In: *Progress in Retinal and Eye Research* 68.August (2019), pp. 1–30.

[25] Christine A. Curcio et al. “Human photoreceptor topography”. In: *Journal of Comparative Neurology* 292.4 (1990), pp. 497–523.

[26] Laura K. Young et al. “Compact, modular and in-plane AOSLO for high-resolution retinal imaging”. In: *Biomedical Optics Express* 9.9 (2018), p. 4275.

[27] Enrique J. Fernández, Ignacio Iglesias, and Pablo Artal. “Closed-loop adaptive optics in the human eye”. In: *Optics Letters* 26.10 (2001), p. 746.

[28] Jasdeep S. Gill, Mariya Moosajee, and Adam M. Dubis. “Cellular imaging of inherited retinal diseases using adaptive optics.” In: *Eye (London, England)* (June 2019).

[29] Junzhong Liang, David R. Williams, and Donald T. Miller. “Supernormal vision and high-resolution retinal imaging through adaptive optics”. In: *Journal of the Optical Society of America A* 14.11 (1997), p. 2884.

[30] Jungtae Rha et al. “Adaptive optics flood-illumination camera for high speed retinal imaging”. In: *Optics Express* 14.10 (2006), p. 4552.

[31] Austin Roorda and David R. Williams. “The arrangement of the three cone classes in the living human eye”. In: *Nature* 397.6719 (1999), pp. 520–522.

[32] Austin Roorda et al. “Adaptive optics scanning laser ophthalmoscopy.” In: *Optics express* 10.9 (May 2002), pp. 405–12.

[33] Stephen A. Burns et al. “Contrast improvement of confocal retinal imaging by use of phase-correcting plates”. In: *Optics Letters* 27.6 (2002), p. 400.

[34] Daniel X. Hammer et al. “Adaptive optics scanning laser ophthalmoscope for stabilized retinal imaging”. In: *Optics Express* 14.8 (2006), p. 3354.

[35] Mircea Mujat et al. “Compact adaptive optics line scanning ophthalmoscope”. In: *Optics Express* 17.12 (2009), p. 10242.

[36] B Hermann et al. “Adaptive-optics ultrahigh-resolution optical coherence tomography.” In: *Optics letters* 29.18 (Sept. 2004), pp. 2142–4.

[37] Robert J. Zawadzki et al. “Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal in vivo imaging”. In: *Optics Express* 13.21 (2005), p. 8532.

[38] Yan Zhang et al. “Adaptive optics parallel spectral domain optical coherence tomography for imaging the living retina”. In: *Optics Express* 13.12 (2005), p. 4792.

[39] Brian Krawitz, Shelley Mo, and Richard B. Rosen. “Adaptive Optics Retinal Imaging: Applications and Clinical Implications”. In: *Retinal Physician* 13.May (2016), pp. 50–54.

[40] Chris Dainty. “Adaptive Optics”. In: *Optical Imaging and Microscopy: Techniques and Advanced Systems*. Ed. by Peter Török and Fu-Jen Kao. 2003rd ed. Springer Series in Optical Sciences. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. Chap. 10, pp. 235–255.

[41] Alfredo Dubra et al. “Noninvasive imaging of the human rod photoreceptor mosaic using a confocal adaptive optics scanning ophthalmoscope”. In: *Biomedical Optics Express* 2.7 (2011), p. 1864.

[42] David Merino et al. “Observation of cone and rod photoreceptors in normal subjects and patients using a new generation adaptive optics scanning laser ophthalmoscope”. In: *Biomedical Optics Express* 2.8 (2011), p. 2189.

[43] Donald T. Miller et al. “Images of cone photoreceptors in the living human eye”. In: *Vision Research* 36.8 (1996), pp. 1067–1079.

[44] Nicole M. Putnam et al. “Modeling the foveal cone mosaic imaged with adaptive optics scanning laser ophthalmoscopy”. In: *Optics Express* 18.24 (2010), p. 24902.

[45] Katie M. Litts et al. “Photoreceptor-Based Biomarkers in AOSLO Retinal Imaging”. In: *Investigative ophthalmology & visual science* 58.6 (2017), BIO255–BIO267.

[46] Drew Scoles et al. “In Vivo Imaging of Human Cone Photoreceptor Inner Segments”. In: *Investigative Ophthalmology and Visual Science* 55.7 (2014), pp. 4244–4251.

[47] Drew Scoles, Yusufu N. Sulai, and Alfredo Dubra. "In vivo dark-field imaging of the retinal pigment epithelium cell mosaic". In: *Biomedical Optics Express* 4.9 (2013), p. 1710.

[48] A. Guevara-Torres, D. R. Williams, and J. B. Schallek. "Origin of cell contrast in offset aperture adaptive optics ophthalmoscopy". In: *Optics Letters* 45.4 (2020), p. 840.

[49] Gunnar Farnebäck. "Two-frame motion estimation based on polynomial expansion". In: *Lecture Notes in Computer Science* 2749.1 (2003), pp. 363–370.

[50] Kurt Spiteri Cornish et al. "The Epidemiology of Stargardt Disease in the United Kingdom". In: *Ophthalmology Retina* 1.6 (2017), pp. 508–513.

[51] Preena Tanna et al. "Stargardt disease: Clinical features, molecular genetics, animal models and therapeutic options". In: *British Journal of Ophthalmology* 101.1 (Jan. 2017), pp. 25–30.

[52] Heidi L. Schulz et al. "Mutation spectrum of the ABCA4 gene in 335 stargardt disease patients from a multicenter German cohort—impact of selected deep intronic variants and common SNPs". In: *Investigative Ophthalmology and Visual Science* 58.1 (2017), pp. 394–403.

[53] Yaroslav Tsybovsky, Robert S Molday, and Krzysztof Palczewski. "The ATP-binding cassette transporter ABCA4: structural and functional properties and role in retinal disease." In: *Advances in experimental medicine and biology* 703 (2010), pp. 105–25.

[54] Robert K. Koenekoop. "The gene for Stargardt disease, ABCA4, is a major retinal gene: A mini-review". In: *Ophthalmic Genetics* 24.2 (2003), pp. 75–80.

[55] Robert S. Molday. "ATP-binding cassette transporter ABCA4: Molecular properties and role in vision and macular degeneration". In: *Journal of Bioenergetics and Biomembranes* 39.5-6 (2007), pp. 507–517.

[56] Curtis R. Vogel et al. "Retinal motion estimation in adaptive optics scanning laser ophthalmoscopy". In: *Optics InfoBase Conference Papers* 5688.2 (2005), pp. 145–151.

[57] Girish Kumar and Susana T.L. Chung. "Characteristics of fixational eye movements in people with macular disease". In: *Investigative Ophthalmology and Visual Science* 55.8 (2014), pp. 5125–5133.

[58] Michael D. Crossland, David P. Crabb, and Gary S. Rubin. "Task-specific fixation behavior in macular disease". In: *Investigative Ophthalmology and Visual Science* 52.1 (2011), pp. 411–416.

[59] Michael D. Crossland, Louise E. Culham, and Gary S. Rubin. "Fixation stability and reading speed in patients with newly developed macular disease." In: *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)* 24.4 (2004), pp. 327–333.

[60] Preena Tanna et al. "Reliability and repeatability of cone density measurements in patients with stargardt disease and RPGR-associated retinopathy". In: *Investigative Ophthalmology and Visual Science* 58.9 (2017), pp. 3608–3615.

[61] Pooja Godara et al. "Adaptive optics retinal imaging: Emerging clinical applications". In: *Optometry and Vision Science* 87.12 (2010), pp. 930–941.

[62] Letizia Mariotti and Nicholas Devaney. "Performance analysis of cone detection algorithms." In: *Journal of the Optical Society of America. A, Optics, image science, and vision* 32.4 (Apr. 2015), pp. 497–506.

[63] Kaccie Y. Li and Austin Roorda. "Automated identification of cone photoreceptors in adaptive optics retinal images". In: *Journal of the Optical Society of America A* 24.5 (2007), p. 1358.

[64] Bai Xue et al. "Photoreceptor counting and montaging of en-face retinal images from an adaptive optics fundus camera". In: *Journal of the Optical Society of America A* 24.5 (2007), p. 1364.

[65] Danuta M. Bukowska et al. "Semi-automated identification of cones in the human retina using circle Hough transform". In: *Biomedical Optics Express* 6.12 (2015), p. 4676.

[66] Robert F. Cooper et al. "Automatic detection of modal spacing (Yellott's ring) in adaptive optics scanning light ophthalmoscope images". In: *Ophthalmic and Physiological Optics* 33.4 (2013), pp. 540–549.

[67] Stephanie J. Chiu et al. "Automatic cone photoreceptor segmentation using graph theory and dynamic programming". In: *Biomedical Optics Express* 4.6 (2013), p. 924.

[68] Mital Shah et al. "Investigating the clinical use of adaptive optics scanning laser ophthalmoscopy in patients with Stargardt disease". In: *Investigative Ophthalmology & Visual Science*. Vol. 60. 9. July 2019, p. 4580.

[69] K. Stargardt. "Über familiäre, progressive Degeneration in der Maculagegend des Auges". In: *Albrecht von Græfe's Archiv für Ophthalmologie* 71.3 (1909), pp. 534–550.

[70] Catherine A. Cukras et al. "Centrifugal expansion of fundus autofluorescence patterns in Stargardt disease over time". In: *Archives of Ophthalmology* 130.2 (2012), pp. 171–179.

[71] Artur V. Cideciyan et al. "ABCA4 disease progression and a proposed strategy for gene therapy". In: *Human Molecular Genetics* 18.5 (2009), pp. 931–941.

[72] Stanley Lambertus et al. "Early-onset stargardt disease: Phenotypic and genotypic characteristics". In: *Ophthalmology* 122.2 (2015), pp. 335–344.

[73] Ruofan Connie Han et al. "Optimisation of dark adaptation time required for mesopic microperimetry". In: *British Journal of Ophthalmology* 103.8 (2019), pp. 1092–1098.

[74] Selig Hecht, Charles Haig, and Aurin M. Chase. "The influence of light adaptation on subsequent dark adaptation of the eye". In: *Journal of General Physiology* 20.6 (1937), pp. 831–850.

[75] Gehua Yang et al. "Registration of challenging image pairs: Initialization, estimation, and decision". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.11 (2007), pp. 1973–1989.

[76] Monika Fleckenstein et al. "Fundus autofluorescence and spectral-domain optical coherence tomography characteristics in a rapidly progressing form of geographic atrophy". In: *Investigative Ophthalmology and Visual Science* 52.6 (2011), pp. 3761–3766.

[77] Anthony G. Robson et al. “Comparison of fundus autofluorescence with photopic and scotopic fine matrix mapping in patients with retinitis pigmentosa: 4- to 8-year follow-up”. In: *Investigative Ophthalmology and Visual Science* 53.10 (2012), pp. 6187–6195.

[78] Rupert W. Strauss et al. “Comparison of Short-Wavelength Reduced-Illuminance and Conventional Autofluorescence Imaging in Stargardt Macular Dystrophy”. In: *American Journal of Ophthalmology* 168 (2016), pp. 269–278.

[79] Janet R Sparrow et al. “Progress in Retinal and Eye Research Lessons learned from quantitative fundus auto fl uorescence”. In: *Progress in Retinal and Eye Research* May (2019), p. 100774.

[80] Giovanni Staurenghi et al. “Proposed lexicon for anatomic landmarks in normal posterior segment spectral-domain optical coherence tomography: The IN • OCT consensus”. In: *Ophthalmology* 121.8 (2014), pp. 1572–1578.

[81] Richard F. Spaide and Christine A. Curcio. “Anatomic Correleate to the Bands Seen in Outer Retina by Optical Coherence Tomograpyhy”. In: *Retina* 31.8 (2011), pp. 1609–1619.

[82] Mital Shah et al. “Association of Clinical and Genetic Heterogeneity With BEST1 Sequence Variations”. In: *JAMA Ophthalmology* (Apr. 2020), pp. 1–8.

[83] G. Mylonas et al. “Comparison of retinal thickness measurements and segmentation performance of four different spectral and time domain OCT devices in neovascular age-related macular degeneration”. In: *British Journal of Ophthalmology* 93.11 (2009), pp. 1453–1460.

[84] Robert M. Steinman. “Effect of Target Size, Luminance, and Color on Monocular Fixation”. In: *Journal of the Optical Society of America* 55.9 (1965), p. 1158.

[85] G. Querques et al. “Analysis of retinal flecks in fundus flavimaculatus using optical coherence tomography”. In: *British Journal of Ophthalmology* 90.9 (2006), pp. 1157–1162.

[86] Niloofar Piri, Brooke L.W. Nesmith, and Shlomit Schaal. “Choroidal hyperreflective foci in stargardt disease shown by spectral-domain optical coherence tomography imaging: Correlation with disease severity”. In: *JAMA Ophthalmology* 133.4 (2015), pp. 398–405.

[87] Monika Voigt et al. “Analysis of retinal flecks in fundus flavimaculatus using high-definition spectral-domain optical coherence tomography”. In: *American Journal of Ophthalmology* 150.3 (2010), pp. 330–337.

[88] Gabriel Coscas et al. “Hyperreflective dots: A new spectral-domain optical coherence tomography entity for follow-up and prognosis in exudative age-related macular degeneration”. In: *Ophthalmologica* 229.1 (2012), pp. 32–37.

[89] Matthias Bolz et al. “Optical Coherence Tomographic Hyperreflective Foci. A Morphologic Sign of Lipid Extravasation in Diabetic Macular Edema”. In: *Ophthalmology* 116.5 (2009), pp. 914–920.

[90] Stela Vujosevic et al. “Imaging retinal inflammatory biomarkers after intravitreal steroid and anti-VEGF treatment in diabetic macular oedema”. In: *Acta Ophthalmologica* 95.5 (2017), pp. 464–471.

[91] Ken Ogino et al. “Characteristics of optical coherence tomographic hyperreflective foci in retinal vein occlusion”. In: *Retina* 32.1 (2012), pp. 77–85.

[92] Sönke Baumüller et al. “Outer retinal hyperreflective spots on spectral-domain optical coherence tomography in macular telangiectasia type 2”. In: *Ophthalmology* 117.11 (2010), pp. 2162–2168.

[93] Akihito Uji et al. “Association between hyperreflective foci in the outer retina, status of photoreceptor layer, and visual acuity in diabetic macular edema”. In: *American Journal of Ophthalmology* 153.4 (2012), 710–717.e1.

[94] Carsten Framme, Sebastian Wolf, and Ute Wolf-Schnurrbusch. “Small dense particles in the retina observable by spectral-domain optical coherence tomography in age-related macular degeneration”. In: *Investigative Ophthalmology and Visual Science* 51.11 (2010), pp. 5965–5969.

[95] Charles D. Birnbach et al. “Histopathology and Immunocytochemistry of the Neurosensory Retina in Fundus Flavimaculatus”. In: *Ophthalmology* 101.7 (1994), pp. 1211–1219.

[96] Robert L Steinmetz et al. “Histopathology of Incipient Fundus Flavimaculatus”. In: *Ophthalmology* 98.6 (1991), pp. 953–956.

[97] Artur V. Cideciyan et al. “ABCA4-associated retinal degenerations spare structure and function of the human parapapillary retina”. In: *Investigative Ophthalmology and Visual Science* 46.12 (2005), pp. 4739–4746.

[98] Noemi Lois et al. “Fundus autofluorescence in stargardt macular dystrophy-fundus flavimaculatus”. In: *American Journal of Ophthalmology* 138.1 (2004), pp. 55–63.

[99] Maarjaliis Paavo et al. “Photoreceptor cells as a source of fundus autofluorescence in recessive Stargardt disease.” In: *Journal of neuroscience research* 97.1 (2019), pp. 98–106.

[100] Gurugirijha Rathnasamy et al. “Retinal microglia – A key player in healthy and diseased retina”. In: *Progress in Neurobiology* 173.June 2018 (2019), pp. 18–40.

[101] Hideo Kohno et al. “Photoreceptor proteins initiate microglial activation via toll-like receptor 4 in retinal degeneration mediated by all-trans-retinal”. In: *Journal of Biological Chemistry* 288.21 (2013), pp. 15326–15341.

[102] Qiang Chen et al. “Semi-automatic geographic atrophy segmentation for SD-OCT images”. In: *Biomedical Optics Express* 4.12 (2013), p. 2729.

[103] Kamron N. Khan et al. “Early Patterns of Macular Degeneration in ABCA4-Associated Retinopathy”. In: *Ophthalmology* 125.5 (2018), pp. 735–746.

[104] Simone Kellner et al. “Lipofuscin- and Melanin-related Fundus Autofluorescence in Patients with ABCA4-associated Retinal Dystrophies”. In: *American Journal of Ophthalmology* 147.5 (2009), 895–902.e1.

[105] Tobias Duncker et al. “Correlations among near-infrared and short-wavelength autofluorescence and spectral-domain optical coherence tomography in recessive stargardt disease”. In: *Investigative Ophthalmology and Visual Science* 55.12 (2014), pp. 8134–8143.

[106] Vivienne C. Greenstein et al. “Near-infrared autofluorescence: Its relationship to short-wavelength autofluorescence and optical coherence tomography in recessive stargardt disease”. In: *Investigative Ophthalmology and Visual Science* 56.5 (2015), pp. 3226–3234.

[107] Janet R. Sparrow et al. “Flecks in recessive stargardt disease: Short-wavelength autofluorescence, near-infrared autofluorescence, and optical coherence tomography”. In: *Investigative Ophthalmology and Visual Science* 56.8 (2015), pp. 5029–5039.

[108] Artur V. Cideciyan et al. “Reduced-illuminance autofluorescence imaging in ABCA4-associated retinal degenerations”. In: *Journal of the Optical Society of America A* 24.5 (2007), p. 1457.

[109] Jessica I.W. Morgan et al. “Light-induced retinal changes observed with high-resolution autofluorescence imaging of the retinal pigment epithelium”. In: *Investigative Ophthalmology and Visual Science* 49.8 (2008), pp. 3715–3729.

[110] Shannon M. Conley et al. “Increased cone sensitivity to ABCA4 deficiency provides insight into macular vision loss in Stargardt’s dystrophy”. In: *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1822.7 (2012), pp. 1169–1179.

[111] Nuno L. Gomes et al. “A comparison of fundus autofluorescence and retinal structure in patients with stargardt disease”. In: *Investigative Ophthalmology and Visual Science* 50.8 (2009), pp. 3953–3959.

[112] Hongxin Song et al. “Cone and rod loss in stargardt disease revealed by adaptive optics scanning light ophthalmoscopy”. eng. In: *JAMA Ophthalmology* 133.10 (Oct. 2015), pp. 1198–1203.

[113] Hongxin Song et al. “High-Resolution Adaptive Optics in Vivo Autofluorescence Imaging in Stargardt Disease”. In: *JAMA Ophthalmology* 137.6 (Mar. 2019), pp. 603–609.

[114] Etienne M. Schönbach et al. “Macular sensitivity measured with microperimetry in stargardt disease in the progression of atrophy secondary to stargardt disease (ProgStar) study report No. 7”. In: *JAMA Ophthalmology* 135.7 (2017), pp. 696–703.

[115] Dmitry Romashchenko, Robert Rosén, and Linda Lundström. “Peripheral refraction and higher order aberrations”. In: *Clinical and Experimental Optometry* January (2019), pp. 86–94.

[116] Jie Shen and Larry N. Thibos. “Peripheral aberrations and image quality for contact lens correction”. In: *Optometry and Vision Science* 88.10 (2011), pp. 1196–1205.

[117] C Blehm, S Vishnu, and A Khattak. “Computer Vision Syndrome: A Review, Survey of Ophthalmology”. In: 50.3 (2005), pp. 253–262.

[118] R. Tutt et al. “Optical and visual impact of tear break-up in human eyes”. In: *Investigative Ophthalmology and Visual Science* 41.13 (2000), pp. 4117–4123.

[119] Antonio Benito et al. “Objective optical assessment of tear-film quality dynamics in normal and mildly symptomatic dry eyes”. In: *Journal of Cataract and Refractive Surgery* 37.8 (2011), pp. 1481–1487.

[120] Manoj V. Subbaram and Mark A. Bullimore. “Visual acuity and the accuracy of the accommodative response”. In: *Ophthalmic and Physiological Optics* 22.4 (2002), pp. 312–318.

[121] Ramkumar Sabesan, Heidi Hofer, and Austin Roorda. “Characterizing the human cone photoreceptor mosaic via dynamic photopigment densitometry”. In: *PLoS ONE* 10.12 (2015), pp. 1–12.

[122] Alexander E. Salmon et al. “An Automated Reference Frame Selection (ARFS) Algorithm for Cone Imaging with Adaptive Optics Scanning Light Ophthalmoscopy”. In: *Translational Vision Science & Technology* 6.2 (2017), p. 9.

[123] Gang Huang et al. “Lucky averaging: quality improvement of adaptive optics scanning laser ophthalmoscope images”. In: *Optics Letters* 36.19 (2011), p. 3786.

[124] Michael C. Roggemann. “Image-spectrum signal-to-noise-ratio improvements by statistical frame selection for adaptive-optics imaging through atmospheric turbulence”. In: *Optical Engineering* 33.10 (1994), p. 3254.

[125] Gomathy Ramaswamy and Nicholas Devaney. “Pre-processing, registration and selection of adaptive optics corrected retinal images”. In: *Ophthalmic and Physiological Optics* 33.4 (2013), pp. 527–539.

[126] Gomathy Ramaswamy, Marco Lombardo, and Nicholas Devaney. “Registration of adaptive optics corrected retinal nerve fiber layer (RNFL) images”. In: *Biomedical Optics Express* 5.6 (2014), p. 1941.

[127] Mircea Mujat et al. “High resolution multimodal clinical ophthalmic imaging system”. In: *Optics Express* 18.11 (2010), p. 11607.

[128] Radim Kolar et al. “Registration of retinal sequences from new video-ophthalmoscopic camera”. In: *BioMedical Engineering Online* 15.1 (2016), pp. 1–17.

[129] Richard A. Muller and Andrew Buffington. “Real-Time Correction of Atmospherically Degraded Telescope Images Through Image Sharpening.” In: *J Opt Soc Am* 64.9 (1974), pp. 1200–1210.

[130] J. R. Fienup and J. J. Miller. “Aberration correction by maximizing generalized sharpness metrics”. In: *Journal of the Optical Society of America A* 20.4 (2003), p. 609.

[131] C. E. Shannon. “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.3 (1948), pp. 379–423.

[132] Laura Young and Hannah E. Smithson. “A Self-Organising Photoreceptor Mosaic Simulation for Studying Spatial Sampling in the Retina”. In: *Perception* 47.5 (2018), pp. 559–560.

[133] Lucie Sawides, Alberto de Castro, and Stephen A. Burns. “The organization of the cone photoreceptor mosaic measured in the living human retina”. In: *Vision Research* 132 (2017), pp. 34–44.

[134] Richard Legras, Alain Gaudric, and Kelly Woog. “Distribution of cone density, spacing and arrangement in adult healthy retinas with adaptive optics flood illumination”. In: *PLoS ONE* 13.1 (2018), pp. 1–13.

[135] Jessica Jarosz et al. “High temporal resolution aberrometry in a 50-eye population and implications for adaptive optics error budget”. In: *Biomedical Optics Express* 8.4 (2017), p. 2088.

[136] John I. Yellott. “Spectral analysis of spatial sampling by photoreceptors: Topological disorder prevents aliasing”. In: *Vision Research* 22.9 (1982), pp. 1205–1210.

[137] Nancy J. Coletta and David R. Williams. “Psychophysical estimate of extrafoveal cone spacing”. In: *Journal of the Optical Society of America A* 4.8 (1987), p. 1503.

[138] Nobuyuki Otsu. “Threshold Selection Method From Gray-Level Histograms.” In: *IEEE Trans Syst Man Cybern* SMC-9.1 (1979), pp. 62–66.

[139] Anna-Katharina Hauperich, Laura K Young, and Hannah E Smithson. “What makes a microsaccade? A review of 70 years research prompts a new detection method”. In: *Journal of Eye Movement Research* 12.6 (2020), pp. 1–22.

[140] Robert F. Cooper et al. “Effects of intraframe distortion on measures of cone mosaic geometry from adaptive optics scanning light ophthalmoscopy”. In: *Translational Vision Science and Technology* 5.1 (2016).

[141] Eero P Simoncelli and William T Freeman. “The steerable pyramid: a flexible architecture for multi-scale derivative computation”. In: *IEEE International Conference on Image Processing* 3 (1995), pp. 444–447.

[142] David Cunefare et al. “Automatic detection of cone photoreceptors in split detector adaptive optics scanning light ophthalmoscope images”. In: *Biomedical Optics Express* 7.5 (2016), p. 2036.

[143] Christos Bergeles et al. “Unsupervised identification of cone photoreceptors in non-confocal adaptive optics scanning light ophthalmoscope images”. In: *Biomedical Optics Express* 8.6 (2017), p. 3081.

[144] Carson Sestili. *Deep Learning: Going Deeper toward Meaningful Patterns in Complex Data*. 2018. URL: `https://insights.sei.cmu.edu/sei%7B%5C_%7Dblog/2018/02/deep-learning-going-deeper-toward-meaningful-patterns-in-complex-data.html` (visited on 03/04/2020).

[145] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-Decem (2016), pp. 770–778. eprint: `arXiv:1512.03385v1`.

[146] Christian Szegedy et al. “Going deeper with convolutions”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 07-12-June (2015), pp. 1–9. eprint: `arXiv:1409.4842v1`.

[147] Kaiming He et al. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”. In: *Proceedings of the IEEE International Conference on Computer Vision* 2015 Inter (2015), pp. 1026–1034. eprint: `1502.01852`.

[148] Christian Szegedy et al. “Rethinking the Inception Architecture for Computer Vision”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-Decem (2016), pp. 2818–2826. eprint: `1512.00567`.

[149] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (Sept. 2015), pp. 1–14. eprint: 1409.1556.

[150] Kevin Jarrett et al. "What is the best multi-stage architecture for object recognition?" In: *Proceedings of the IEEE International Conference on Computer Vision* (2009), pp. 2146–2153.

[151] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6 (May 2017), pp. 84–90.

[152] Weidi Xie, J. Alison Noble, and Andrew Zisserman. "Microscopy cell counting and detection with fully convolutional regression networks". In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization* 6.3 (2018), pp. 283–292.

[153] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351 (2015), pp. 234–241. eprint: arXiv:1505.04597v1.

[154] François; et al. Chollet. *Keras*. 2015. URL: https://keras.io.

[155] Martín Abadi et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems". In: (2016). eprint: 1603.04467.

[156] Jianfei Liu et al. "Cone photoreceptor cell segmentation and diameter measurement on adaptive optics images using circularly constrained active contour model". In: *Investigative Ophthalmology and Visual Science* 59.11 (2018), pp. 4639–4652.

[157] Diederik P. Kingma and Jimmy Lei Ba. "Adam: A method for stochastic optimization". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015), pp. 1–15. eprint: 1412.6980.

[158] Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *32nd International Conference on Machine Learning, ICML 2015* 1 (2015), pp. 448–456. eprint: 1502.03167.

[159] Shibani Santurkar et al. "How does batch normalization help optimization?" In: *Advances in Neural Information Processing Systems* 2018-Decem.NeurIPS (2018), pp. 2483–2493. eprint: 1805.11604.

[160] Andrew M. Saxe, James L. McClelland, and Surya Ganguli. "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks". In: *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings* (2014), pp. 1–22. eprint: 1312.6120.

[161] Wei Hu, Lechao Xiao, and Jeffrey Pennington. "Provable Benefit of Orthogonal Initialization in Optimizing Deep Linear Networks". In: January (Jan. 2020), pp. 99–118. eprint: 2001.05992.

[162] N. E. Knox Cartwright et al. "The Cataract National Dataset electronic multicentre audit of 55 567 operations: When should IOLMaster biometric measurements be rechecked?" In: *Eye* 24.5 (2010), pp. 894–900.

[163] Hui Chen et al. "Distribution of axial length, anterior chamber depth, and corneal curvature in an aged population in South China". In: *BMC Ophthalmology* 16.1 (2016), pp. 1–7.

[164] Peter Christian Hoffmann and Werner W. Hütz. "Analysis of biometry and prevalence data for corneal astigmatism in 23 239 eyes". In: *Journal of Cataract and Refractive Surgery* 36.9 (2010), pp. 1479–1485.

[165] Stephen A. Burns et al. "Large-field-of-view, modular, stabilized, adaptive-optics-based scanning laser ophthalmoscope". In: *Journal of the Optical Society of America A* 24.5 (2007), p. 1313.

[166] Alfredo Dubra and Yusufu Sulai. "Reflective afocal broadband adaptive optics scanning ophthalmoscope". In: *Biomedical Optics Express* 2.6 (2011), p. 1757.

[167] David Cunefare et al. "Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks". In: *Scientific Reports* 7.1 (2017), pp. 1–11.

[168] Stephanie J. Chiu et al. "Automatic segmentation of closed-contour features in ophthalmic images using graph theory and dynamic programming". In: *Biomedical Optics Express* 3.5 (2012), p. 1127.

[169] Benjamin S. Liu et al. "The reliability of parafoveal cone density measurements". In: *British Journal of Ophthalmology* 98.8 (2014), pp. 1126–1131.

[170] Christopher S. Langlo et al. "Residual foveal cone structure in CNGB3-associated achromatopsia". In: *Investigative Ophthalmology and Visual Science* 57.10 (2016), pp. 3984–3995.

[171] Lynn W. Sun et al. "Multimodal imaging of photoreceptor structure in Choroideremia". In: *PLoS ONE* 11.12 (2016), pp. 1–21.

[172] David Cunefare et al. "Deep learning based detection of cone photoreceptors with multimodal adaptive optics scanning light ophthalmoscope images of achromatopsia". In: *Biomedical Optics Express* 9.8 (2018), p. 3740.

[173] David Cunefare et al. "RAC-CNN: multimodal deep learning based automatic detection and classification of rod and cone photoreceptors in adaptive optics scanning light ophthalmoscope images". In: *Biomedical Optics Express* 10.8 (2019), p. 3815.

[174] Benjamin Davidson et al. "Automatic Cone Photoreceptor Localisation in Healthy and Stargardt Afflicted Retinas Using Deep Learning". In: *Scientific Reports* 8.1 (2018), pp. 1–13.

[175] Yingming Chen et al. "Cone photoreceptor abnormalities correlate with vision loss in patients with Stargardt Disease". eng. In: *Investigative Ophthalmology and Visual Science* 52.6 (May 2011), pp. 3281–3292.

[176] Moataz M. Razeen et al. "Correlating Photoreceptor Mosaic Structure to Clinical Findings in Stargardt Disease". In: *Translational Vision Science & Technology* 5.2 (2016), p. 6.

[177] Michel Michaelides et al. “High-Resolution Imaging in Stargardt Disease: Preliminary Observations In Preparation for Intervention”. In: *Investigative Ophthalmology & Visual Science* 55.13 (Apr. 2014), p. 5016.

[178] Fouad R Zakka et al. “Disambiguation of photoreceptor structure in transition zones of retinal degenerative diseases”. In: *Investigative Ophthalmology & Visual Science* 55.13 (Apr. 2014), p. 1591.

[179] Moataz M Razeen et al. “Correlating Photoreceptor Abnormalities on Adaptive Optics Scanning Light Ophthalmoscopy to Conventional Clinical Findings in Patients with Stargardt Disease”. In: *Investigative Ophthalmology & Visual Science* 56.7 (June 2015), p. 2781.

[180] Grace Han et al. “Adaptive optics imaging of ABCA4 Retinal Degeneration”. In: *ARVO Abstract 2014*. 2014 Annual Meeting of the Association for Research in Vision and Ophthalmology, ARVO 2014. United States. 55.13 (2014), p. 2616.

[181] I S Audo et al. “Early findings in a phase I/IIa clinical program for stargardt disease (STGD1, MIM #248200)”. In: *Investigative ophthalmology and visual science. Conference: 2015 annual meeting of the association for research in vision and ophthalmology, ARVO 2015. United states*. 2015 Annual Meeting of the Association for Research in Vision and Ophthalmology, ARVO 2015. United States. 56.7 (2015), p. 3819.

[182] Melissa Kasilian et al. “Reliability of cone density measurements on adaptive optics images in Stargardt disease”. In: *Investigative Ophthalmology & Visual Science* 56.7 (June 2015), p. 4928.

[183] Maxwell A Reback et al. “Longitudinal adaptive optics imaging reveals regional variation in cone and rod loss in Stargardt disease”. In: *Investigative Ophthalmology & Visual Science* 56.7 (June 2015), p. 4929.

[184] Tara L Favazza et al. “AO-SLO imaging of diseased retina using offset and confocal apertures”. In: *Investigative Ophthalmology & Visual Science* 57.12 (Sept. 2016), p. 1662.

[185] LUCIA L E E FERRARO et al. “Validation of an adaptive optics scanning laser ophthalmoscope prototype in patients with degenerative retinal diseases.The LITE study.” In: *Investigative Ophthalmology & Visual Science* 58.8 (June 2017), p. 309.

[186] Hongxin Song et al. “Autofluorescence of the photoreceptors in Stargardt disease (SD) identified using fluorescence adaptive optics scanning light ophthalmoscopy (FAOSLO)”. In: *Investigative Ophthalmology & Visual Science* 59.9 (July 2018), p. 4635.

[187] Hongxin Song et al. “Adaptive Optics Scanning Laser Ophthalmoscopy in Stargardt Disease Reveals Decreased Cone and Rod Densities”. In: *Investigative Ophthalmology & Visual Science* 54.15 (June 2013), p. 1743.

[188] Joy Hirsch and Christine A. Curcio. “The spatial resolution capacity of human foveal retina”. In: *Vision Research* 29.9 (1989), pp. 1095–1101.

[189] Edward A.H. Mallen and Priti Kashyap. “Technical Note: Measurement of retinal contour and supine axial length using the Zeiss IOLMaster”. In: *Ophthalmic and Physiological Optics* 27.4 (2007), pp. 404–411.

[190] Kalev Nõupuu et al. “Structural and genetic assessment of the ABCA4-associated optical gap phenotype”. In: *Investigative ophthalmology & visual science* 55.11 (2014), pp. 7217–7226.

[191] Muneo Yamaguchi et al. “High-Resolution Imaging by Adaptive Optics Scanning Laser Ophthalmoscopy Reveals Two Morphologically Distinct Types of Retinal Hard Exudates”. In: *Scientific Reports* 6.April (2016), pp. 1–14.

[192] Sonja G. Karst et al. “Characterization of in Vivo Retinal Lesions of Diabetic Retinopathy Using Adaptive Optics Scanning Laser Ophthalmoscopy”. In: *International Journal of Endocrinology* 2018 (2018).

[193] Ryan N. Vogel et al. “High-resolution imaging of intraretinal structures in active and resolved central serous chorioretinopathy”. In: *Investigative Ophthalmology and Visual Science* 58.1 (2017), pp. 42–49.

[194] Shagun K. Bhatia et al. “Analysis of RPE morphometry in human eyes”. In: *Molecular Vision* 22.March (2016), pp. 898–916.

[195] C. K. Dorey et al. “Cell loss in the aging retina. Relationship to lipofuscin accumulation and macular degeneration”. In: *Investigative Ophthalmology and Visual Science* 30.8 (1989), pp. 1691–1699.

[196] Mark O.M. Ts’o and Ephraim Friedman. “The Retinal Pigment Epithelium: III. Growth and Development”. In: *Archives of Ophthalmology* 80.2 (Aug. 1968), pp. 214–216.

[197] Fritz Krombach et al. “Cell Size of Alveolar Macrophages: An Interspecies Comparison”. In: *Environmental Health Perspectives* 105.September (Sept. 1997), p. 1261.

[198] Austin Roorda, Yuhua Zhang, and Jacque L. Duncan. “High-resolution in vivo imaging of the RPE mosaic in eyes with retinal disease”. In: *Investigative Ophthalmology and Visual Science* 48.5 (2007), pp. 2297–2303.

[199] Jessica I.W. Morgan et al. “In vivo autofluorescence imaging of the human and macaque retinal pigment epithelial cell mosaic”. In: *Investigative Ophthalmology and Visual Science* 50.3 (2009), pp. 1350–1359.

[200] Charles E. Granger et al. “Human retinal pigment epithelium: In vivo cell morphometry, multispectral autofluorescence, and relationship to cone mosaic”. In: *Investigative Ophthalmology and Visual Science* 59.15 (2018), pp. 5705–5716.

[201] Tao Liu et al. “Noninvasive near infrared autofluorescence imaging of retinal pigment epithelial cells in the human retina using adaptive optics”. In: *Biomedical Optics Express* 8.10 (2017), p. 4348.

[202] Zhuolin Liu, Omer P. Kocaoglu, and Donald T. Miller. “3D imaging of retinal pigment epithelial cells in the living human retina”. In: *Investigative Ophthalmology and Visual Science* 57.9 (2016), pp. 533–543.

[203] Zhuolin Liu et al. “In vivo measurement of organelle motility in human retinal pigment epithelial cells”. In: *Biomedical Optics Express* 10.8 (2019), p. 4142.

[204] M Boulton and P Dayhaw-Barker. “The role of the retinal pigment epithelium: topographical variation and ageing changes.” In: *Eye (London, England)* 15.Pt 3 (June 2001), pp. 384–9.

[205] Thomas Ach et al. “Lipofuscin redistribution and loss accompanied by cytoskeletal stress in retinal pigment epithelium of eyes with age-related macular degeneration”. In: *Investigative Ophthalmology and Visual Science* 56.5 (2015), pp. 3242–3252.

[206] S A Burns, J C He, and F Delori. “Do the cones see light scattered from the deep retinal layers”. In: *Vision Science and Its Applications* 1 (1997), pp. 94–97.

[207] Kady S. Bruce et al. “Normal perceptual sensitivity arising from weakly reflective cone photoreceptors”. In: *Investigative Ophthalmology and Visual Science* 56.8 (2015), pp. 4431–4438.

[208] Michalis Georgiou et al. “Adaptive optics imaging of inherited retinal diseases”. eng. In: *British Journal of Ophthalmology* 102.8 (Aug. 2018), pp. 1028–1035.

[209] Ann H. Milam, Zong Yi Li, and Robert N. Fariss. “Histopathology of the human retina in retinitis pigmentosa”. In: *Progress in Retinal and Eye Research* 17.2 (1998), pp. 175–205.

[210] Nasreen Syed et al. “Evaluation of retinal photoreceptors and pigment epithelium in a female carrier of choroideremia”. In: *Ophthalmology* 108.4 (2001), pp. 711–720.