·UNIVERSITY OF OXFORD·

·DEPARTMENT OF ENGINEERING SCIENCE·

Information optimised sensor-less adaptive optics for multiphoton fluorescence microscopy

· A thesis for the degree of Doctor of Philosophy ·

Biwei Zhang

Mansfield College

Supervised by Professor Martin Booth

January 2026

Table of contents

Abstract

Multiphoton fluorescence microscopy is an important technique to image biological structures especially through deep tissue. The microscope images can be degraded by aberrations, so adaptive optics (AO) is normally required to correct aberrations using adaptive elements introduced into the microscope. In many cases, AO is operated without a wavefront sensor to directly measure the aberration but with images taken sequentially for different pre-defined modulations called biases to indirectly estimate the aberration. The sensor-less AO allows a simple optical design thus favoured in a wide range of applications, but it tends to be time-consuming due to the sequential measurement of multiple images and the need for iterations to reach satisfactory estimation. For more time-efficient performance, sensor-less AO methods have been developed to estimate the aberration in more suitable representations and by more powerful algorithms. The state-of-the-art performance has been achieved with a deep learning (DL) based algorithm, which can estimate the aberration from at least two images. However, these methods generally operate image measurement in a heuristic way hence the images are not necessarily in the best condition for aberration estimation. This has prevented sensor-less AO achieving its optimal performance.

In this thesis, sensor-less AO was optimised by considering the information contained in the images. First, a framework was introduced to analyse the information in the images by Fisher information and derive a loss function based on Cramér-Rao

bound to optimise the biases; this framework is not specific to any sensor-less AO method or imaging scenario; the results showed that the optimised biases brought better sensor-less AO performance and the loss function derived from information analysis served a good performance estimator. Then, a new sensor-less AO method was proposed to parallelise the sequential image measurement by using an array of focal spots for imaging with each spot containing a different bias optimised by the framework introduced above, so that the information used to be separated in multiple images was concentrated into one single superimposed image; with a DL based algorithm to estimate the aberration from the superimposed image, the method could finish in milliseconds with even a single image measurement; experiments on a two-photon (2-P) microscope showed that the method generally maintained effective performance in a wide range of scenarios. This optimised sensor-less AO has great potential to improve the imaging performance of multiphoton fluorescence microscopy, especially in challenging situations with fast-changing aberrations. The benefits may also be extended to other imaging modalities.

Chapter 1

Introduction

In biomedical research, a lot of findings rely on the support of optical microscopy. Compared to other imaging modalities, optical microscopy has many advantages: (1) It can image samples in a non-invasive way thus suitable for living sample or in vivo; (2) It reveals structures and functions on a sub-cellular scale; (3) It is fast and induces little damage to the sample [1-3]. Fluorescence microscopy has become one of the most important variants of optical microscopy for its ability to selectively image the structures of interest through the use of targeted fluorescent labels [4-6]. For deep tissue imaging, multiphoton fluorescence microscopy is commonly applied due to its ability to image deep through scattering media [7,8].

Generally, the imaging results of multiphoton fluorescence microscopy are degraded by aberrations. Aberrations can be modelled as variations in the light wavefront. They come from imperfections in the optics of the microscope, and more importantly from the inhomogeneity in the refractive index of the sample. While diffraction limited images can be expected when no aberration exists in the microscope, aberrations will distort the point spread function (PSF) of the microscope and blur the images [9-13].

To restore the image quality, adaptive elements can be applied in the microscope to correct aberrations. The adaptive elements are normally reconfigurable devices such as

a deformable mirror (DM) or a spatial light modulator (SLM). To realise aberration correction, the aberration is measured and the adaptive elements are configured according to the measurement to provide a wavefront modulation that compensates for the aberration. Such a technique is widely known as adaptive optics (AO) [9-13].

There are different methods to measure aberrations in AO. These methods can be divided into two categories based on whether a separate wavefront sensor is required [9-13]. Wavefront sensor-less methods are often favoured as they allow simpler optical design and are applicable to a much wider range of scenarios. [9-13] In sensor-less AO methods, the aberration is estimated iteratively from a sequence of images, each acquired with a different wavefront modulation called a bias aberration (or “bias” for short) [12,14]. The sequential image acquisition and the iterations cause sensor-less AO methods to be time-consuming, which becomes the main problem limiting their application [9-13].

To solve this problem, many works have increased the efficiency of aberration estimation by using more suitable aberration representations and more powerful estimation algorithms [13,14]. Recently, with the use of deep learning (DL) based algorithms, the required number of images and iterations for aberration estimation has been greatly reduced, whereas there is still much room for improvement to really conquer the problem [13]. At the same time, most sensor-less AO methods, including the state-of-the-art ones, operate image measurement based on empirical experience, which means that there is a high chance the measured images are not optimally

conditioned for aberration estimation. Thus, this thesis has explored different ways to optimise image measurement so that faster sensor-less AO methods can be realised to better serve multiphoton fluorescence microscopy and potentially other imaging modalities as well.

1.1 Research contributions

The research in this thesis improved the time efficiency and correction ability of sensor-less AO for multiphoton fluorescence microscopy by optimising the information collected through image measurement for better aberration estimation.

To start with, a review of existing sensor-less AO methods for multiphoton fluorescence microscopy was provided to summarise common principles and discuss varied components leading to different performances. Based on this, two major directions for improvement were proposed as (1) to optimise the biases to maximise aberration information encoded in the images; (2) to parallelise the sequential measurement to concentrate information separated in multiple images into one image.

Following the first direction, a framework was introduced to provide information guided optimisation of the biases. The framework can accommodate different sensor-less AO methods and imaging scenarios. The optimised biases were demonstrated to effectively improve sensor-less AO performance.

Following the second direction, a new sensor-less AO method was presented to measure multiple biased images in parallel so that much faster sensor-less AO was achieved with even a single image measurement. Experimental results showed that the

method was effective in a wide range of scenarios.

1.2 Thesis outline

The thesis is organised as follows:

In Chapter 1, an introduction of this DPhil research with the research contributions listed and a brief outline of the thesis given;

In Chapter 2, the background knowledge involved in the thesis;

In Chapter 3, a review of existing sensor-less AO methods for multiphoton fluorescence microscopy;

In Chapter 4, a framework for information guided optimisation of the biases in sensor-less AO;

In Chapter 5, a new sensor-less AO method for multiphoton fluorescence microscopy based on only a single image;

In Chapter 6, the conclusion and outlook.

Chapter 2

Background

This chapter provides background knowledge that is involved in the thesis. This includes necessary basics of wave optics, a selected introduction to optical microscopy, the main principle of AO, a digest of information theory and some important details of deep learning.

2.1 Wave optics basics

Light is a form of electromagnetic radiation [15]. It exhibits both wave-like and particle-like properties. The study of light and its behaviour is called optics [16]. As it is difficult to apply complete electromagnetic description of light in practice, different simplified models are used in optics for different circumstances. Geometrical optics treats light as rays that travel in straight lines, which only works when the wavelength of light is small compared to the size of objects that light interacts with [15,16]. Wave optics studies light as waves and explains phenomena such as interference and diffraction that geometric optics fails to account for [15,16]. Quantum optics considers light not only as waves but also a stream of particles called photons, each of which carries one quantum of energy that light can only gain or lose as a whole [17]. Quantum optics is required to model effects related to quantum mechanics. Here, we focus on wave optics and provide the basics used for modelling in this thesis.

2.1.1 Wave model of light

In the 19th century, Maxwell first concluded theoretically that light is electromagnetic waves, which was then confirmed by Hertz through experiments [15]. Electromagnetic waves can be described by Maxwell's equations:

$$\nabla \cdot \boldsymbol{D} = \rho \tag{2-1a}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2-1b}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{2-1c}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t} \tag{2-1d}$$

with $\boldsymbol{D}$ the displacement field, $\boldsymbol{B}$ the magnetic field, $\boldsymbol{E}$ the electric field, $\boldsymbol{H}$ the magnetizing field, ρ the electric charge density, $\boldsymbol{J}$ the conduction current density and t the time. In a static isotropic medium, there are relations:

$$\boldsymbol{J} = \sigma \boldsymbol{E} \tag{2-2a}$$

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \tag{2-2b}$$

$$\boldsymbol{B} = \mu \boldsymbol{H} \tag{2-2c}$$

where σ is the conductivity, ε is the permittivity and μ is the permeability. For simplicity, if we assume an isotropic homogeneous medium ($\sigma = 0$ and ε, μ being constants) with no free charge and conduction current in it ($\rho = 0$, $\boldsymbol{J} = 0$), then it can be derived from the equations (2-1) and (2-2) that:

$$\nabla^2 \boldsymbol{E} - \frac{1}{v^2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0 \tag{2-3a}$$

$$\nabla^2 \boldsymbol{B} - \frac{1}{v^2}\frac{\partial^2 \boldsymbol{B}}{\partial t^2} = 0 \tag{2-3b}$$

The equations (2-3) indicate that the electromagnetic field propagates through space in

the form of waves at a speed $v = 1/\sqrt{\varepsilon\mu}$. In a vacuum, electromagnetic waves travel at a speed $c = 1/\sqrt{\varepsilon_0\mu_0}$, where ε_0 is the vacuum permittivity and μ_0 is the vacuum permeability. The ratio of c and v defines the refractive index of a medium as:

$$n = c/v \tag{2-4}$$

While different forms of solutions exist for the equations (2-3), the planar harmonic solutions are normally taken to study the waves as they provide the fundamental components of other solutions. The planar harmonic solutions can be written in a complex form as:

$$\boldsymbol{E} = \boldsymbol{A}e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} \tag{2-5a}$$

$$\boldsymbol{B} = \boldsymbol{A}'e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} \tag{2-5b}$$

Here, $\boldsymbol{A}$ and $\boldsymbol{A}'$ are amplitude vectors of the electric field and the magnetic field respectively; $\boldsymbol{k}$ denotes the wave vector and is defined as $\boldsymbol{k} = k\boldsymbol{k_0}$, where $\boldsymbol{k_0}$ is a unit vector along the direction of wave propagation and k is a magnitude representing the angular wavenumber; $\boldsymbol{r}$ is the position vector; ω is the angular frequency. The wavelength λ is related to k by:

$$k = 2\pi/\lambda \tag{2-6}$$

By applying the planar harmonic solutions into Maxwell's equations, we can get:

$$\boldsymbol{k}\cdot\boldsymbol{E} = 0 \tag{2-7a}$$

$$\boldsymbol{k}\cdot\boldsymbol{B} = 0 \tag{2-7b}$$

$$\boldsymbol{B} = \frac{1}{v}(\boldsymbol{k_0}\times\boldsymbol{E}) = \sqrt{\varepsilon\mu}(\boldsymbol{k_0}\times\boldsymbol{E}) \tag{2-7c}$$

The equations (2-7) suggest that $\boldsymbol{E}$, $\boldsymbol{B}$ and $\boldsymbol{k}$ are all orthogonal to each other, and that

given $\boldsymbol{E}$ and $\boldsymbol{k}$, $\boldsymbol{B}$ is also determined. Thus, the electromagnetic field can be represented by the electric field $\boldsymbol{E}$ alone.

When we only consider the spatial distribution of the waves at a certain moment, the temporal term in the equation (2-5a) can be ignored, so $\boldsymbol{E}$ can be represented in the form of the complex amplitude:

$$\boldsymbol{E} = \boldsymbol{A}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = \boldsymbol{A}e^{i\phi} \tag{2-8}$$

In this case, the phase ϕ is defined by $\boldsymbol{k}\cdot\boldsymbol{r}$. The set of all the spatial points with the same phase consist of the wavefront, whose normal direction is the direction of $\boldsymbol{k}$. It is easy to see that $\boldsymbol{E}$ oscillates as the waves travel forward. The orientation of the oscillation defines the polarisation. It means linear polarisation if the orientation does not change.

The electromagnetic waves carry radiant energy through space. The energy flux can be represented by Poynting vector:

$$\boldsymbol{S} = \frac{1}{\mu}\boldsymbol{E}\times\boldsymbol{B} \tag{2-9}$$

Here, the direction of $\boldsymbol{S}$ is the same as the propagation direction for the planar waves. The temporal average of $\boldsymbol{S}$ defines the intensity. Normally, the relative intensity is calculated by:

$$I = |\boldsymbol{E}|^2 \tag{2-10}$$

where for simplicity multiplying constants have been omitted.

By frequency or wavelength, the electromagnetic waves can be classified into separate bands that together form the electromagnetic spectrum shown in Figure 2-1

[18]. In such spectrum, the visible band spans the wavelength range roughly from 380 to 750 nm with different wavelength corresponding to different colour. Specifically, the visible band along with its adjacent bands, the infrared (IR) and the ultraviolet (UV), defines the light in optics, which can be manipulated by optical elements [19].

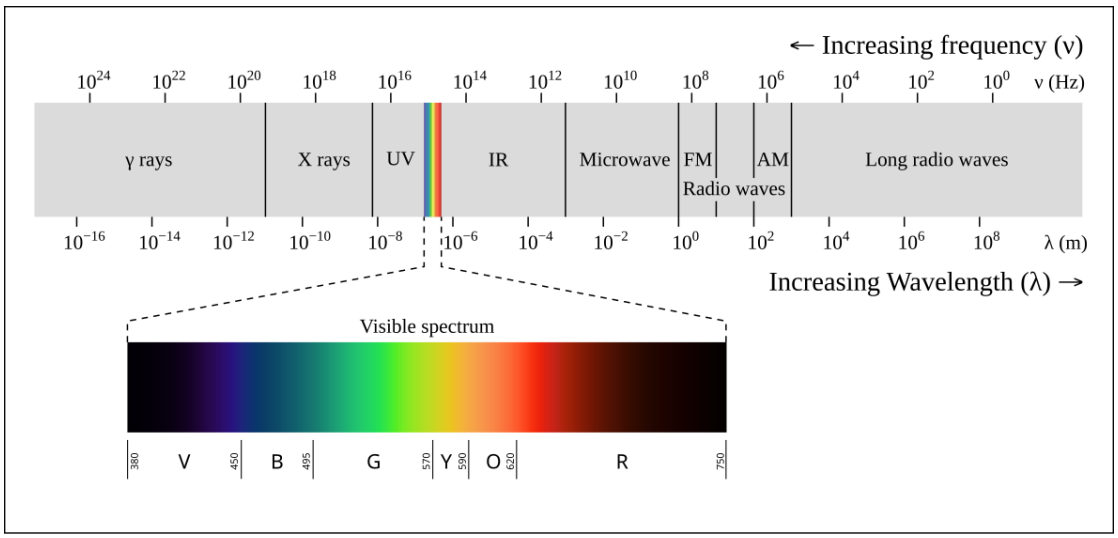


Figure 2-1 [18]: The electromagnetic spectrum, with the visible spectrum highlighted.

2.1.2 Interference and diffraction

When two or more light waves meet at a point in space, the combined light field at that point equals to the vector sum of the light field of each individual wave. For two light waves of the same frequency and polarisation, the combined light field can be written in a scalar form as:

$$E = E_1 + E_2 = A_1 e^{i\phi_1} + A_2 e^{i\phi_2} \tag{2-11}$$

with the first light wave represented by $E_1 = A_1 e^{i\phi_1}$ and the second by $E_2 = A_2 e^{i\phi_2}$. The combined light intensity can be derived as:

$$I = |E|^2 = {A_1}^2 + {A_2}^2 + 2A_1 A_2 \cos(\phi_1 - \phi_2) \tag{2-12}$$

This suggests that the combined light intensity is determined by the phase difference $\phi_1 - \phi_2$ at that point and such phenomenon is called interference. The potential for two light waves to interfere is defined as coherence. When constructive interference occurs, the combined intensity is larger than the sum of the individual intensities; when destructive interference takes place, the combined intensity is smaller than the sum of the individual intensities. The combined intensity of two incoherent light waves however equals to the sum of the individual intensities [15].

Diffraction is the spreading out of light waves as they pass around an obstacle or through a narrow opening. It can be described by Huygens' principle, which states that each point on a propagating wavefront can be treated as a source of spherical wavelets, and the envelop of the secondary wavelets from different points forms a new wavefront [15]. As the secondary wavelets originated from the same wavefront are coherent with each other, a mathematical expression of Huygens' principle can be provided by Fresnel-Kirchhoff diffraction formula based on mutual interference among wavelets. If we assume a light wave of wavelength λ propagating along the vector $\boldsymbol{z}$ through an aperture on a plane of coordinates (x_1, y_1), the light field on a plane of coordinates (x_2, y_2) at a distance z_1 after the aperture can be expressed by:

$$E(x_2, y_2) = \frac{1}{i\lambda} \iint E(x_1, y_1) \left[\frac{1+\cos(\boldsymbol{r},\boldsymbol{z})}{2}\right] \frac{e^{ikr}}{r} dx_1 dy_1 \qquad (2\text{-}13)$$

Here, $\boldsymbol{r}$ is the vector between each pair of (x_1, y_1) and (x_2, y_2), whose size can be calculated by $r = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + {z_1}^2}$; $\cos(\boldsymbol{r}, \boldsymbol{z})$ is the cosine of the angle between $\boldsymbol{r}$ and $\boldsymbol{z}$.

Different approximations can be applied to further simplify the equation (2-13) upon different scales of z_1. Common approximations include taking $\cos(\boldsymbol{r},\boldsymbol{z})$ as 1 and replacing the r in the denominator with z_1, as they generally cause little difference to the result [15]. On top of these, the Fresnel approximation assumes that z_1 is large enough to let $r \approx z_1 + \frac{(x_2-x_1)^2+(y_2-y_1)^2}{2z_1}$, which simplifies the equation (2-13) into:

$$E(x_2, y_2) = \frac{e^{ikz_1}}{i\lambda z_1} \iint E(x_1, y_1)\, e^{\frac{ik}{2z_1}[(x_2-x_1)^2+(y_2-y_1)^2]} dx_1 dy_1 \qquad (2\text{-}14)$$

The Fraunhofer approximation considers even larger z_1 that satisfies $z_1 \gg ({x_1}^2 + {y_1}^2)_{max}/\lambda$ with $({x_1}^2 + {y_1}^2)_{max}$ representing the largest radius of the aperture and allows further simplifications by:

$$E(x_2, y_2) = \frac{e^{ikz_1}}{i\lambda z_1} e^{\frac{ik}{2z_1}({x_2}^2+{y_2}^2)} \iint E(x_1, y_1)\, e^{-\frac{ik}{z_1}(x_1x_2+y_1y_2)} dx_1 dy_1 \qquad (2\text{-}15)$$

As a parameter to indicate which approximation should be applied, Fresnel number F is defines by:

$$F = \frac{({x_1}^2+{y_1}^2)_{max}}{z_1\lambda} \qquad (2\text{-}16)$$

When $F \approx 1$, the diffracted light is in the near field and the Fresnel approximation is suitable, while $F \ll 1$ means the diffracted light in the far field and the Fraunhofer approximation becomes valid [15].

2.1.3 Fourier optics

A general light wave can be represented by a combination of planar waves, with each of them considered as a natural mode of spatial frequency [20]. This allows us to use Fourier transform to study light of any wave form as a spatial frequency spectrum.

Fourier optics assists analysis of various optical imaging systems [16,20,21].

2.1.3.1 Thin lens model

Lenses are the most fundamental components in optical imaging systems [16]. A common convex lens of a thickness can be approximately modelled as a thin lens which changes the phase of transmission light differently at different positions on a plane, along with a pupil which defines the limited aperture [21]. If the pupil is a circle of radius a on a plane of coordinates (x_1, y_1), the pupil function can be given as:

$$P(x_1, y_1) = \begin{cases} 1, & \sqrt{{x_1}^2 + {y_1}^2} \le a \\ 0, & \sqrt{{x_1}^2 + {y_1}^2} > a \end{cases} \tag{2-17}$$

When the thin lens is located on the same plane, the transmission function can be derived by applying a paraxial approximation to be:

$$t(x_1, y_1) = P(x_1, y_1) e^{-\frac{ik}{2f}({x_1}^2 + {y_1}^2)} \tag{2-18}$$

where f is the focal length of the lens.

If the pupil is illuminated by a light wave of wavelength λ propagating along the vector $\boldsymbol{z}$, according to equations (2-14) and (2-18), the light field on the back focal plane of coordinates (x_2, y_2) at a distance f after the thin lens can be calculated by:

$$E(x_2, y_2) = \frac{e^{ikf}}{i\lambda f} e^{\frac{ik}{2f}({x_2}^2 + {y_2}^2)} \iint E(x_1, y_1) P(x_1, y_1)\, e^{-\frac{ik}{f}(x_1 x_2 + y_1 y_2)} dx_1 dy_1 \tag{2-19}$$

By introducing a pair of spatial frequencies $u_x = \frac{x_2}{\lambda f}$ and $u_y = \frac{y_2}{\lambda f}$, the equation (2-19) can be turned into a Fourier transform with an extra phase factor:

$$E(x_2, y_2) = \frac{e^{ikf}}{i\lambda f} e^{i\pi\lambda f({u_x}^2 + {u_y}^2)} \mathcal{F}[E(x_1, y_1) P(x_1, y_1)]_{u_x, u_y} \tag{2-20}$$

where $\mathcal{F}$ denotes the two-dimensional (2-D) Fourier transform. When the pupil is

moved to the front focal plane at a distance f before the thin lens, a new phase factor $e^{ikf}e^{-i\pi\lambda f(u_x{}^2+u_y{}^2)}$ will have to be introduced into the equation (2-20) so that [21]:

$$E(x_2, y_2) = \frac{e^{i2kf}}{i\lambda f}\mathcal{F}[E(x_1, y_1)P(x_1, y_1)]_{u_x,u_y} \tag{2-21}$$

In this case, by ignoring the constant factor, the light field on the back focal plane is exactly the Fourier transform of that on the pupil.

For any defocused plane at a distance of $f + \Delta z$ after the thin lens with Δz representing the amount of defocus, the light field can be similarly calculated as in the equation (2-21) by [21]:

$$E(x_2, y_2, \Delta z) = \frac{e^{i2k(f+\Delta z)}}{i\lambda(f+\Delta z)}\mathcal{F}\left[E(x_1, y_1)P(x_1, y_1)e^{\frac{ik}{2}\left(\frac{1}{f}-\frac{1}{f+\Delta z}\right)(x_1{}^2+y_1{}^2)}\right]_{u_x,u_y} \tag{2-22}$$

Obviously, the light field on a defocused plane is still equivalent to the Fourier transform of that on the pupil but with an extra phase factor $e^{\frac{ik}{2}\left(\frac{1}{f}-\frac{1}{f+\Delta z}\right)(x_1{}^2+y_1{}^2)}$. The equation (2-22) provides the expression of the three-dimensional (3-D) light field in the back focal region.

The limiting pupil determines the numerical aperture (NA) of the lens, which is defined by:

$$NA = n\sin\alpha \tag{2-23}$$

where n is the refractive index of the medium in which the lens works and α is the half-angle of the maximum cone of light that can enter or exit the lens. Based on this, a pair of normalised optical coordinates can be defined as:

$$u_\rho = \frac{2\pi NA\rho}{\lambda} \tag{2-24a}$$

$$u_z = \frac{8\pi n\Delta z}{\lambda}\sin^2\left(\frac{\alpha}{2}\right) \approx \frac{2\pi\Delta z NA^2}{\lambda n} \tag{2-24b}$$

where $\rho = \sqrt{{x_2}^2 + {y_2}^2}$. With the optical coordinates, the light intensity on the back focal plane and along the optical axis can be derived respectively from the equation (2-22) as:

$$I(u_\rho, u_z = 0) = \left|E(u_\rho, u_z = 0)\right|^2 = \left|\frac{2\mathcal{J}_1(u_\rho)}{u_\rho}\right|^2 \quad \text{(2-25a)}$$

$$I(u_\rho = 0, u_z) = \left|E(u_\rho = 0, u_z)\right|^2 = \left|\frac{\sin(u_z/4)}{u_z/4}\right|^2 \quad \text{(2-25b)}$$

where the constant factors are ignored and $\mathcal{J}_1$ representing the first-order Bessel function. This suggests that the lens focuses the light into a diffraction spot in which most of the light intensity concentrates in the central bright part. Specially, the lateral profile of the central bright part on the back focal plane is called Airy disc. If the resolution of an imaging system is limited by the diffraction of a lens, as an empirical criterion, the lateral and axial resolution can be theoretically estimated by the full width half maximum (FWHM) of the lateral and axial profile of central bright part, which can be roughly given by [22]:

$$\sigma_\rho = \frac{0.51\lambda}{NA} \quad \text{(2-26a)}$$

$$\sigma_z = \frac{1.8n\lambda}{NA^2} \quad \text{(2-26b)}$$

2.1.3.2 4f system

An imaging system consisting of a pair of confocal lenses is called a 4f system [16]. A 4f system is shown in Figure 2-2, where the pupil of the system is assumed on the confocal plane in the middle of the two lenses. A 2-D object placed on the front focal plane of the first lens (L_1, focal length f_1) generates the light field $E(x_0, y_0)$ of wavelength λ. If the light from the object is coherent, the light field on the confocal

plane can be derived according to the equation (2-21) as:

$$E(x_1, y_1) = \mathcal{F}[E(x_0, y_0)]_{u_{x_1}, u_{y_1}} \tag{2-27}$$

where the constant factors are ignored. Apparently, it is the spatial frequency spectrum that shows on the confocal plane with $u_{x_1} = \frac{x_1}{\lambda f_1}$ and $u_{y_1} = \frac{y_1}{\lambda f_1}$. Likewise, the light field of the 2-D image on the back focal plane of the second lens (L_2, focal length f_2) can be computed by:

$$E(x_2, y_2) = \mathcal{F}[E(x_1, y_1)P(x_1, y_1)]_{u_{x_2}, u_{y_2}} \tag{2-28}$$

with $u_{x_2} = \frac{x_2}{\lambda f_2}$ and $u_{y_2} = \frac{y_2}{\lambda f_2}$. From the equations (2-27) and (2-28), we can get:

$$E(x_2, y_2) = \mathcal{F}\left[\mathcal{F}[E(x_0, y_0)]_{u_{x_1}, u_{y_1}}\right]_{u_{x_2}, u_{y_2}} \otimes \mathcal{F}[P(x_1, y_1)]_{u_{x_2}, u_{y_2}} \tag{2-29}$$

with $\otimes$ calculating the 2-D convolution. Here, the term before $\otimes$ can be solved as $E(Mx_0, My_0)$ with $M = -\frac{f_2}{f_1}$.

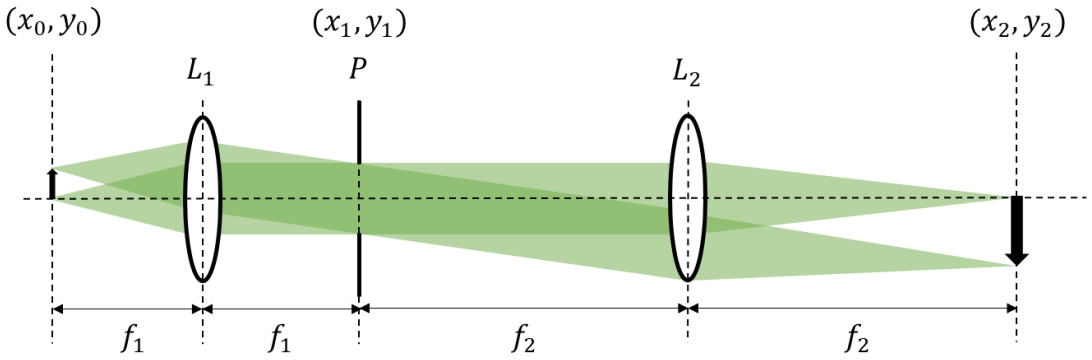


Figure 2-2: A 4f system. The focal lengths of the two lenses L_1 and L_2 are f_1 and f_2 respectively. The pupil P is on the confocal plane. An object is placed on the front focal plane of L_1, and its image is located on the back focal plane of L_2.

If we assume an infinite pupil, the image is a magnified replica of the object with M being the magnification, which suggests that the image plane is a conjugate plane of the object plane. The minus M means an inverted image. On the other hand, if we

reduce the object to be a single point, the image becomes an amplitude point spread function (PSF) representing the light complex amplitude defined by:

$$h(x_2, y_2) = \mathcal{F}[P(x_1, y_1)]_{u_{x_2}, u_{y_2}} \tag{2-30}$$

It can be seen that the PSF comes from the pupil filtering the spatial frequency spectrum.

As a more common case, the light field of the 3-D image volume of a 3-D object can be similarly computed by [21]:

$$E(x_2, y_2, \Delta z) = \left[E(Mx_0, My_0, M^2\Delta z_0)e^{i\zeta_0(\Delta z_0)}\right] \otimes_{3-D} h(x_2, y_2, \Delta z) \tag{2-31}$$

Here, Δz is the amount of defocus from the back focal plane of L_2; Δz_0 is the amount of defocus from the front focal plane of L_1; $E(Mx_0, My_0, M^2\Delta z_0)$ represents a magnified replica of the object with M the lateral magnification and M^2 the axial magnification; $e^{i\zeta_0(\Delta z_0)}$ is a phase factor caused by Δz_0; $\otimes_{3-D}$ calculates the 3-D convolution; $h(x_2, y_2, \Delta z)$ represents a 3-D PSF, which is defined by:

$$h(x_2, y_2, \Delta z) = \mathcal{F}\left[P(x_1, y_1)e^{i\zeta(\Delta z)}\right]_{u_{x_2}, u_{y_2}} \tag{2-32}$$

where $e^{i\zeta(\Delta z)}$ is a phase factor caused by Δz. When the light from the object is incoherent, a similar expression as the equation (2-31) can be used to compute the light intensity of the 3-D image volume by [21]:

$$I(x_2, y_2, \Delta z) = |E(x_2, y_2, \Delta z)|^2 =$$

$$\left|E(Mx_0, My_0, M^2\Delta z_0)e^{i\zeta_0(\Delta z_0)}\right|^2 \otimes_{3-D} |h(x_2, y_2, \Delta z)|^2 \tag{2-33}$$

In this case, $|h(x_2, y_2, \Delta z)|^2$ defines a new intensity PSF, which represents the light intensity rather than the light complex amplitude.

2.1.3.3 General model of optical imaging systems

From the representative study of a 4f system, a general model can be further concluded to describe optical imaging systems. In this model, optical imaging systems take the object as the input and transform it into the image as the output. The transform can be deemed as linear and spatially invariant, which means that the complete image is a linear combination of images of different points on the object, and that different object points form their own images following the same function. Hence, without considering the system magnification, the image can be represented by a convolution between the object and the PSF, with the PSF characterising the transform behaviour of the system [16,20,21].

For coherent imaging such as what we assumed in previous sections, the system linearity is in terms of the light complex amplitude, so the PSF is a function of the light complex amplitude. In this case, the Fourier transform of the PSF is defined as the coherent transfer function (CTF), which describes how the system processes the spatial frequency spectrum when transforming the object into the image. In systems of limited aperture, the CTF is the same as the pupil function.

When it comes to incoherent imaging, it is easy to deduce that the system model still holds whereas the system linearity lies in the light intensity. Accordingly, the PSF becomes a function of the light intensity, and the Fourier transform of the PSF now defines the optical transfer function (OTF), which describes the change in the spatial frequencies of the light intensity distribution through the system. In systems of limited aperture, the OTF is the autocorrelation of the pupil function, which means that the

spatial frequency cut-off of the OTF is twice that of the CTF.

2.2 Optical microscopy

Optical microscopy is an optical imaging technique that produces enlarged images of small objects with a system of lenses. Modern optical microscopes are commonly built in a compound form to achieve higher magnification [3,23]. For specialised purposes, there are many variants of the optical compound microscopy applying different illumination setups to introduce various contrast mechanisms such as phase [24-29], fluorescence [4-8,30], harmonic generation [31-34], etc.

2.2.1 Optical compound microscopy

A diagram is given in Figure 2-3 to show the standard setups of an optical compound microscope. Conventionally, an objective lens alone forms an enlarged inverted real image of the object at an intermediate position within the microscope. The intermediate image is further magnified by an eyepiece to produce an inverted virtual image of the object that viewed by eyes. As a result, the total magnification of the optical compound microscope equals to the magnification of the objective lens times the magnification of the eyepiece. Instead, most modern optical compound microscopes use an infinity-corrected system where the objective lens projects a virtual image at infinity and an extra tube lens after the objective lens refocuses the virtual image into the real image [1]. Technically, the objective lens and the tube lens form a 4f system as shown in Figure 2-2. Such a design allows for the addition of optical accessories between the objective lens and the tube lens with minimal effects in the image quality. Moreover, a digital

camera is often used instead of the eyepiece to record the real image for viewing [1-3,23].

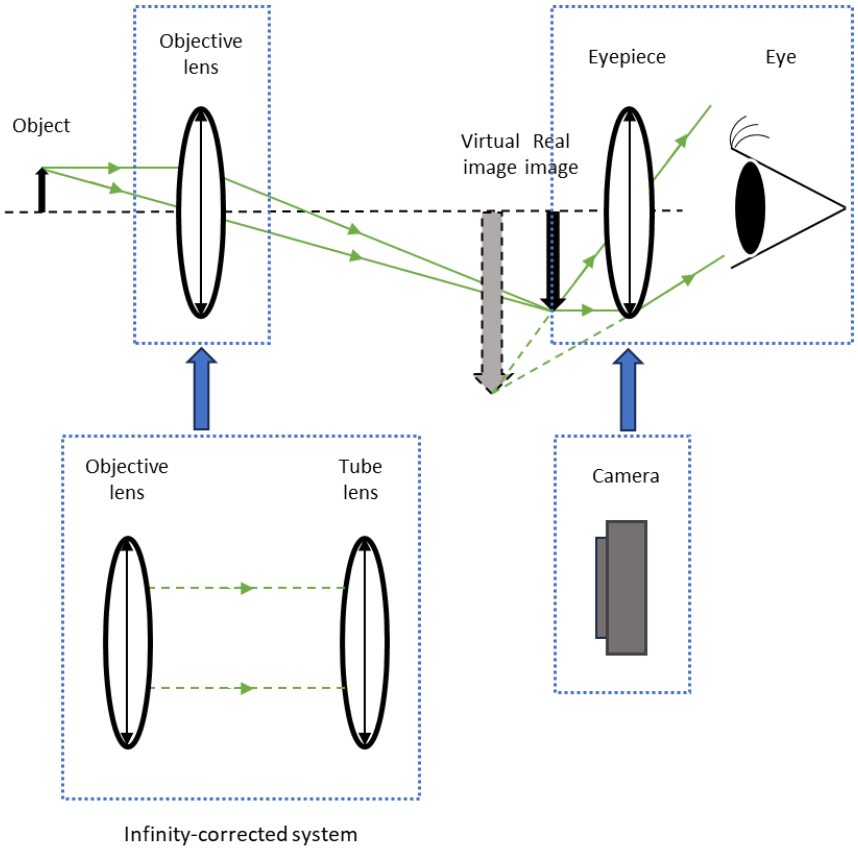


Figure 2-3: A diagram of standard setups of an optical compound microscope. The objective lens forms a magnified inverted real image from the object, which is further magnified by the eyepiece to form a virtual image viewed by the eye. In modern designs, the single objective lens is mostly replaced by an infinity-corrected system consisting of an objective lens and an extra tube lens, and a digital camera is often used instead of the eyepiece to record the real image for viewing.

In a standard optical compound microscope with an infinity-corrected system and a digital camera, the image can be computed either by the equation (2-31) for coherent imaging or the equation (2-33) for incoherent imaging. Since a digital camera consists of a 2-D array of pixels each detecting the light intensity independently with a limited signal-to-noise ratio (SNR) and a potential background [35], the detected image can be

described as:

$$D = R(\xi I_D + \beta) \tag{2-34}$$

where R represents a distribution that characterising the random noise; ξ is the signal amplification; I_D is a 2-D array of the image light intensity signal detected by each pixel; β represents the background. The lateral resolution of the microscope is approximately the estimation calculated by the equation (2-26a) with the NA of the objective lens or the pixel size of the camera whichever is larger, while the axial resolution is generally poor as the camera records a superposition of image components from the object slices at different axial depths.

2.2.2 Fluorescence microscopy

For objects in biological samples, the structures of interest can be selectively labelled with fluorophores, which are molecules that emit fluorescence upon absorption of light [36]. As a variant of optical compound microscopy, fluorescence microscopy collects the fluorescence emitted only by the labelled structures for imaging, so the image background is greatly reduced. The excitation of fluorescence includes a process where the fluorophore absorbs a photon to jump from a ground state to an unstable excited state when illuminated by excitation light of a specific wavelength λ_e, and then emits a photon as fluorescence of a longer wavelength λ_f to return to the ground state. The difference between λ_e and λ_f is called the Stokes shift [5]. Fluorescence microscopy requires intense illumination of excitation light which is commonly provided by such as a xenon arc lamps with an excitation filter, a laser or a high-power light emitting

diode (LED) [37]. For coherent excitation light from a laser, the light coherence is lost through the fluorescence excitation process, so that it is incoherent imaging that performed by fluorescence microscopy. If there are no saturation effects, the light intensity of the emitted fluorescence is in proportion to that of the excitation light.

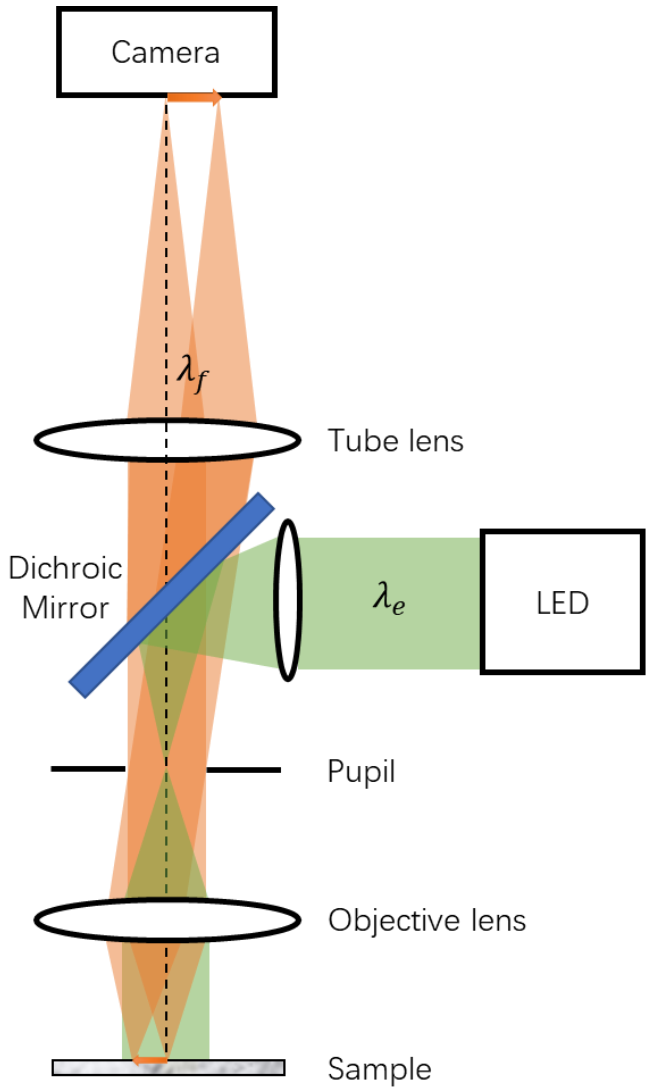


Figure 2-4: A diagram of a widefield epifluorescence microscope. The excitation light of a wavelength λ_e from a LED is reflected by the dichroic mirror and passes through the objective lens of a focal length f_1 to evenly illuminate the sample; the emitted fluorescence of a longer wavelength λ_f is collected by the objective lens, transmitted by the dichroic mirror, and refocused by the tube lens of a focal length f_2 on the camera to be recorded.

Most fluorescence microscopes adopt the epifluorescence design as the one shown in Figure 2-4 where the objective lens is not only used for imaging but also for providing the illumination [5]. This causes that the excitation light and the fluorescence overlap with each other in the light path. To separate them, a dichroic mirror is inserted into the light path and acts as a wavelength specific filter that reflects the excitation light with a shorter wavelength and transmits the fluorescence with a longer wavelength. In the widefield epifluorescence microscope presented in Figure 2-4, the sample is evenly illuminated and the image is detected by a digital camera. For a fluorophore distribution O in the sample as the object, the image fluorescence intensity signal can be described based on the equation (2-33) as:

$$I_D = \left(O I_e \left| e^{i\zeta_0} \right|^2 \right) \otimes_{3-D} |h|^2 \tag{2-35}$$

with the magnification ignored and I_e the constant excitation light intensity. The detected image D can be further acquired by the equation (2-34). The resolution of the microscope shares a similar situation as the standard optical compound microscope in the last section, with $\lambda = \lambda_f$ when calculating the equation (2-26a).

Fluorescence microscopy normally suffers from photodamage problems [38]. One of them is photobleaching, which is due to fluorophores accumulating chemical damage from the fluorescence excitation and gradually losing their ability to fluoresce as they are illuminated [39]. Also, the fluorescence excitation is prone to inducing phototoxicity to living biological samples [40]. These problems severely limit the sample exposure over which images can be acquired by fluorescence microscopy.

2.2.3 Confocal fluorescence microscopy

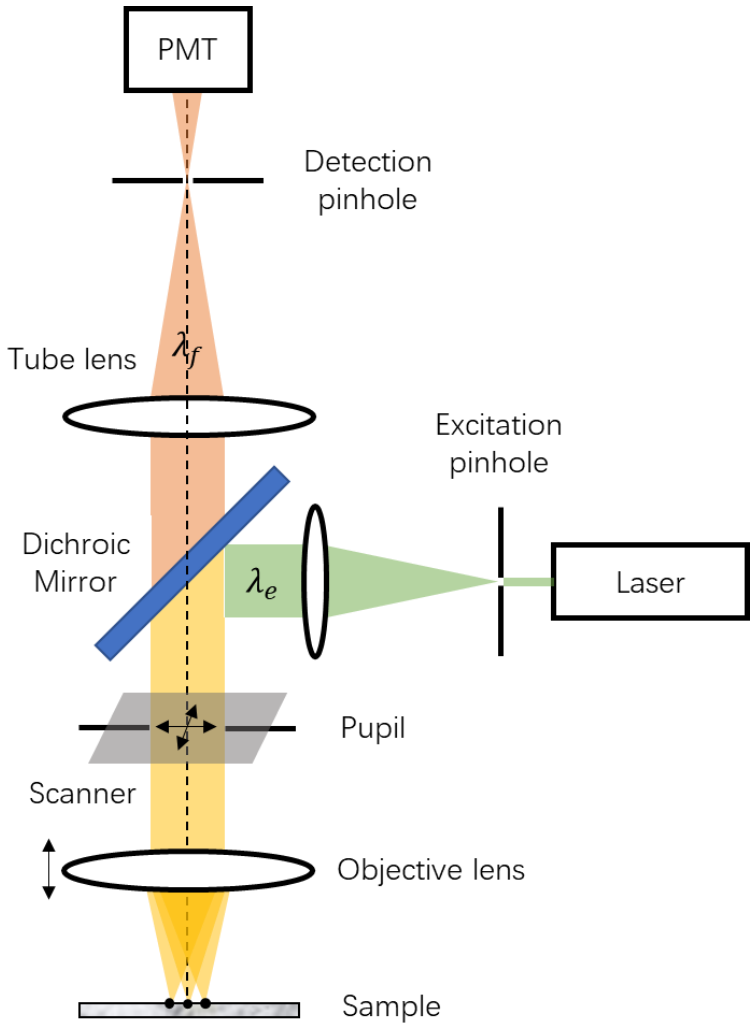


Figure 2-5: A diagram of a confocal fluorescence microscope. The laser beam of a wavelength λ_e passes through the excitation pinhole before reflected by the dichroic mirror and horizontally scanned by a scanner to be focused by the objective lens onto different spots in the sample for illumination; the emitted fluorescence of a longer wavelength λ_f is collected by the objective lens, de-scanned by the scanner, transmitted by the dichroic mirror, refocused by the tube lens and filtered by the detection pinhole until finally reaching the PMT for detection. Signal detected from different scanned positions can collectively build up a 2-D image of an optical section centred with the scanned plane. By moving the objective lens vertically, multiple optical sections can be imaged at different depths, which allows for the 3-D reconstruction of a volume in the sample. For simplicity, the scanner is drawn at the pupil here, while it normally consists of a pair of galvanometer mirrors positioned at conjugate planes of the pupil that relayed by 4-f systems in real practice.

Conventional fluorescence microscopy takes a widefield form, which means that the entire volume of interest in the sample is illuminated. This means that when imaging a plane of interest, the fluorophores in both the focal plane and out-of-focus planes will be excited simultaneously to emit fluorescence. The detected signal will hence contain a large part of background signal from the out-of-focus parts of the sample. This problem can be addressed by the use of confocal fluorescence microscopy [41].

A diagram of a confocal fluorescence microscope is presented in Figure 2-5. A pinhole is placed in front of the detector on a conjugate plane of the front focal plane of the objective lens so that only fluorescence from a small area on a thin section of the sample centred with the focal plane can be detected. This significantly reduces the detected signal. To offset the signal drop, a laser beam is focused by the objective lens into a tight diffraction spot for illumination; the size of the pinhole is controlled approximately the size of the Airy disc of the diffraction spot on the focal plane; a photomultiplier tube (PMT) is used as a more sensitive detector to produce a single image pixel from all the detected signal [42]. As only a single image pixel is produced at a time, a scanner (usually a pair of galvanometer mirrors) is required to scan the illumination spot across the focal plane in a raster pattern, while a 2-D image can be constructed pixel-by-pixel according to each scanning location. The constructed 2-D image represents the entire thin sample section. By further moving the objective lens along the optical axis, a series of 2-D images at different depths are stacked into a 3-D image of the volume of interest.

By assuming constant excitation light field on the pupil, the light intensity of the illumination focal spot can be derived from the equation (2-22) as:

$$I_e = \left|\mathcal{F}\left[Pe^{i\zeta}\right]\right|^2 \tag{2-36}$$

where the constant factors are ignored and $e^{i\zeta}$ represents the extra defocus phase factor. Referring to the equation (2-32), I_e takes the same form as $|h|^2$ with $\lambda = \lambda_e$. The pinhole filtering the fluorescence resulted from the focal spot on a conjugate plane can also be modelled as $|h|^2$ but with $\lambda = \lambda_f$ according to the equation (2-33) [21]. Thus, the detected image signal along with the scanning of the illumination focal spot can be computed as a pixel array by:

$$I_D = O \otimes_{3-D} \left[|h_e|^2 \left|h_f\right|^2\right] \tag{2-37}$$

Here, $|h_e|^2\left|h_f\right|^2$ defines the PSF of the microscope with $|h_e|^2$ from I_e and $\left|h_f\right|^2$ from the pinhole. The pixel dimensions are defined by the scanning dimensions and the pixel size is the scanning step which also decides the magnification of the microscope. As the PMT is also a digital detector, the detected image D can be similarly acquired by the equation (2-34).

Compared to the widefield fluorescence microscope, the confocal fluorescence microscope provides the important optical sectioning capacity which significantly enhances the axial resolution [43]. If the scanning step is sufficiently small, both the lateral and axial resolution can be estimated by the equation (2-26) but with $\lambda = \lambda_e$. As λ_e is shorter than λ_f, there is also a marginal improvement in the lateral resolution. For a large scanning step, the resolution is decided by the scanning step. Also, the

confocal fluorescence microscope further suppresses the image background [43].

2.2.4 Multiphoton fluorescence microscopy

Different from conventional fluorescence microscopy where fluorescence is excited by a single photon, multiphoton fluorescence microscopy performs imaging based on fluorescence from multiphoton excitation, which is a type of non-linear optical effect [44,45]. Multiphoton excitation involves a process that the fluorophore can absorb multiple photons of much lower energy than needed for single-photon excitation simultaneously to get enough energy to jump to the excited state and emit fluorescence. In this case, the emitted fluorescence is of a shorter wavelength λ_f than the wavelength λ_e of the excitation light. Since the probability of multiphoton excitation is extremely low, an excitation light source such as a femtosecond pulsed laser is required to provide a high peak flux of photons to make sure detectable fluorescence [7,8]. Also, due to the absorption of multiple photons, the probability of multiphoton excitation, which means the light intensity of the fluorescence, increases exponentially with the excitation light intensity. Multiphoton fluorescence microscopy that commonly used includes two-photon (2-P) microscopy [46-48] and three-photon (3-P) microscopy [49-51], where the light intensity of the fluorescence is respectively proportional to the square and the cube of the excitation light intensity.

A multiphoton fluorescence microscope is depicted in Figure 2-6. Similar to the design of the confocal fluorescence microscope, the laser beam for illumination is tightly focused and scanned across the focal plane, with the fluorescence detected by a

highly sensitive detector like a PMT to finally construct a full image in a pixel-by-pixel manner. Nevertheless, rather than using a pinhole, the rejection of the background signal from the defocus planes is realised by the localisation of fluorescence excitation here, which means that most fluorescence comes from the tiny focal region of illumination. Likewise, 3-D imaging can be accomplished by moving the objective lens to image optical sections at different depths.

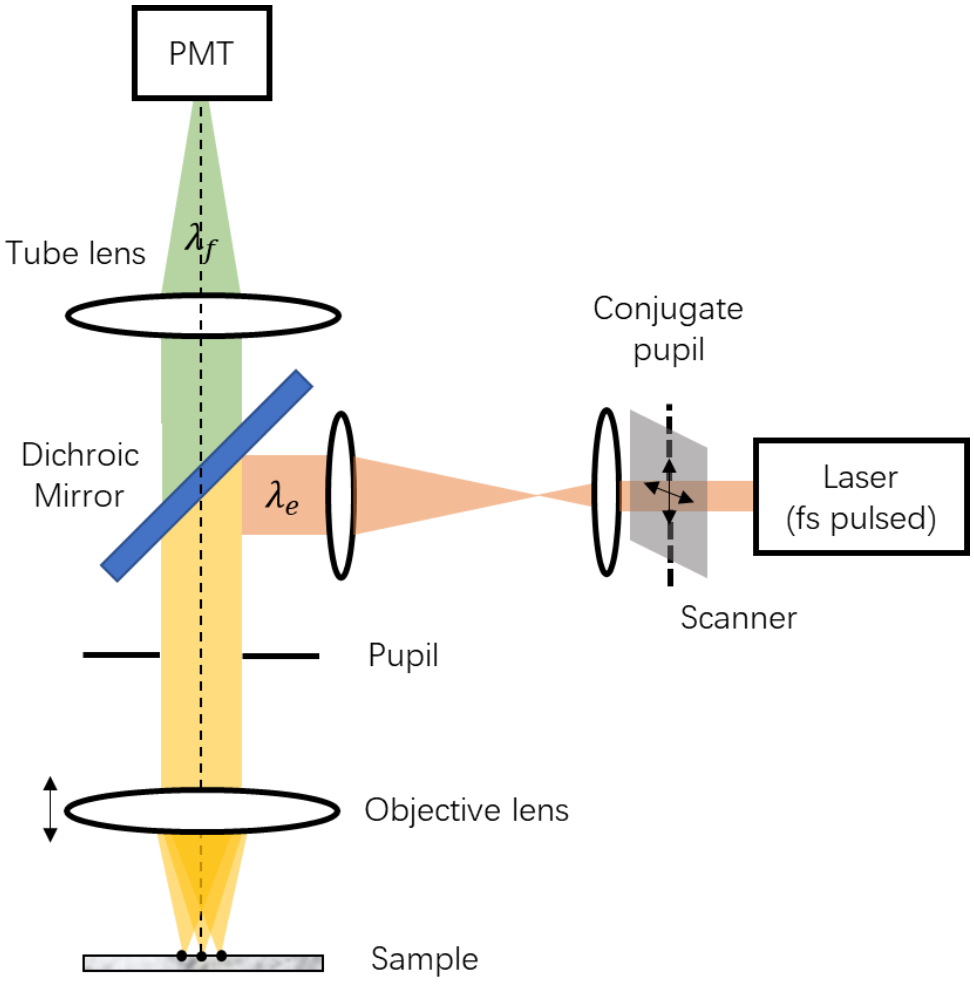


Figure 2-6: A diagram of a multiphoton fluorescence microscope. The femtosecond pulsed laser beam of a wavelength λ_e is scanned by a scanner at a conjugate pupil before reflected by the dichroic mirror and focused by the objective lens onto different spots in the sample for illumination; the emitted fluorescence of a wavelength λ_f is collected by the objective lens, transmitted by the dichroic mirror and refocused by the tube lens to be

detected by a PMT. Signal detected from different scanned spots can collectively build up a 2-D image of an optical section centred with the scanned plane. By moving the objective lens vertically, multiple optical sections can be imaged at different depths and used for reconstruction of a 3-D volume in the sample. For simplicity, the scanner is compacted on a single plane here while in real practice it normally consists of a pair of galvanometer mirrors that positioned on a couple of conjugate planes relayed by 4-f systems.

With constant excitation light field on the pupil, the light intensity of the excitation focal spot can be again described by the equation (2-36), which is in the same form as $|h|^2$ with $\lambda = \lambda_e$. Due to multiphoton excitation, the detected image signal along with the focal spot scanning can be computed as a pixel array by:

$$I_D = O \otimes_{3-D} |h|^{2\kappa} \tag{2-38}$$

where $|h|^{2\kappa}$ defines the multiphoton PSF and κ is the order of non-linearity with $\kappa = 2$ for 2-P and $\kappa = 3$ for 3-P. The pixel dimensions and the pixel size are still defined respectively by the scanning dimensions and the scanning step. Also, the detected image D can be obtained by the equation (2-34).

As the multiphoton fluorescence microscope equally provides the optical sectioning, the resolution is also decided in the same way as the confocal fluorescence microscope, though the estimation from the equation (2-26) might be slightly worse as λ_e here is longer [43]. Apart from this, the multiphoton fluorescence microscope has many advantages compared to the confocal fluorescence microscope. It typically uses IR excitation light, which is scattered less in typical biological tissues and allows deeper imaging [52,53]. In addition, the excitation light of lower energy along with the localised excitation leads to less photodamage outside of the focal region [54,55].

Furthermore, as the fluorescence is no longer filtered by the pinhole, the signal detection is much more effective and the de-scanning is not required.

2.3 Adaptive optics

In an optical imaging system, if the image quality is only limited by the diffraction of the pupil, then the imaging is diffraction-limited. It however marks a relatively ideal state, while the actual image is commonly further degraded by aberrations. Adaptive optics (AO) is a technique that works by measuring aberrations and compensating for them with adaptive elements introduced into the imaging system. It helps the system to approach diffraction-limited imaging [9-13].

2.3.1 Aberration

In wave optics, aberrations can be defined as distortions in the light wavefront. They generally come from imperfections of optics, refractive index mismatches in the light path and inhomogeneous sample structures [9-13]. Although aberrations may vary with light wavelength (chromatic aberrations) [15], monochromatic light is assumed here to consider aberrations.

A monochromatic aberration can be represented by a distorted phase wavefront ϕ, which can be approximately described by a combination of different kinds of modes such as:

$$\phi = \sum_j^N \phi_j \tag{2-39}$$

where N is the total number of modes and ϕ_j is the j-th mode. Hence, there are various aberration representations [14]. An aberration representation can be orthogonal

or non-orthogonal. The orthogonality is generally defined as the inner product between each pair of different modes in the representation being zero, while the inner product may be calculated in different ways [56]. The size of the aberration can be measured by the root mean square (RMS) of ϕ [15]. If orthogonal modes are used to describe ϕ, the RMS can be calculated by:

$$RMS_{\phi} = \sqrt{\frac{1}{N}\sum_{j}^{N} {\phi_j}^2} \tag{2-40}$$

When the aberration exists in an imaging system, ϕ can be considered as the wavefront in the pupil. This will cause the pupil function multiplied with an extra phase factor $e^{i\phi}$ when calculating the PSF, so that the PSF tends to become wider and have a lower peak. As a result, the resolution of the imaging system is prone to decrease and the image quality drops [9-13]. A common metric used to evaluate the image quality degradation induced by the aberration is the Strehl ratio, which is defined as the ratio of the peak aberrated image intensity from a point source compared to its diffraction-limited counterpart [15]. For small aberrations, the peak image intensity can be assumed at the image centre. When the image centre is on the optical axis, the Strehl ratio can be roughly calculated by:

$$S = \left|\langle e^{i\phi}\rangle\right|^2 \tag{2-41}$$

where $\langle e^{i\phi}\rangle$ represents the average of $e^{i\phi}$ taken over the pupil. When $RMS_{\phi} \leq 1$ rad, the relation between the Strehl ratio and the aberration size can be roughly given by [15]:

$$S = 1 - {RMS_{\phi}}^2 \tag{2-42}$$

When the imaging system is free of aberration, $RMS_\phi = 0$ so that the wavefront in the pupil is flat and $S = 1$ which means the diffraction-limited image.

2.3.2 Zernike polynomials

Among all the aberration representations, Zernike polynomials are the most commonly used [15]. They describe aberrations over a circular wavefront, such as the one in the pupil of a typical imaging system using circular lenses. Zernike polynomials are orthogonal to each other over a unit disk. By applying a normalisation to make sure zero mean and unity variance over the unit disk, Zernike polynomials can be defined as [57]:

$$Z_l^m(\rho,\varphi) = \begin{cases} \sqrt{2}R_l^{-m}(\rho)\sin(-m\varphi), & m < 0 \\ R_l^0(\rho), & m = 0 \\ \sqrt{2}R_l^m(\rho)\cos(m\varphi), & m > 0 \end{cases} \tag{2-43a}$$

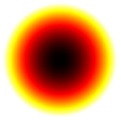
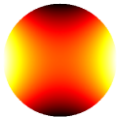
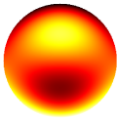
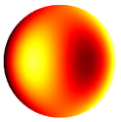
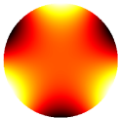
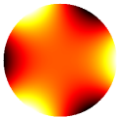
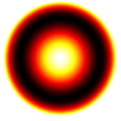
$$R_l^m(\rho) = \sqrt{l+1}\sum_{s=0}^{(l-m)/2} \frac{(-1)^s(l-s)!}{s!\left(\frac{(l+m)}{2}-s\right)!\left(\frac{(l-m)}{2}-s\right)!}\rho^{l-2s} \tag{2-43b}$$

where (ρ,φ) defines the polar coordinates on the unit disk with ρ the radial distance and φ the azimuthal angle; l and m are nonnegative integer indices representing the radial order and azimuthal order respectively and constrained to the conditions $l \geq |m|$ and $l - |m|$ being even. The double indices of Zernike polynomials can be further mapped to a single positive integer Noll's index j [58].

The first 11 Zernike polynomials according to the Noll's index are listed in Table 2-1. It can be seen that each Zernike polynomial represents a different aberration mode, so the overall wavefront can be described by a linear combination of N Zernike modes:

$$\phi = \sum_j^N a_j Z_j \tag{2-44}$$

Table 2-1. Zernike polynomials (the first 11 in Noll's index)

j	l	m	$Z_l^m(\rho,\varphi)$	Aberration term	Shape
1	0	0	1	Piston	
2	1	1	$2\rho\cos(\varphi)$	Horizontal tilt	
3	1	-1	$2\rho\sin(\varphi)$	Vertical tilt	
4	2	0	$\sqrt{3}(2\rho^2-1)$	Defocus	
5	2	-2	$\sqrt{6}\rho^2\sin(2\varphi)$	Oblique astigmatism	
6	2	2	$\sqrt{6}\rho^2\cos(2\varphi)$	Vertical astigmatism	
7	3	-1	$2\sqrt{2}(3\rho^3-2\rho)\sin(\varphi)$	Vertical coma	
8	3	1	$2\sqrt{2}(3\rho^3-2\rho)\cos(\varphi)$	Horizontal coma	
9	3	-3	$2\sqrt{2}\rho^3\sin(3\varphi)$	Vertical trefoil	
10	3	3	$2\sqrt{2}\rho^3\cos(3\varphi)$	Oblique trefoil	
11	4	0	$\sqrt{5}(6\rho^4-6\rho^2+1)$	Primary spherical	

where a_j is the coefficient of normalised Z_j. Due to the orthogonality of Zernike modes, the coefficients are linearly independent and the aberration can be represented

by a unique set of coefficients. Moreover, the RMS of ϕ can be simply calculated by:

$$RMS_{\phi} = \sqrt{\sum_j^N {a_j}^2} \tag{2-45}$$

It is notable that piston (Z_1) and tilt (Z_2 and Z_3) are mostly not considered for aberration correction as they do not deform the image. Defocus (Z_4) is technically the lowest-order aberration to be considered when a thin object is imaged, while it is often left out for a thick object as it simply equals to imaging at a different depth [57,59].

2.3.3 Adaptive elements

Adaptive elements are reconfigurable devices introduced into the imaging system for wavefront modulation. Common adaptive elements used in AO include deformable mirrors (DM) and spatial light modulators (SLM) [9-13,60].

A DM is a mirror whose surface can be deformed by electrical or mechanical actuators, so it can modulate light of arbitrary polarisation or wavelength with the number of actuators determining the number of degrees of freedom [61]. There are different DM concepts. A segmented DM consists of many flat mirror segments each controlled independently by different actuators [62]. The advantage of this DM concept lies in absence of crosstalk between actuators, while it works poorly to modulate continuous wavefront as gaps between mirror segments lead to light scattering and diffraction losses. The gaps can be minimised by introducing extra actuators to allow the tilt of each mirror segment [60]. Another common DM concept is based on a continuous reflective membrane whose shape depends on the combined influence from a number of actuators [63]. Such a DM concept works better modulating continuous

wavefront and allows a great number of actuators when it is driven by micro-electromechanical systems (MEMS), though calibration is often needed before use to map the control voltages to the corresponding change in the membrane shape [64].

A SLM is typically a pixelated liquid crystal (LC) device. Each SLM pixel is individually controlled by applying a voltage to rotate the local LC molecules. As the LC molecules are birefringent, the rotation will cause a phase shift to the passing light along the extraordinary axis of LC molecules [65]. In this way, a SLM can modulate the intensity, phase or polarisation of light with high precision. When it is used exclusively for phase modulation, the input light is required to be of linear polarisation along the extraordinary axis of the LC molecules. Also, the performance of SLM is dependent on the light wavelength [66].

2.3.4 Aberration measurement

A fundamental part of AO is the ability to measure aberrations. There are mainly two categories of methods to implement aberration measurement. The methods in the first category employ a separate wavefront sensor such as a Shack-Hartmann sensor [67-70] or an interferometer [57,59,64]. These methods are direct and fast, whereas they require a complex optical design and suffer from non-common path errors. Also, they generally fail to work when it is impossible to place a sensor at the target location or when the light for sensing is weak [10-13]. On top of this, different wavefront sensors have their own specific limitations. For example, a Shack-Hartmann requires the sensed light from a confined small area and an interferometer requires the sensed light to be coherent.

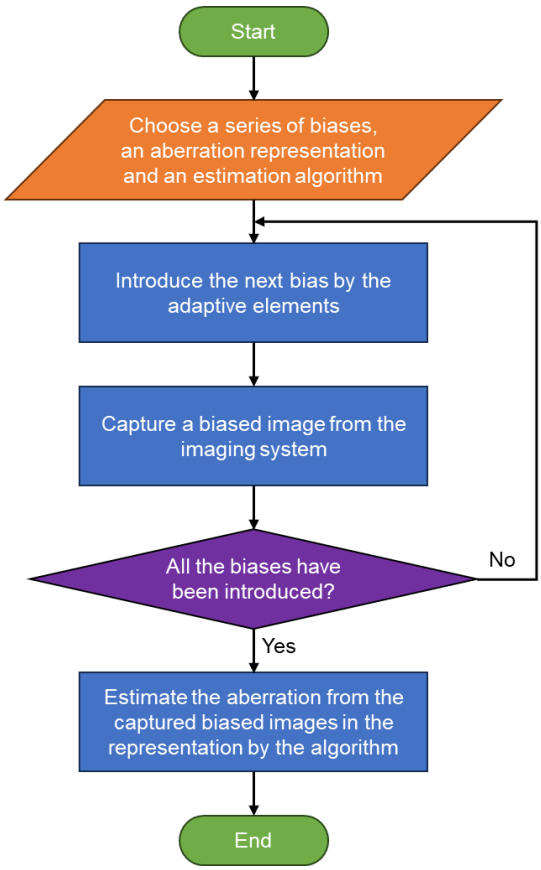


Figure 2-7: A flow chart of the sensor-less aberration measurement process.

The other category of methods operates without a wavefront sensor, so they are known as sensor-less methods [9-14]. Sensor-less methods evolve from phase retrieval schemes such as in [71] where the phase was deduced from intensity measurements in multiple planes of different amount of defocus. In sensor-less methods, the adaptive elements are reconfigured multiple times to deliberately introduce a set of pre-defined wavefront modulations into the imaging system. These wavefront modulations are called bias aberrations (or "biases" for short) and they are determined by a specific scheme to bring so-called phase diversity [72, 73]. For each bias, a biased image is taken, so a sequence of differently biased images can be obtained corresponding to the whole set of biases. The aberration is then estimated indirectly from the biased images

in a certain type of representation by a chosen algorithm. A flow chart of the sensor-less aberration measurement process is presented in Figure 2-7. Sensor-less methods are generally preferred as they function with a simpler optical design and are more easily adapted for a wide range of scenarios. However, they tend to be more time-consuming because they need to take multiple images sequentially and they normally rely on iterations to get a good estimation of the aberration. This is coupled with long and repetitive exposure, which potentially means motion related errors and photodamage [9-13]. A more detailed discussion of sensor-less AO methods can be found in the next chapter.

2.3.5 Aberration correction

According to the measured aberration, the adaptive elements are controlled for aberration correction in AO. Typically, the adaptive elements are positioned at the pupil or its conjugate planes of the imaging system [9-14], so the measured aberration should be firstly converted into a wavefront $\hat{\phi}$ in the pupil. Then, the adaptive elements are controlled to provide a wavefront modulation equivalent to $-\hat{\phi}$ therefore compensate for the aberration.

2.4 Information theory

Information is an abstract concept that can be thought of as the resolution of uncertainty [74]. It can be mathematically quantified by information theory, which was established and formalised by Shannon in the 1940s [75]. Information theory is based on probability theory and statistics, so it associates measures of information with random variables

[74].

2.4.1 Information content

In information theory, the information content is also called self-information or Shannon information [74]. It is derived from the probability of a particular event occurring from a random variable. Given a random variable X of a probability distribution $p_X(x)$, the information content of measuring X as outcome x is defined by:

$$\mathcal{I}_X(x) = -\log_b[p_X(x)] \tag{2-46}$$

This means that the occurrence of a less probable event provides more information. The choice of the logarithmic base b determines the unit of information: when $b = 2$, the unit is the bit or the Shannon; when $b = e$, the unit is the natural unit of information; when $b = 10$, the unit is the Hartley [74]. Since an arbitrary unit can be taken, b is hereafter ignored. For multiple independent events, the total information content is the sum of the information content of individual events.

2.4.2 Entropy

The entropy or the Shannon entropy is the expected value of the information content [74]. For the same random variable X, the entropy can be calculated by:

$$\mathcal{H}(X) = \mathbb{E}[\mathcal{I}_X(x)] = -\sum_{\Omega_x} p_X(x) \log p_X(x) \tag{2-47}$$

where $\mathbb{E}$ calculates the expected value and Ω_x is the outcome space of x. The entropy measures the amount of information in a random variable. A high entropy suggests that the outcome is more evenly distributed, thereby making it harder to predict.

When another random variable Y of a probability distribution $p_Y(y)$ is considered, the joint entropy of X and Y can be calculated by:

$$\mathcal{H}(X,Y) = \mathbb{E}\big[-\log p_{X,Y}(x,y)\big] = -\sum_{\Omega_{x,y}} p_{X,Y}(x,y)\log p_{X,Y}(x,y) \quad (2\text{-}48)$$

where $p_{X,Y}(x,y)$ is the joint probability distribution of (X,Y) and $\Omega_{x,y}$ is the joint outcome space of (x,y). If X and Y are independent, their joint entropy becomes the sum of their individual entropies. If the value of Y is known, the conditional entropy of X given Y is defined as:

$$\mathcal{H}(X|Y) = -\sum_{\Omega_{x,y}} p_{X,Y}(x,y)\log p_{X|Y}(x|y) \quad (2\text{-}49)$$

where $p_{X|Y}(x|y)$ is the conditional probability distribution of X given Y. A basic property can be given by:

$$\mathcal{H}(X|Y) = \mathcal{H}(X,Y) - \mathcal{H}(Y) \quad (2\text{-}50)$$

2.4.3 Relative entropy

The entropy provides the basis of many other measures of information. Among them, the relative entropy or the Kullback-Leibler (KL) divergence defines the expected difference in the information content from using $q_X(x)$ as a model probability distribution of X instead of the true probability distribution $p_X(x)$ [74]. Mathematically, it can be expressed as:

$$D_{KL}\big(p_X(x)||q_X(x)\big) = \sum_{\Omega_x} p_X(x)\log\frac{p_X(x)}{q_X(x)} \quad (2\text{-}51)$$

Since the relative entropy measures the difference between the model and the ground truth, one can find a $q_X(x)$ that is closest to $p_X(x)$ by minimising the relative entropy. Such a process is called information projection as the relative entropy is treated as a

distance metric between the two probability distributions. However, the relative entropy is not a true metric because it is not symmetric, i.e. $D_{KL}\big(p_X(x)||q_X(x)\big) \neq D_{KL}\big(q_X(x)||p_X(x)\big)$, and fails to satisfy the triangle inequality [74].

2.4.4 Fisher information

The Fisher information is the expected value of the observed information. If X is observable and its probability distribution $p_X(x)$ is parameterised by Θ, the Fisher information measures the amount of information carried by X about Θ [74]. It can be computed by:

$$\Im_X(\Theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\Theta}\log p_X(x;\Theta)\right)^2\right] \tag{2-52}$$

Here, $p_X(x;\Theta)$ is used to represent $p_X(x)$ parameterised by Θ. Intuitively, the Fisher information can be interpreted as the changes in $p_X(x;\Theta)$ with respect to changes in Θ. A larger change to $p_X(x;\Theta)$ resulting from a change of Θ means that X carries more information about Θ. When Θ is a vector of multiple parameters, the Fisher information becomes a matrix in which:

$$[\Im_X(\Theta)]_{l,m} = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_l}\log p_X(x;\Theta)\right)\left(\frac{\partial}{\partial\theta_m}\log p_X(x;\Theta)\right)\right] \tag{2-53}$$

Here, θ_l and θ_m are elements of Θ indexed by l and m respectively. Such a matrix is equivalent to the infinitesimal form, specifically the Hessian, of the relative entropy $D_{KL}\big(p_X(x;\Theta)||p_X(x;\Theta')\big)$, where Θ' is another set of parameters close to Θ [74].

2.4.5 Cramér-Rao bound

The Cramér-Rao bound states that the inverse of the Fisher information is a lower bound

on the variance of any unbiased estimator $\widehat{\Theta}$ of Θ from X [74], that is:

$$CRLB_{\widehat{\Theta}} = \Im_X(\Theta)^{-1} \leq \mathbb{V}\left(\widehat{\Theta}\right) \tag{2-54}$$

Here, $CRLB_{\widehat{\Theta}}$ denotes the Cramér-Rao lower bound (CRLB) of $\widehat{\Theta}$ and $\mathbb{V}$ calculates the variance. When Θ is a vector of multiple parameters, the CRLB of each element $\hat{\theta}_m$ of $\widehat{\Theta}$ can be calculated by:

$$CRLB_{\hat{\theta}_m} = [\Im_X(\Theta)^{-1}]_{m,m} \leq \mathbb{V}\left(\hat{\theta}_m\right) \tag{2-55}$$

When it is difficult to calculate the inverse of the Fisher information matrix, a possibly loose lower bound can be found by $[\Im_X(\Theta)]_{m,m}{}^{-1}$. An unbiased estimator that achieves its CRLB is fully efficient and of the lowest possible mean squared error (MSE) [74].

2.5 Deep learning

Conventional algorithms are explicitly programmed for each case, so they tend to rely on strong assumptions, perform poorly with high-dimensional data and struggle to capture complex non-linear relationship. In contrast, learning-based algorithms are designed to learn patterns implicitly from data and generalise to unseen data, therefore they are much more flexible without strong reliance on assumptions, well-suited for high-dimensional data and powerful at capturing complex non-linear relationship automatically [76]. For image processing, most high-performance learning-based algorithms are based on deep learning (DL), which uses neural networks (NN) of multiple layers [77].

2.5.1 Neural network

The NN or artificial NN is a computational model inspired by the structure of biological NN [76]. It is composed of connected units or nodes called artificial neurons. Each neuron receives inputs and processes them to produce a single output that can be sent to other neurons connected to it as an input. The initial input U includes feature values of an example of external data, and the ultimate output $\hat{V}$ is expected to fulfil the task. Thus, a NN establishes the mapping from U to $\hat{V}$ as:

$$\hat{V} = NN(U) \tag{2-56}$$

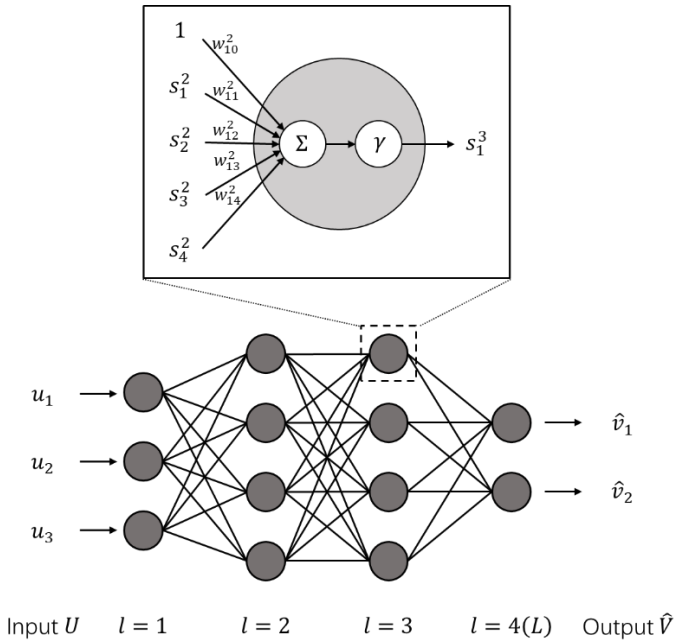


Figure 2-8: The model architecture of a traditional NN or MLP. The neurons are represented by circles and the connection between neurons are represented by edges. The NN takes elements of the input U, passes them through 4 (L) fully connected layers of neurons indexed by l, and generates elements of the output $\hat{V}$. The example neuron in the box is expanded to illustrate the calculation in it.

The model architecture of a traditional NN, also known as a multilayer perceptron (MLP) [77], is shown in Figure 2-8. In this NN, the neurons are organised into fully connected layers, where each neuron in one layer connects to each neuron in the next

layer. With l as the layer index, U is fed into the input layer ($l = 1$) such as:

$$s_k^{(1)} = u_k \tag{2-57}$$

where $s_k^{(1)}$ denotes the output of the k-th neuron of the input layer and u_k denotes the k-th element of U. Similarly, $\hat{V}$ emerges from the output layer ($l = L$) such as:

$$s_k^{(L)} = \hat{v}_k \tag{2-58}$$

where $s_k^{(L)}$ denotes the output of the k-th neuron of the output layer and $\hat{v}_k$ denotes the k-th element of $\hat{V}$. The layers between the input and output layer are the hidden layers ($l = 2, \cdots, L-1$). A NN with more than one hidden layer are typically called a deep NN, which are used for DL [77]. The number of layers L gives the NN depth. Other than the input layer, the neurons in the rest of the layers ($l = 2$ to L) calculate their output by:

$$s_k^{(l)} = \gamma\left(\sum_j w_{kj}^{(l-1)} s_j^{(l-1)} + w_{k0}^{(l-1)}\right) \tag{2-59}$$

Here, $s_k^{(l)}$ denotes the output of the k-th neuron of the l-th layer; γ computes an activation function that is mostly nonlinear; the part in the parentheses are called the activation, in which $\sum_j w_{kj}^{(l-1)} s_j^{(l-1)}$ is a weighted sum of all the inputs, with $w_{kj}^{(l-1)}$ the weight of the connection from the j-th neuron of the previous layer and $s_j^{(l-1)}$ the input from the same neuron, and $w_{k0}^{(l-1)}$ is a bias term. The weights and biases are the parameters of the NN, which control the mapping relationship.

2.5.2 Neural network training and testing

Training a NN is teaching it to generate output that can better accomplish the task [77]. It is realised by an iterative process where the parameters of the NN are adjusted to

optimise a loss or cost function evaluating the performance. The methods for training can be separated into three main paradigms: supervised learning, unsupervised learning and reinforcement learning. Supervised learning feeds labelled data into the NN, with the label in each example providing the desired output corresponding to the input, and trains the NN to minimise the difference between its output and the label [77]; Unsupervised learning gives the NN unlabelled data and trains the NN to find hidden patterns or properties in the data [77]; In reinforcement learning, the NN acts as an agent that learns by making trials and getting feedbacks from an environment until it develops a policy to minimise the long-term cost [78]. The choice of the training method depends on the task and the state of data. After the NN has been trained, its performance is tested on data or cases that were never presented during training [77].

For the DL tasks involved in this thesis, the training of NN is mostly implemented in the fashion of supervised learning. As the preparation for each training, a labelled dataset is collected and split into three parts: the training set, the validation set and the testing set. The training set is the largest one which is used to train the NN; the validation set is also used during training but to monitor progress; the testing set is used after training to evaluate the final performance of the trained NN.

The training starts with the initialisation of the NN parameters. Next, a batch of input data $\{U_m\}$ $(m = 1, \cdots, M)$ from the training set is fed into the NN, with M the batch size, and passed through it to output $\{\hat{V}_m\}$ based on the equation (2-56). Such a step is called forward propagation. With the corresponding labels of $\{U_m\}$ being $\{V_m\}$,

a loss is calculated by integrating the difference between each pair of V_m and $\hat{V}_m$ as:

$$\mathcal{L} \equiv \sum_m \mathcal{E}\left(V_m, \hat{V}_m\right) \tag{2-60}$$

where $\mathcal{E}$ is a distance metric. A backward propagation step is then operated by calculating the gradients of the loss $\mathcal{L}$ with respect to each NN parameter w to determine its contribution to $\mathcal{L}$. As the activation functions in the NN are often non-differentiable, approximate gradients $\left[\frac{\partial \mathcal{L}}{\partial w}\right]$ are calculated instead. Based on the gradients, a gradient descent optimisation algorithm updates the NN parameters to reduce the loss by:

$$\widetilde{w} = w - \eta \left[\frac{\partial \mathcal{L}}{\partial w}\right] \tag{2-61}$$

where $\widetilde{w}$ represents the updated w and η defines the learning rate. Except for the initialisation, all these procedures are iterated with different batches of data until a full pass is made through the entire training set, which is called an epoch. The iterations are normally carried on for multiple epochs as required.

To evaluate the performance of the NN, another dataset completely separated from the training set is required to calculate a validation loss according to the equation (2-60). If both the validation loss and the training loss are high, the NN is underfitting which happens when the NN learns not enough from the training set. To solve this, the NN should be trained longer or built with more complex architecture. If the validation loss is high but the training loss is low, the NN is overfitting which means that the NN learns the training set too well but generalises poorly to unseen data. To prevent overfitting, regularisation can be applied during the training. Common approaches for

regularisation include early stopping, which means to stop training before the NN memorises the training set, weight decay (L1 and L2 regularisation), which adds a penalty term onto the training loss to suppress the NN parameters, and dropout, which randomly deactivates a certain percentage of neurons during the training [77].

In addition to the NN parameters, the performance of the NN also depends on how the NN is configured. The high-level parameters that are set to define the configuration of the NN are called hyperparameters. Hyperparameters generally come from the model architecture, the training process and the regularisation [77]. Typical hyperparameters of the model architecture are the NN depth, the neuron number or the width per layer, the choice of each activation function. Hyperparameters of the training process include the choice of the initialisation, the batch size, the choice of the loss function, the learning rate, the choice of the optimisation algorithm, the number of epochs, etc. Regularisation hyperparameters are such as the timing for early stopping, the regularisation coefficient for weight decay and the dropout rate. Unlike the NN parameters that can be learned from the training data, hyperparameters must be tuned separately to optimise the NN performance.

To ensure optimised NN performance, the validation set is used repeatedly after each training epoch to evaluate the NN performance, which serves as the basis to tune the hyperparameters. Once the NN is finalised and all hyperparameter tuning is completed, its performance is checked for the one last time with the testing set consisting of data that has never appeared throughout the training.

2.5.3 Convolutional neural network

Convolutional neural network (CNN) is a type of NN designed to process structured grid data like images. Although a traditional MLP can also process images by taking each pixel as an input feature, it ignores the spatial structure in images and becomes computationally intractable with images of higher resolution. In contrast to an MLP, a CNN exploits the spatially local correlation presented in images to mitigate these challenges [77].

The model architecture of a typical CNN is presented in Figure 2-9. In this CNN, the neurons are organised into 3-D layers and the hidden layers consist of convolutional layers, pooling layers and traditional fully connected layers, while different layers may be found in a different CNN.

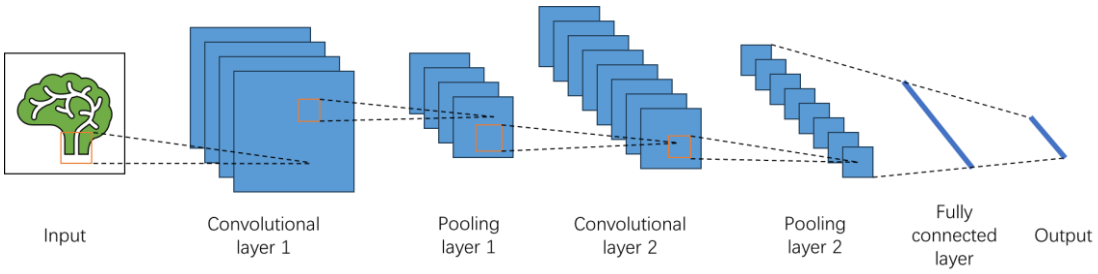


Figure 2-9: The model architecture of a typical CNN. The CNN consists of different layers whose dimensions are illustrated by blue shapes. The black dotted lines between layers indicate their connections. The orange box linked to each layer marks either the receptive field of each neuron in a convolutional layer or the pooling window of a pooling layer. From the input image, the convolutional layer 1 generates feature maps (squares) by convolutions with different filters and the pooling layer 1 reduces the size of feature maps. The feature maps are further combined and reduced in size by the convolutional layer 2 and the pooling layer 2, until they are flattened into a vector (line) and processed by a fully connected layer to generate the output.

In a convolutional layer, each neuron is only connected to neurons in a restricted region local in width and height but extending through the full depth of the previous layer. Such a region is called the receptive field of the neuron. The output of each neuron can still be calculated following the equation (2-59). For all neurons in a single 2-D depth slice, they share the same set of parameters calculating their output, so the calculation is equivalent to a convolution between the input volume of the layer and the set of parameters as a filter. Assuming the input volume of both the width and the height w in addition to padding around the borders of a size p and the filter of both the width and the height f moving in a stride of s during the convolution, the calculation result is a 2-D array of both the width and the height $\left\lfloor \frac{(w-f+2p)}{s} + 1 \right\rfloor$, where $\lfloor \cdot \rfloor$ is a round-down operator. Such a 2-D array is called a feature map which represents a particular feature of the input volume. The output volume of the layer is hence a stack of 2-D feature maps generated by different filters along the depth of the layer. For a stack of convolutional layers, the initial layers extract local simple features from the input, which will be progressively combined into global complex features in the deeper layers [79].

A pooling layer contains no neuron but performs a fixed pooling operation on the feature maps in the output volume of a preceding convolutional layer. It works by sliding a window of both the width and the height f' in a stride of s' greater than 1 over each feature map independently, and aggregating the values within the window into a single value which is commonly their maximum (max pooling) or average

(average pooling). Thus, the depth of the volume remains unchanged while the width and the height is reduced by a factor of s' after the pooling. A global pooling reduces both the width and the height to 1 thereby turning the volume into a vector along the depth.

After several convolutional and pooling layers, the output volume is flattened into a vector and passed on to one or more fully connected layers. The fully connected layers in a CNN function in the same way as those in an MLP. They are responsible for generating the ultimate output to fulfil the task based on the features provided by the earlier layers.

The training and the testing of the CNN follows the same process as that of the MLP. It is noted that the pooling layers have no parameters so they are not adjusted during the training. Compared to the MLP, the CNN drastically decreases the number of parameters required for processing images by the local connectivity and the parameter sharing in convolutional layers and the use of pooling layers, so it is less likely suffering from overfitting and has lower computational cost. On top of this, the CNN is more robust to spatial translations of features in the input images as the features are extracted regardless of their positions [77].

2.5.4 Residual neural network

To deal with complex tasks, more layers tend to be stacked in a NN to make it deeper so that it can model a more complex mapping relationship and achieve better performance [80]. However, as the NN depth keep increasing, the performance will

start to saturate and then degrade rapidly [81]. Such a degradation problem is not a result of overfitting but the difficulty in training a very deep NN. The main reason behind the difficulty is the vanishing / exploding gradients, namely that the gradients of the loss become extremely small or large after propagating backward through many layers [82,83]. This causes the parameters in earlier layers fail to be updated effectively during the training.

To address the problem, residual neural networks (ResNet) organise layers into residual blocks [84]. In each residual block, a shortcut connection is used to bypass the layers in the block. The short connection performs identity mapping of the block input S, and its output is added to the output of the block layers to form the block output $\Gamma(S)$. This means that the block layers only model a residual mapping $\Gamma(S) - S$. Such a structure enables a more direct path for propagating through layers thereby allowing for successful training of extra deep ResNet to obtain higher performance on complex tasks.

Among different versions of ResNet, ResNet-18 is a commonly used one whose architecture is provided in Figure 2-10. There are in total 18 layers containing trainable parameters in this model. Aside from the initial convolutional layer and the last fully connected layer, all the other convolutional layers have 3×3 filters and are grouped in pairs and each pair forms a residual block. As common rules to design the convolutional layers, the size of the feature map is either maintained the same or halved. If the size maintained the same, the number of filters is also kept the same; if the size halved, the

number of filters is doubled. The activation function for all the convolutional layers is rectified linear unit (ReLU) [85]. Also, batch normalisation (BN) is adopted in each convolutional layers after the convolution and before the activation function, which makes training faster and more stable [86]. By adapting the output layer, this network can be easily used for different tasks.

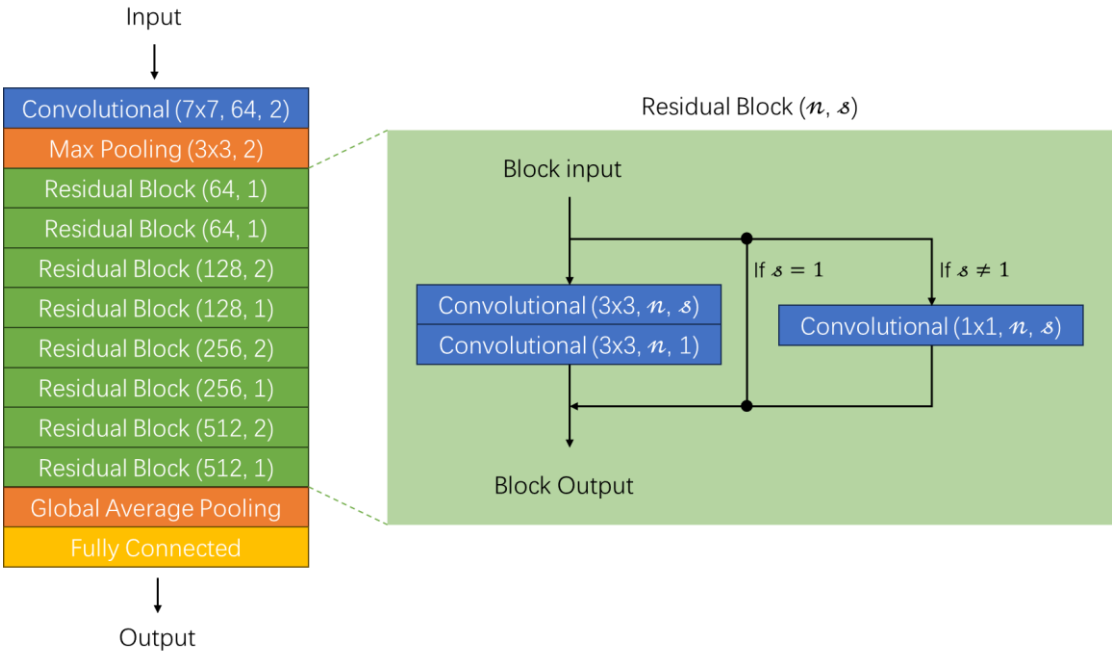


Figure 2-10: The model architecture of ResNet-18. All the layers are illustrated by boxes arranged according to their order in the network. A blue box labelled as "Convolutional ($f \times f$, n, s)" represents a convolutional layer where the convolutions are operated with n filters of size $f \times f$ moving in a stride of s. An orange box labelled as "Max Pooling ($f' \times f'$, s')" represents a max pooling layer with a $f' \times f'$ window moving in a stride of s'. A green box labelled as "Residual Block (n, s)" represents a residual block of two convolutional layers as shown in the right. If $s = 1$, the shortcut connection performs identity mapping; If $s \neq 1$, a convolutional layer of 1×1 filters is inserted into the shortcut connection to match the dimensions. The network ends with a global average pooling layer and a fully connected layer to generate the final output.

Chapter 3

Sensor-less adaptive optics methods for multiphoton fluorescence microscopy

This chapter provides a review of existing sensor-less AO methods for multiphoton fluorescence microscopy. This starts with a summary of common principles, followed by variations of these methods that induce different method performances. Some methods originally demonstrated in other microscopes are also included as they are suitable for multiphoton fluorescence microscopy with little adaptation.

3.1 Common principles

A sensor-less AO method normally assumes that the aberration keeps invariant in a limited range of time and space. It works in two main stages: image measurement and aberration estimation. To measure each image, a bias aberration is introduced by the adaptive elements into the microscope, which makes the total aberration in the microscope become:

$$\phi = \phi_A + \phi_B \tag{3-1}$$

where ϕ_A denotes the aberration to be estimated and ϕ_B denotes the bias aberration. If the pupil of the microscope is uniformly illuminated, based on the equation (2-36) and the pupil wavefront distortion from the aberration discussed in Section 2.3.1, the light intensity of the excitation focal spot can be calculated by:

$$I_e = |h|^2 = \left|\mathcal{F}\left[Pe^{i(\phi+\zeta)}\right]\right|^2 \tag{3-2}$$

with the notation system kept the same as in Section 2.2. The detected image signal I_D can be obtained by the equation (2-38) with $|h|^{2\kappa}$ defining the multiphoton PSF and the measured image D can be further obtained by the equation (2-34). According to a specific biasing scheme, a set of bias aberrations $\{\phi_B^t\}$ $(t = 1, \cdots, T)$ is introduced sequentially so a series of biased images $\{D^t\}$ are measured. In the stage of aberration estimation, ϕ_A is estimated as $\hat{\phi}_A$ from $\{D^t\}$. The estimation is based on a chosen aberration representation, which means:

$$\hat{\phi}_A = \sum_j^N \hat{\phi}_{Aj} \tag{3-3}$$

where $\hat{\phi}_{Aj}$ is the j-th in a set of N modes. A chosen algorithm G is employed to compute the set of modes from $\{D^t\}$ as:

$$\left\{\hat{\phi}_{Aj}\right\}_{j=1,\cdots,N} = G\left(\{D^t\}_{t=1,\cdots,T}\right) \tag{3-4}$$

If the set of modes are orthogonal to each other, the computation can be conducted mode by mode, otherwise the whole set must be computed together [87-89]. The estimation error $\phi_A - \hat{\phi}_A$ represents the residual aberration after correction, which indicates the performance of the method. For a lower estimation error, the two stages of the method can be iterated multiple times by taking the residual aberration as the new ϕ_A in the next iteration. As the number of iterations increases, the method takes longer time and more image measurements which potentially means more photodamage to the sample. Thus, the time consumption and the required number of image measurements are also considered for a fair evaluation of the method performance. With respect to a

specific imaging scenario, the method performance varies with different choices of three method components: the biasing scheme, the aberration representation and the estimation algorithm.

3.2 Aberration representations

Sensor-less AO methods have used different aberration representations for effective estimation. As mentioned in Section 2.3.1, aberration representations can be separated into orthogonal ones and non-orthogonal ones based on whether the inner product between each pair of different modes is zero. Non-orthogonal modes are typically used to match the control of a continuous membrane DM in which each actuator has an influence function overlapping with that of the others [90-92]. Since non-orthogonal modes must be computed as a whole, the computation is often implemented by a stochastic search algorithm which can require many iterations to converge [87,88,93]. Works have been conducted to convert the non-orthogonal modes into orthogonal ones through DM calibration [56,64,94-96].

In contrast, orthogonal modes are much more commonly used for higher estimation efficiency. There are continuous orthogonal modes each spanning the entire pupil such as Zernike modes [95,97-126], deformable mirror modes [56,64,127] and Lukosz modes [128,129]. The latter two are variants derived from Zernike modes specifically for different scenarios. Alternatively, orthogonal modes can be piecewise. On each piece, the phase is defined by only a constant piston component [106,130-134,143-149] or that with extra two tilt components [101,135-142]. Each piecewise mode can contain

only one single piece confined within a local zone on the pupil [101,130-142] or multiple pieces filling up the whole pupil as in Hadamard-Walsh modes [106,143-147] and polar Walsh modes [148,149].

Research has been undertaken to compare the performances of methods using different kinds of orthogonal modes. In [56], two different sets of deformable mirror modes were compared working with DM of different actuator arrangements whereas only minor differences were observed when the DM had a large number of actuators. In [150], the performances were compared using different variants of Zernike modes including the Lukosz modes, singular value decomposition (SVD) modes and Neumann modes, but the result showed similar performances. A systematic comparison was not made until the work in [14] where a generalised aberration representation covering both continuous modes and piecewise zonal modes was defined to allow a fair comparison. It was concluded that modes each spanning a larger zone on the pupil resulted in better estimation of low-order aberrations of simpler wavefront shape and more robust performance to noise, while modes each spanning a smaller zone worked better with high-order aberrations of more complex wavefront shape but caused the performance more sensitive to the noise. The work in [141] further elaborated the relation between Walsh modes and the other modes. Also, it demonstrated that the use of Walsh modes helped estimating aberrations of more complex wavefront shape with the estimation space reduced into a fundamental Voronoi cell of a lattice geometry while keeping the performance robust to noise. To combine the advantages of different modes in

estimating different types of aberrations, attempts have been made to estimate low-order aberration components in continuous modes and high-order aberration components in piecewise modes [101,106].

3.3 Estimation algorithms

Estimation algorithms used in sensor-less AO methods include both conventional algorithms and learning-based algorithms. Conventional algorithms generally estimate the aberration by optimising an image quality metric, while learning-based algorithms estimate the aberration directly from the measured images.

3.3.1 Conventional algorithms

A conventional algorithm estimates the aberration through an optimisation of a quality metric calculated from each measured image that achieves its global optimum when the residual aberration after correction is minimised. The algorithm is hence defined through two aspects: the choice of the metric and the way to search for the global optimum.

For a multiphoton fluorescence microscope, the most common metric is the average (or total) image intensity [64,87-89,93,95,97-100,103,104,106,107,118,130-140,143,148], while other single-valued choices include the size of the PSF [89,127,128] and the image sharpness [102,129], which is often defined as a function of spatial frequency components in the image. A multi-valued metric has also been introduced in [114] based on a multi-scale wavelet analysis to track spatial frequency variations in the image. These metrics reflect mostly the size but unable to show the shape of the

aberration, so a large number of measurements are required to collect enough information for the algorithm to provide good estimation. On top of this, although these metrics only have one global optimum (assuming the aberration modulo 2π and without piston) when the aberration is small [95], apart from the one used in [114], most of them tend to generate multiple extra local optimums and show little change when the aberration become large [15]. This causes difficulty for the algorithm to estimate a large aberration especially when the SNR is low.

To search for the global optimum of the metric, the algorithm can be based on the image formation model or model-free. A model-free algorithm is generally effective to converge but requires a large number of measurements. When the aberration is estimated in non-orthogonal modes, a model-free stochastic search algorithm such as a genetic algorithm [88,89,93,135], a hill-climbing algorithm [87,88] or a random search algorithm [88] is normally used to provide a global optimisation with respect to the whole aberration. As mentioned in the last section, such an algorithm has a less reliable convergence speed. When the aberration is estimated in orthogonal modes, the optimisation can be realised individually with respect to each mode by a model-free systematic search algorithm which can more efficiently converge [131,134,136-138,142-149].

Since the image formation model is mostly well understood, the function of a metric with respect to the aberration can be derived approximately as a quadratic polynomial [64,95,97,129]. Thus, a model-based algorithm estimates the aberration by

resolving all the parameters in the quadratic function with the measurements and finding the aberration corresponding to the global optimum analytically. Such an algorithm is direct and deterministic so that it requires much fewer measurements compared to a model-free algorithm. When the aberration is represented by N orthogonal modes, a minimum of $N+1$ measurements (from N biases each in a different mode plus a zero bias) is required for a point object (i.e. a focusing system) if all the modes are both estimated and corrected at the same time [99,103,130]; a minimum of $2N+1$ measurements (from N pairs of opposite biases each pair in a different mode plus a zero bias) is required if each mode is estimated separately but corrected at the same time [98,100,104,129]; a minimum of $3N$ measurements (from N groups of biases with a pair of opposite biases in a different mode plus a zero bias in each group) is required if each mode is both estimated and corrected separately [104,118,132,133]. In the case of a large aberration or a low SNR, more than the minimum number of measurements are often required to achieve good estimation, which can cause the strategy with a larger minimum number providing a better performance [104].

3.3.2 Learning-based algorithms

In a learning-based algorithm, the aberration is estimated directly from the measured images by a computational model whose parameters are iteratively adjusted through a learning process to minimise the estimation error. With no need of calculating any image quality metric, the algorithm can employ information about the aberration

contained in measured images much more comprehensively than a conventional algorithm, therefore it can provide good estimation based on only a few measurements.

Different computational models with different learning approaches have been applied to implement learning-based algorithms. Typically, a deep NN model for DL is used as it is powerful and flexible as a universal function approximator [108-113,115-117,119-122,124-126], while there are also algorithms fitting a traditional regression model with a Gauss-Newton method for learning [105,123]. In some DL based algorithms, the NN is designed as a pipeline of distinct modules by incorporating knowledge from the image formation model [117,124]. This helps the interpretation and debugging of the NN. Also, the training of the whole NN can be broken down to the training of modules, which is much easier and can be conducted without a training dataset. However, as the NN is limited in flexibility by artificial constraints, these algorithms share an important drawback - they need to retrain the NN upon each experiment so typically take minutes to run.

In contrast, other DL based algorithms follow an end-to-end NN design to build a direct mapping from the measured images to the estimated aberration [108-113,115,116,119-122,125,126]. Thus, the NN is trained as a whole and automatically learns from data to optimise the estimation. In [119] and [122], the NN was trained by deep reinforcement learning, however these algorithms relied on multiple iterations and showed no advantage in performance compared to conventional algorithms. A more prevalent way to train an end-to-end NN is through supervised learning, which typically

involves the collection of a large dataset for the training [108-113,115,116,120,121,125,126]. As it is impractical to collect such a large dataset through experiments, simulations are conducted based on the image formation model to generate a synthetic dataset in most cases, while in [125] an extra small dataset was collected through experiments to fine-tune the NN. After the NN is properly trained on the dataset, it can estimate the aberration from at least two image measurements and finish the computation almost instantly without the need of retraining for each different experiment. The problems of using an end-to-end NN are its "black box" nature, which makes it less interpretable, and the difficulty in its training, as a large amount of data is required and the process is computationally expensive. When supervised learning is implemented, it is usually challenging to collect a fully eligible dataset that well-represents different possible imaging scenarios, especially when the object is of arbitrary extended structures. This means that the NN in many works is insufficiently trained, so that it can only work on images of point-like objects or PSF measurements [108-113,115,116,120,125,126]. Such a limitation was addressed in [121] where pseudo-PSFs were calculated from pairs of measured images to remove the influence from the object structures based on knowledge from the image formation model. In this case, a simpler NN that is easier to train was used to estimate the aberration with pseudo-PSFs as the input, which allowed the algorithm to work on images of extended objects.

3.4 Biasing schemes

The biasing scheme in a sensor-less AO method is generally determined in a heuristic manner. When a conventional algorithm is used in the method, the biasing scheme is normally restricted to be compatible with the algorithm, which leaves the biases defined only by a few parameters free to adjust. For example, as mentioned in Section 3.3.1, N pairs of opposite biases with each pair in a different mode plus a zero bias is specified for a model-based algorithm to estimate the aberration in N orthogonal modes from $2N + 1$ measurements, so the biasing scheme only depends on the choice of bias amplitude for each mode [98,100,104,129]. While the optimal choice of the free bias parameters was explored in [98], these parameters are mostly determined based on empirical experience.

On the other hand, the biasing scheme can be chosen much more freely when it works with a learning-based algorithm. Yet in most methods applying a learning-based algorithm, the biases are still chosen with the common choices used for conventional algorithms as references. For the biases based on Zernike modes, it has been reported that defocus (if imaging a thin object) and astigmatism generally bring better performance when serving as the biases, while coma and trefoil are less desirable [121,123,125].

3.5 Summary and discussion

In general, a sensor-less AO method takes image measurements according to a chosen biasing scheme and estimates the aberration in a certain representation from the measured images by a specific algorithm. The performance of the method is evaluated

based on not only the estimation error, but also the time consumption and the required number of image measurements. Different methods are defined by different choices in the biasing scheme, the aberration representation and the estimation algorithm. The biasing scheme is widely determined in a heuristic way. About the aberration representation, orthogonal modes are more common choices in which continuous modes are more suitable for estimating low-order aberrations while piecewise modes are more suitable for estimating high-order aberrations. In terms of the estimation algorithm, learning-based algorithms can be more efficient than conventional algorithms as they can extract useful information contained in the measured images more comprehensively rather than only make use of limited information contained in the image quality metric. The state-of-the-art performance has been achieved with a DL based algorithm in which an end-to-end NN is trained on a large dataset through supervised learning and hence able to estimate the aberration from at least two images without the need of retraining for each different experiment.

The major problem shared by existing sensor-less AO methods is the large time consumption as they need to take multiple images in a sequential manner and often rely on iterations to refine their estimation. Even though some recent methods have demonstrated the potential of estimating the aberration based on only two images by a DL based algorithm, much more images are usually required in many scenarios to reach a satisfactory estimation. Therefore, it is important to further accelerate sensor-less AO. To achieve this, a key step is to optimise the information contained in the measured

images for aberration estimation, which can be implemented in two major directions: (1) to maximise information about the aberration encoded in the measured images by optimising the biases; (2) to concentrate information separated in multiple images into one image by measuring the images in parallel. In the next two chapters, two studies are presented to optimise sensor-less AO following the two directions correspondingly.

Chapter 4

Information guided optimisation of sensor-less adaptive optics methods for multiphoton fluorescence microscopy

Sensor-less AO generally requires taking multiple images to support the aberration estimation. To reduce the time consumption, a valid way is to increase the efficiency in estimating the aberration from the images. This naturally demands more information about the aberration should be extracted, which depends on how information is encoded in the images. In sensor-less AO, the biasing scheme conditions the image measurement so it determines the encoding of information. Since the biasing scheme has been mostly chosen based on empirical experiences, one should not expect that the images are optimally conditioned for aberration information extraction. Thus, it is important to optimise the biasing scheme objectively to better condition the images so that sensor-less AO can work more efficiently.

As shown in Figure 4-1, a framework was introduced in this chapter to provide information guided optimisation of the biasing scheme in sensor-less AO. In this framework, a numerical model was built to describe sensor-less AO. Based on this model, information encoded in the images was analysed by calculating the Fisher information. According to the Cramér-Rao bound, a loss function could be derived from the Fisher information representing the accuracy of the aberration estimator when it

fully extracted the information about the aberration from the images. With such a loss function, optimisation of the biases could be performed objectively to result in better conditioned images and thus facilitate higher sensor-less AO performance. This framework is not specific to any sensor-less AO method or imaging scenario. However, features from both the sensor-less AO method and imaging scenario should be integrated into each implementation of the framework to make the optimisation more practically relevant. In this chapter, a multiphoton fluorescence microscope was assumed and the framework was demonstrated with different sensor-less AO methods working in different imaging scenarios.

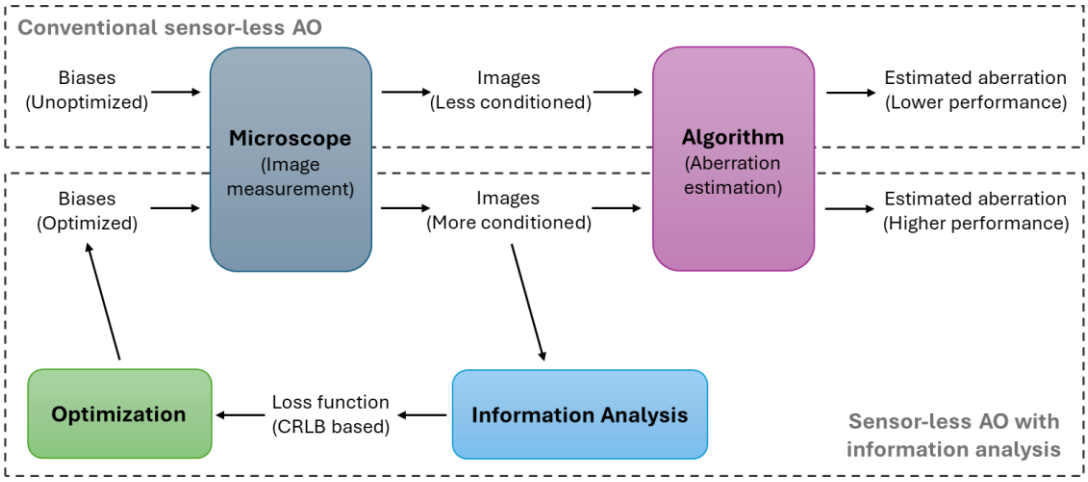


Figure 4-1: A diagram of sensor-less AO with the biases optimised through a framework based on information analysis comparing to the conventional sensor-less AO without optimisation of the biases.

4.1 A numerical model of sensor-less adaptive optics

A general model of sensor-less AO has been provided in Section 3.1. Based on that, the total aberration in the microscope was specifically modelled here by a linear combination of N Zernike modes, so the equation (3-1) became:

$$\phi = \sum_j^N (a_j + b_j) Z_j \tag{4-1}$$

where $a_j + b_j$ forms the coefficient of Zernike mode Z_j, with a_j from the aberration to be estimated ϕ_A and b_j from the bias aberration ϕ_B. For each measured image, the pixel dimensions were set as 128×128 and the pixel size was set as a half of the FWHM of the diffraction-limited multiphoton PSF on the focal plane. To calculate each measured image D, the 3-D convolution in the equation (2-38) was calculated through an approximation that sampled five axial planes with the central plane placed on the focal plane and the spacing between any two adjacent planes as the FWHM of the diffraction-limited multiphoton PSF along the optical axis; Before applying the equation (2-34), the detected image signal I_D was normalised such that its total intensity equalled to 1 when it was aberration free, which means that its total intensity equalled to its Strehl ratio when it was aberrated; The signal amplification ξ in the equation (2-34) was thus defined as the photon budget, which is the expected total signal photon number in D when there was no aberration; The background β was defined as a constant value across D; The distribution R was modelled as a Poisson distribution because photodetection in the multiphoton microscope is mainly affected by shot noise [35]. Moreover, the aberration estimator $\hat{\phi}_A$ was assumed to consist of M Zernike coefficient estimators $\hat{a}_j$ with $M \leq N$, so that the equation (3-3) changed to:

$$\hat{\phi}_A = \sum_j^M \hat{a}_j Z_j \tag{4-2}$$

4.2 Information analysis

As informed in Section 2.4.4, the Fisher information measures the amount of information carried by an observable random variable X about its contributing parameters Θ [74]. It provides a suitable way to analyse the information encoded in the images measured for sensor-less AO, because each pixel value of the images is an independent random variable that depends on ϕ_A. In the information analysis, the images as a whole was considered as X in which each element X_p corresponded to each pixel value. Θ was defined by a vector of multiple parameters including those representing not only ϕ_A but also other potentially unknown factors such as the object structures, the signal level, and the background level. The Fisher information in X_p could be calculated as a matrix based on the equation (2-53). Since X_p was Poisson-distributed, this equation changed into:

$$\left[\Im_{X_p}(\Theta)\right]_{l,m} = \frac{1}{\overline{X_p}}\left(\frac{\partial\overline{X_p}}{\partial\theta_l}\right)\left(\frac{\partial\overline{X_p}}{\partial\theta_m}\right) \tag{4-3}$$

where $\overline{X_p}$ represents the expected value of X_p. It was easy to obtain $\overline{X_p}$ using the numerical model, and the partial derivatives could be approximately calculated by a finite difference method. The step for finite difference calculation was chosen to ensure sufficient accuracy of the partial derivative approximations while keep the computational complexity at an acceptable level. For the partial derivatives with respect to parameters of ϕ_A, the step was set as 0.01. As X_p was independent from each other, the Fisher information in X could be finally calculated by:

$$[\Im_X(\Theta)]_{l,m} = \sum_p \left[\Im_{X_p}(\Theta)\right]_{l,m} \tag{4-4}$$

It is important to note that the independence here was in terms of the noise on each pixel

rather than the expected value on each pixel and the independence came from the pixel-by-pixel detection in the multiphoton microscope. To consider different imaging scenarios where coupling exists among detection of different pixels such as using a typical CCD camera, it might be no longer valid to assume such independence, and assuming such independence will cause an overestimation of Fisher information.

In a specific sensor-less AO method, the directly measured images can be further processed before serving as the input to the estimation algorithm. The processing of the images essentially selects only a part of information in the images to be effectively used in the subsequent calculation. In this case, the information analysis was preferably conducted with X defined by the processed result instead. Since random variables remain random variables after processed, the information in the new X could still be analysed by calculating the Fisher information [74]. For the Fisher information calculation, the distribution of each element X_p in the new X needed to be derived from that of each pixel value in the images. It is noted that such derivation could be difficult to accomplish when the processing was complex. Alternatively, the information in an intermediately processed result or the original images might be more easily analysed, although such analysis might provide less accurate guidance for the optimisation of the specific method.

Another aspect about a specific sensor-less AO method that can influence the information analysis is the usage of prior knowledge for the aberration estimation as the prior knowledge forms an extra information source besides the images. To take the

extra information into account, a simple way was to adjust Θ while still calculating the Fisher information in the same X for the analysis. When no prior knowledge was used, all the contributing factors of X should be deemed unknown therefore their parameters should all be included in Θ. As prior knowledge was introduced into the estimation algorithm either explicitly as additional inputs or implicitly as assumptions, it took effect by eliminating the uncertainty from some parameters so that these parameters could be removed from Θ [74]. When there were a large number of parameters in Θ, it could be difficult to calculate the whole Fisher information matrix. A compromise solution was to ignore some parameters that were not to be estimated, especially when these parameters contributed weakly to X, so a smaller matrix that equalled to a submatrix of the original one could be calculated instead, but at the cost of a potential drop in the accuracy of the information analysis.

4.3 Loss function

As informed in Section 2.4.5, the CRLB defines the lower bound on the variance of any unbiased estimator $\hat{\theta}_m$ of each parameter θ_m in Θ from X [74]. It indicates the accuracy of $\hat{\theta}_m$ when $\hat{\theta}_m$ makes full use of the information about θ_m in X. Thus, the CRLB of the aberration estimator $\hat{\phi}_A$ could be used to derive a desirable loss function for optimisation of the biases in sensor-less AO to achieve the most accurate estimation of ϕ_A. Following the numerical model, the CRLB of $\hat{a}_j$ could be calculated from $\Im_X(\Theta)$ based on the equation (2-55) as:

$$CRLB_{\hat{a}_j} = [\Im_X(\Theta)^{-1}]_{m_{a_j}, m_{a_j}} \leq \mathbb{V}(\hat{a}_j) \tag{4-5}$$

where m_{a_j} denotes the index of a_j in Θ. To combine all the $CRLB_{\hat{a}_j}$, a loss function was defined as:

$$\mathcal{L} = \sqrt{\sum_j^M CRLB_{\hat{a}_j}} \tag{4-6}$$

According to the equation (2-45), this loss function represented a lower bound on the mean error of $\hat{\phi}_A$ in the RMS.

4.4 Optimisation

The optimal biases could be searched by minimising the loss function. However, as a sensor-less AO method is generally expected to work well in not only one single case but a range of scenarios, the mean loss over the range of scenarios should be minimised instead in the optimisation for the best overall performance. The optimisation was implemented in a similar way to the training of a NN. First, all the optimisable parameters of the biases were randomly initialised. In each optimisation step, a batch of examples were randomly drawn from the range of scenarios following a probability distribution and the mean loss over the batch of example scenarios was calculated as the basis for an Adam gradient descent optimiser [151] to update jointly all the optimisable parameters of the biases. Multiple steps were normally required for the optimisation to achieve convergence. The details such as the definition of the scenario range, the optimisable parameters of the biases, the way of initialisation, the batch size, the probability distribution to draw examples, the learning rate of the optimiser and the number of optimisation steps could be different in different implementations and will be specified in later sections. Each implementation in this chapter was programmed

with TensorFlow in Python and run on a GPU (Nvidia GeForce RTX 3070).

4.5 Optimisation with a conventional algorithm

The framework was first demonstrated with sensor-less AO methods using a conventional algorithm. Here, the total aberration ϕ consisted of $N = 5$ Zernike modes including astigmatism (Z_5 and Z_6), coma (Z_7 and Z_8) and spherical (Z_{11}); the measured images were calculated by simulating either a 2-P or 3-P microscope with a point-like object at the focus, the photon budget $\xi = 200{,}000$ and the background $\beta = 100$. A representative sensor-less AO method was considered to estimate each a_j in ϕ_A separately. In this method, there were two images measured for each mode to be estimated with a pair of opposite biases of that mode and an extra image measured with a zero bias, so a total $2N + 1$ images were measured to estimate N modes. For the estimation of each mode, the total image intensity was calculated as an image quality metric from the two images measured for that mode and the common non-biased image respectively. A model-based algorithm was used to fit the three metric readings along with their corresponding biases to a parabola and obtain the coefficient estimator $\hat{a}_j$ from the value where the parabola peaked. The full details of this method can be found in [104].

Since each aberration mode was estimated from different inputs, the information analysis was conducted separately. For each mode, three corresponding metric readings were processed from the images to be the input of the estimation algorithm so they defined X. The distribution of each metric reading X_p could be easily derived from

that of each pixel value in the images as the sum of independent Poisson-distributed random variables still follows a Poisson distribution [152]. Thus, the Fisher information could be calculated similarly by the equation (4-3) and (4-4). The prior knowledge used in the estimation algorithm came from the additional inputs of the biases and the assumption of the parabolic model. This means that Θ only depended on the single a_j, the photon budget ξ and the background β. From $\Im_X(\Theta)$ calculated for each mode, a separate loss was calculated with $M = 1$ as only a single mode was estimated each time.

Correspondingly, an individual branch of optimisation was also implemented for each mode. In each branch, the sensor-less AO method was optimised for a range of scenarios where the RMS of ϕ_A was limited smaller than either 0.5 or 1.0 rad. Such a range was chosen to be within the range where the parabolic model can be deemed valid [95]. Due to the restriction on the choices of biases in the method, i.e. a pair of opposite biases in that mode for each mode plus a common zero bias, the optimisable parameter included only the bias amplitude for each mode. The bias amplitude was initialised by a value randomly drawn from a uniform distribution within 0.75 rad. The batch size was 32 and the ϕ_A in each example scenario was drawn from the surface of an n-sphere as in [153] with the radius of the n-sphere corresponding to an RMS value drawn from a uniform distribution in the limited RMS range of ϕ_A. There were 500 optimisation steps in total. The learning rate started from 0.05 and decreased by half every 100 steps.

To test whether the optimised biases resulted in the best performance of the method, a range of bias amplitudes were sampled with 0.1 rad as the step. For each sampled bias amplitude, 2000 example scenarios were drawn in the same way as in the optimisation for each mode to calculate the mean loss and the mean error of $\hat{a}_j$ by simulations based on the numerical model. Hence, a separate curve could be plotted for both the mean loss and the mean error with respect to the bias amplitude for comparison.

4.5.1 Two-photon microscope case

For a 2-P microscope, the test results were presented in Figure 4-2(a) when the RMS of ϕ_A was within 0.5 rad and Figure 4-2(b) when the RMS of ϕ_A was within 1.0 rad. In both figures, the curves of the mean loss followed the general trends of the curves of the mean error and the optimised bias amplitudes, which were marked by vertical dotted lines, also corresponded approximately to the lowest mean errors.

It is however noted that the mean error could become even lower than the mean loss though the latter represented a lower bound of the former. This is because $\hat{a}_j$ was manually limited according to the RMS range of ϕ_A, thereby leading to a further drop in the mean error. Interestingly, the curves of the two astigmatism modes (Z_5 in blue and Z_6 in red) always followed each other closely and the curves of the two coma modes (Z_7 in yellow and Z_8 in purple) also presented the same trait. This was expected as the two pairs of modes share similar shapes and only differ in orientations, which has been shown in Table 2-1. By comparing the two figures, it was observed that the optimised bias amplitudes increased when scenarios of larger ϕ_A were taken into

consideration. Such an observation agrees with the report in [98].

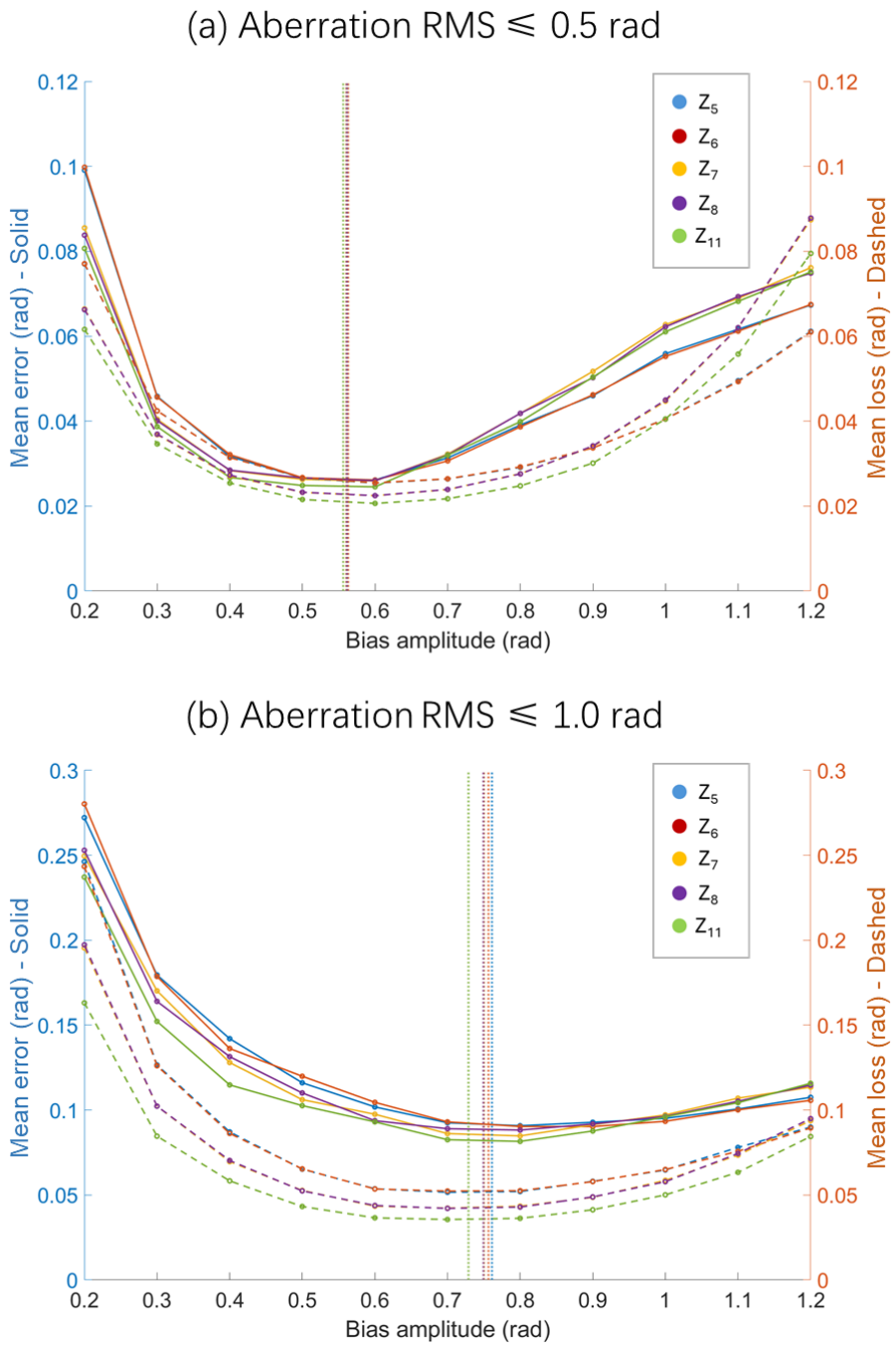


Figure 4-2: The 2-P comparison between the mean loss (dashed curves) and the mean error of each single mode coefficient estimator (solid curves) with respect to varied bias amplitudes when the RMS of ϕ_A was within 0.5 rad (a) and 1.0 rad (b). The optimised bias amplitude for each mode was marked with a vertical dotted line in both (a) and (b).

4.5.2 Three-photon microscope case

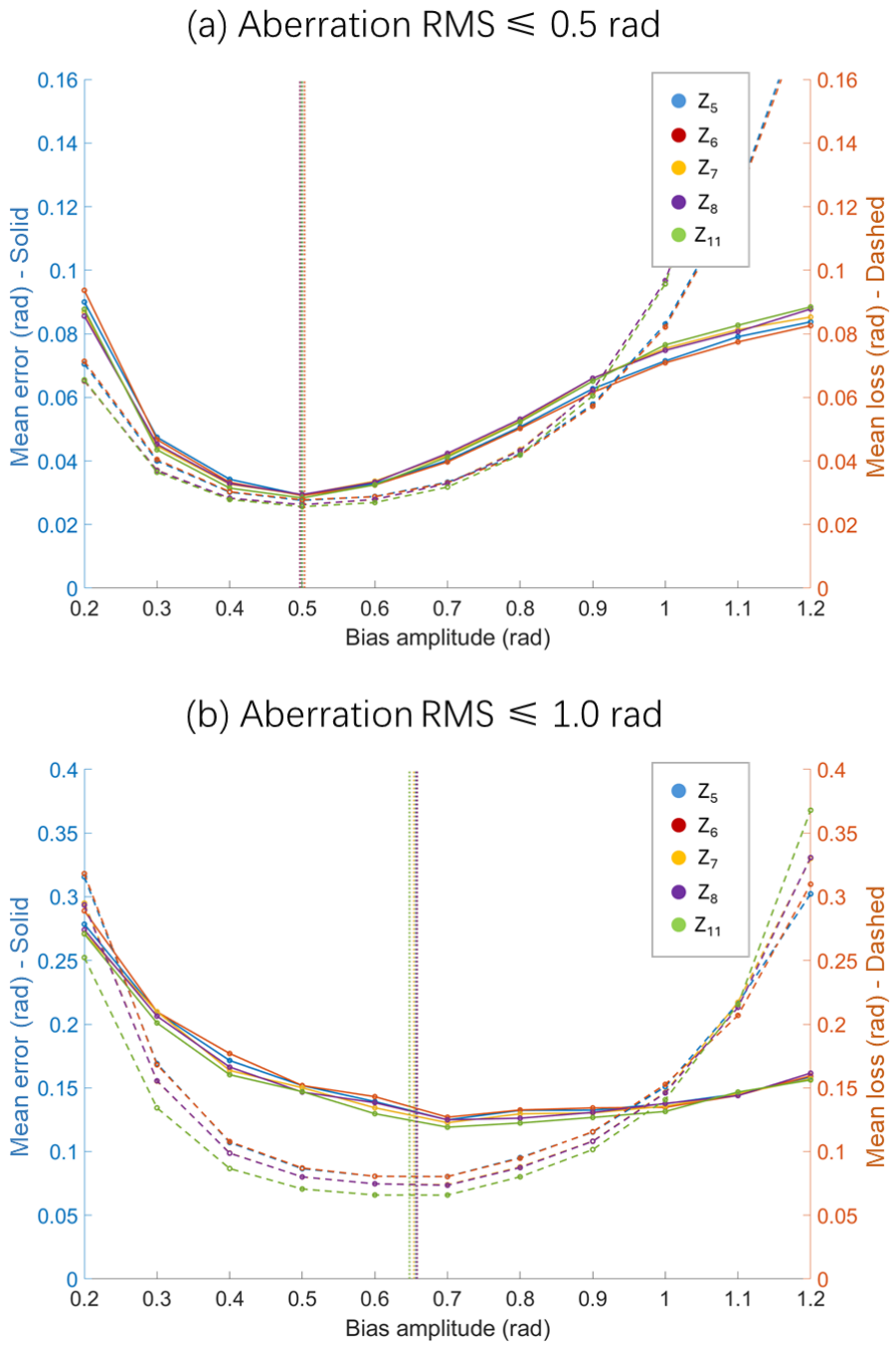


Figure 4-3: The 3-P comparison between the mean loss (dashed curves) and the mean error of each single mode coefficient estimator (solid curves) with respect to varied bias amplitudes when the RMS of ϕ_A was within 0.5 rad (a) and 1.0 rad (b). The optimised bias amplitude for each mode was marked with a vertical dotted line in both (a) and (b).

Figure 4-3 showed the test results for a 3-P microscope, which were generally similar to the 2-P results. It was observed that the mean loss curves started to deviate slightly

from the mean error curves when the RMS upper bound of ϕ_A increased from 0.5 to 1.0 rad, while the optimised bias amplitudes still met around the minima of the mean error curves, which were very flat as the bias amplitude changed between 0.6 and 1.0 rad. This is due to the higher order of the non-linearity of the 3-P microscope, which means that the total image intensity metric only responded effectively to the aberration changes in a narrower range [118]. The parabolic model was hence no longer valid when the RMS of ϕ_A increased to 1.0 rad thereby reducing the correspondence between the mean error and the mean loss. In addition, the optimised bias amplitudes became smaller when changing from 2-P to 3-P, which also reflected the reduced aberration range of the metric response.

4.6 Optimisation with a learning-based algorithm

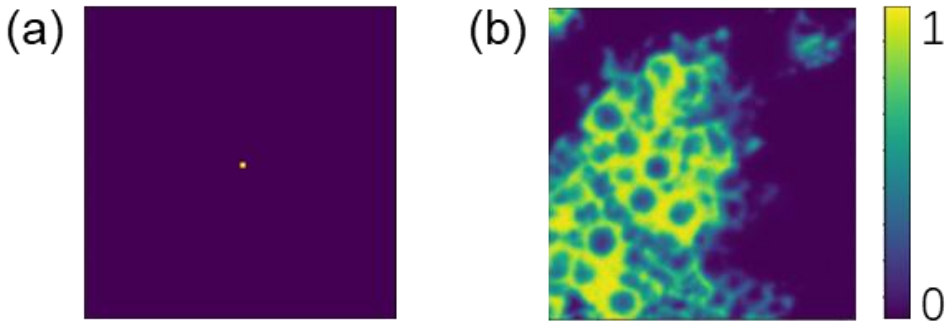


Figure 4-4: Image examples of the point-like object (a) and the extended volumetric object (b) when there was no aberration or noise. Each image was normalised to its own maximum.

Sensor-less methods using a learning-based algorithm were then employed for the demonstration of the framework. In this section, the total aberration ϕ consisted of $N = 7$ Zernike modes including astigmatism (Z_5 and Z_6), coma (Z_7 and Z_8), trefoil

(Z_9 and Z_{10}) and spherical (Z_{11}). The measured images were calculated by simulating a 2-P microscope with either a point-like object at the focus or an extended volumetric object, the photon budget $\xi = 50{,}000$ and the background $\beta = 1$. In the point-like object case, defocus (Z_4) was added exclusively to the biases as an extra mode. The extended volumetric object was generated by down-sampling an image stack of fluorescence labelled cytoplasmic membrane in a biological sample taken from a high-resolution 2-P microscope. The down-sampling helped to restore the sharp features in the sample structures that had been smoothed out by the imaging process. An image example of each object was provided in Figure 4-4 when there was no aberration or noise.

Specifically, a sensor-less AO method was designed to measure two biased images from which a ResNet-18 trained through supervised learning estimated all the seven a_j of ϕ_A together. The two biased images were normalised such that the maximum pixel value in the whole stack became 1 and the minimum became 0 before they were fed into the NN as the input. For the supervised learning, a synthetic dataset containing 20,000 examples for training and 2,000 examples for validation was generated by simulations based on the numerical model. In each example, there contained a stack of two biased images as the input labelled with its corresponding seven a_j of ϕ_A that was randomly drawn from the surface of an n-sphere as in [153] with the radius corresponding to an RMS smaller than 3.0 rad. During the training, the NN parameters were initialised by Glorot uniform initialisation [154]; the batch size was 32; the loss

was defined by the mean error of $\hat{\phi}_A$ in the RMS; the Adam gradient descent optimiser was used to train the NN for 30 epochs [151]; the learning rate started from 10^{-3} and decayed to 1/3 of its previous value when the decrease of the validation loss was less than 10^{-3} for two epochs in a row until the learning rate finally reached 10^{-5}; both L1 and L2 regularisation were applied to prevent overfitting. The training was programmed with TensorFlow in Python and run on a GPU (Nvidia GeForce RTX 3070).

In the information analysis, X was directly defined by the two biased images as the normalisation induced negligible change to the information. The training essentially imparted knowledge about the numerical model that was used for generating the training data to the NN. As only ϕ_A was changing in the numerical model throughout the data generation, Θ could be defined by a vector of $N = 7$ coefficients a_j. Based on $\Im_X(\Theta)$, the loss could be calculated with $M = N$. To implement the optimisation, the scenario range was defined by limiting the RMS of ϕ_A within 3.0 rad, which was the same as the range of the NN training data. As there was no extra restriction on choosing the biases from the method, the optimisable parameters were $2N = 14$ coefficients b_j representing the two biases when the extended volumetric object was imaged. In the point-like object case, the number of b_j was $2(N + 1) = 16$ as defocus was introduced as an extra bias mode. Each bias coefficient was initialised by a value randomly drawn from a uniform distribution within 1.5 rad. To prevent memory overflow, the batch size was set at 16. The ϕ_A in each example scenario was also

drawn from the same distribution as in the generation of each example for NN training. The biases were optimised for 1000 steps with the learning rate starting from 0.05 and reducing by half every 200 steps.

With more degrees of freedom, the biases could be chosen from a much broader space for the method. Also, every time when the biases changed, the NN needed to be trained all over again, which could induce great computational cost. It would be therefore impractical to follow the same way of demonstration as in the last section to sample different choices of the biases exhaustively from the whole space and test their corresponding performances to compare with the optimised biases. Instead, multiple conventional choices of the biases were selected for comparison. Similarly, for each different choice of the biases, 2000 example scenarios were drawn in the same way as in the optimisation to calculate the mean loss and the mean error of $\hat{\phi}_A$ in the RMS for comparison.

4.6.1 Point-like object case

When the point-like object was imaged, the optimised biases were shown in Figure 4-5. It could be seen that the optimised biases were dominated by defocus, which is in line with the result in [125]. In Figure 4-6, the optimised biases were compared with commonly used ± 1 rad single-mode biases in terms of the mean loss and the mean error. The standard derivations of the errors were depicted by the error bars. For reference, the initial ϕ_A was displayed in parallel with the errors. It was observed that the optimised biases resulted in the lowest mean error with the smallest standard

deviation and that the trends shown by the mean loss reflected well those of the mean error.

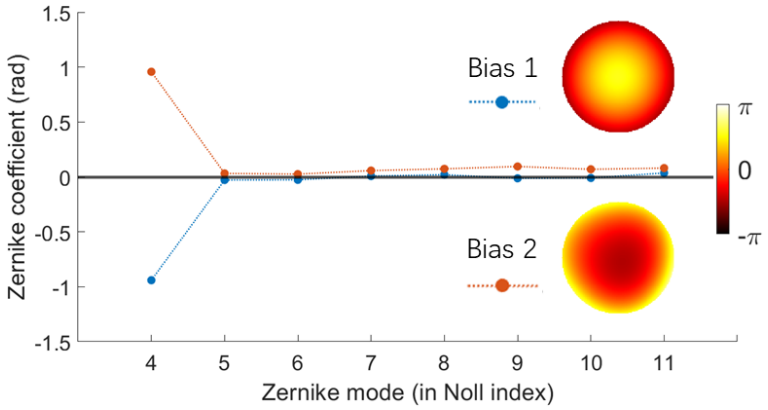


Figure 4-5: The optimised biases when the point-like object was imaged with their phases in rad shown as insets.

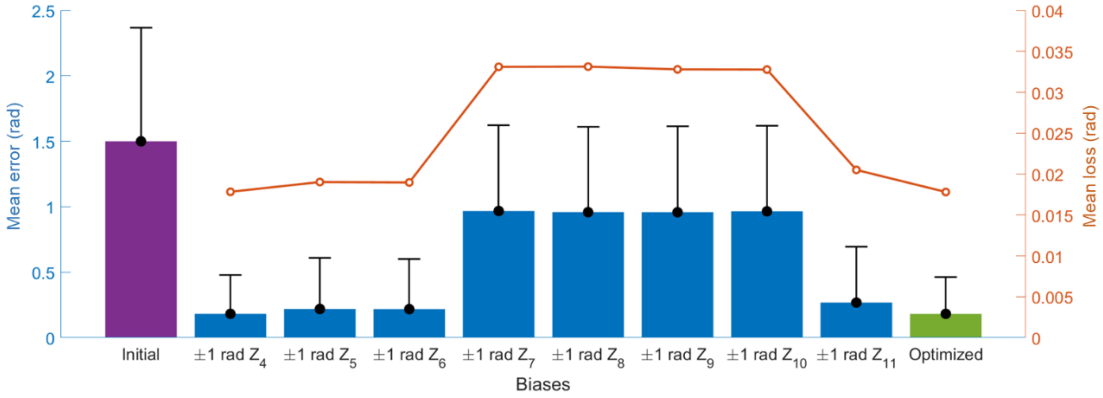


Figure 4-6: The comparison between the optimised biases and different ± 1 rad single-mode biases in terms of the mean loss and the mean error of $\hat{\phi}_A$ in the RMS when the point-like object was imaged. The error bars represented the standard deviations of the errors in the RMS. The initial ϕ_A was provided along with the errors as the reference.

4.6.2 Extended volumetric object case

Figure 4-7 presented the optimised biases when the extended volumetric object was imaged. It could be seen that the optimised biases were mostly in the two astigmatism modes, which agrees with the results in [121] and [123]. Similarly, the optimised biases were compared with commonly used ± 1 rad single-mode biases in Figure 4-8 by

calculating the mean loss and the mean error respectively. As in Figure 4-6, the error bars represented the standard deviation of the errors and the initial ϕ_A was presented for reference. Still, the optimised biases outperformed all the other biases and the mean loss kept consistent with the mean error.

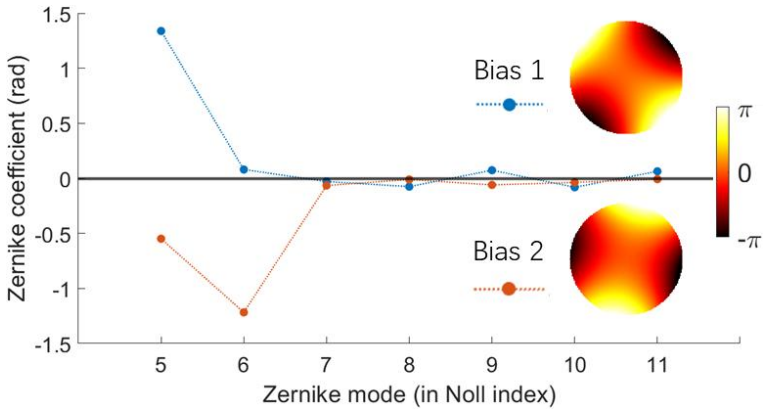


Figure 4-7: The optimised biases when the extended volumetric object was imaged with their phases in rad shown as insets.

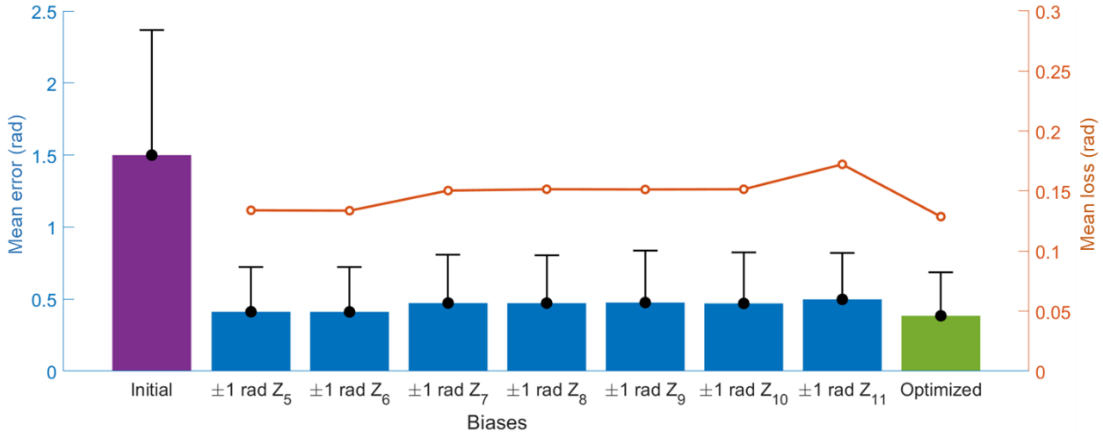


Figure 4-8: The comparison between the optimised biases and different ± 1 rad single-mode biases in terms of the mean loss and the mean error of $\hat{\phi}_A$ in the RMS when the extended volumetric object was imaged. The error bars represented the standard deviations of the errors in the RMS. The initial ϕ_A was provided along with the errors as the reference.

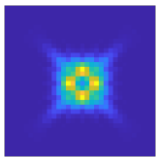
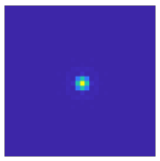
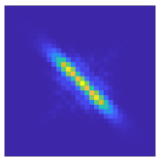
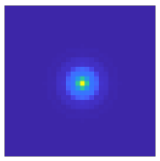
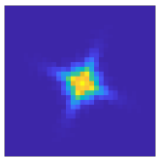
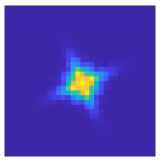
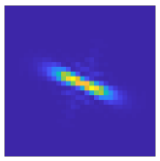
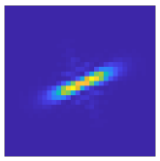
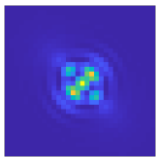
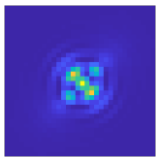
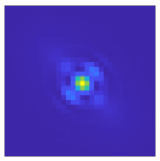
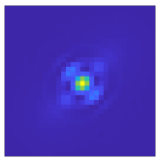
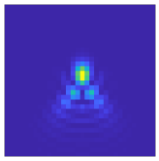
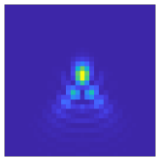
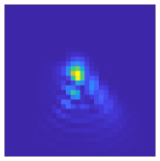
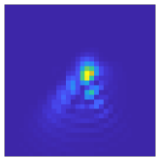
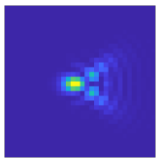
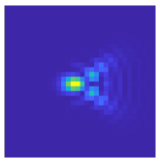
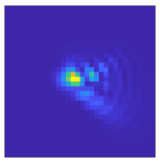
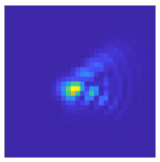
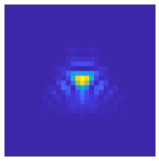
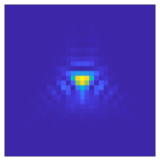
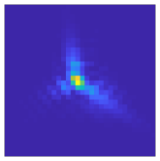
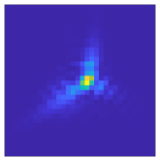
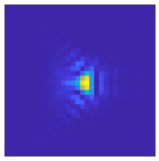
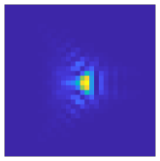
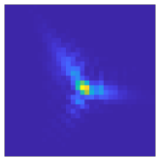
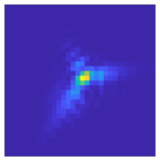
An interesting observation was that the single-mode biases of coma (Z_7 and Z_8) and trefoil (Z_9 and Z_{10}) performed much differently imaging the extended volumetric

object compared to imaging the point-like object. This is due to the fact that the image of the point-like object was roughly the PSF on the focal plane, while the image of the extended volumetric object had contributions from PSFs on both the focal plane and the defocused planes. For two different ϕ_A, if they were opposite in even radial order modes and the same in odd radial order modes, they would induce identical PSFs on the focal plane but different PSFs on the defocused planes when no bias was applied. Hence, it was ambiguous to differentiate such two ϕ_A from the image of the point-like object, while the ambiguity did not exist for the image of the extended volumetric object. When an even radial order bias mode was applied, the even radial order modes would be no longer opposite, so the two ϕ_A would become differentiable from the image of the point-like object as well. In contrast, when an odd radial order bias mode was applied, the condition for the ambiguity would still hold true, so the ambiguity would remain for the image of the point-like object.

An illustration of this phenomenon is provided in Table 4-1. In the illustration, the PSFs on the focal plane and on the +1 rad defocused plane are presented for +1 rad different bias modes with either +1 or -1 rad astigmatism (Z_5) as ϕ_A. As astigmatism (Z_5) is an even radial order mode, the two possible ϕ_A induce the same PSF on the focal plane, which shows the ambiguity. When the even radial order bias modes (Z_5, Z_6 and Z_{11}) is applied, the PSFs on the focal plane resulted from the two possible ϕ_A become different from each other so the ambiguity is avoided. However, when the odd radial order bias modes (Z_7, Z_8 Z_9 and Z_{10}) is applied, the PSFs on the focal plane

resulted from the two possible ϕ_A stay the same so the ambiguity remains. For the PSFs on the defocused plane resulting from the two possible ϕ_A, they are always different so no ambiguity is involved.

Table 4-1. A comparison of PSFs for different bias modes

Bias mode (+1 rad)		Focal plane PSF (+1 rad Z_5 ϕ_A)	Focal plane PSF (-1 rad Z_5 ϕ_A)	+1 rad defocused plane PSF (+1 rad Z_5 ϕ_A)	+1 rad defocused plane PSF (-1 rad Z_5 ϕ_A)
Even radial order	Z_5				
	Z_6				
	Z_{11}				
Odd radial order	Z_7				
	Z_8				
	Z_9				
	Z_{10}				

4.7 Optimisation with more degrees of freedom

Using more optimisable parameters could increase the benefit of the optimisation, which would in turn influence the sensor-less AO performance resulting from use of the optimised biases. As the number of the optimisable parameters depended on the degrees of freedom in the whole set of biases, the demonstration in the last section was extended to more than two biases. The new demonstration could also be considered as an optimisation of the number of biases. Here, most implementation details were still maintained the same as in the last section. The extended volumetric object images were used. To make sure that the sensor-less AO performance would only change with the optimisation result, a fixed total photon budget and fixed total background were maintained across the whole stack of the biased images. This is equivalent to having a fixed amount of sample exposure in a fixed amount of time for the measurement of the whole set of biased images. Assuming that each biased image had an equal share of the total photon budget and total background, the photon budget of each biased image ξ could be derived by the total photon budget divided by the bias number and the background of each biased image β could be derived by the total background divided by the bias number. Thus, the more biases were used, the smaller fraction of the total photon budget and total background was assigned to each biased image. The total photon budget was set at 1,000,000 / 320,000 / 100,000 corresponding to high / medium / low photon budget and the total background was kept at 100. To avoid memory overflow during the optimisation, the batch size was reduced to 8 when the bias number

was larger than 8.

For the purpose of testing the optimisation results with more degrees of freedom and their corresponding performance, the bias number was increased from 2 to $2N = 14$ and the optimisation was performed for each bias number in this range. Based on the optimised biases obtained from each optimisation, 2000 example scenarios were drawn as in the last section to calculate the mean loss, so the mean loss could be plotted against the bias number. Since it was computationally expensive to train the NN and calculate the mean error of $\hat{\phi}_A$ in the 2000 example scenarios for each bias number, three representative cases were selected to calculate the mean error instead. These three cases included (1) 2 biases, which is the minimum required for the sensor-less AO method [116]; (2) $N + 1 = 8$ biases, which is the minimum required to span the aberration space of N coefficients [99,103]; (3) $2N = 14$ biases, which provides symmetrical biasing for each mode [104]. Hence, these three cases formed a logical subset of all possible bias numbers that were sufficient to show any trends.

4.7.1 High photon budget case

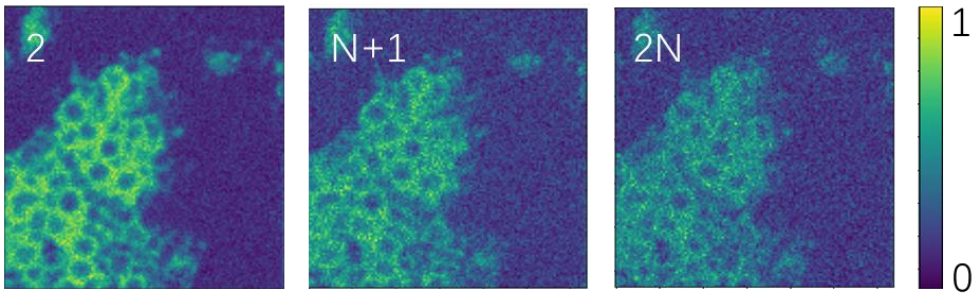


Figure 4-9: Image examples obtained with high photon budget when there was no aberration and the bias number was 2, $N + 1$ and $2N$ respectively. Each image was normalised to its own maximum.

With high photon budget, three diffraction-limited image examples were obtained corresponding to the three representative bias numbers (2, $N+1$ and $2N$) and presented in Figure 4-9 to visualise the typical image SNR in these cases. Clearly, the image SNR dropped with the increase of the bias number as the total signal was divided into more images.

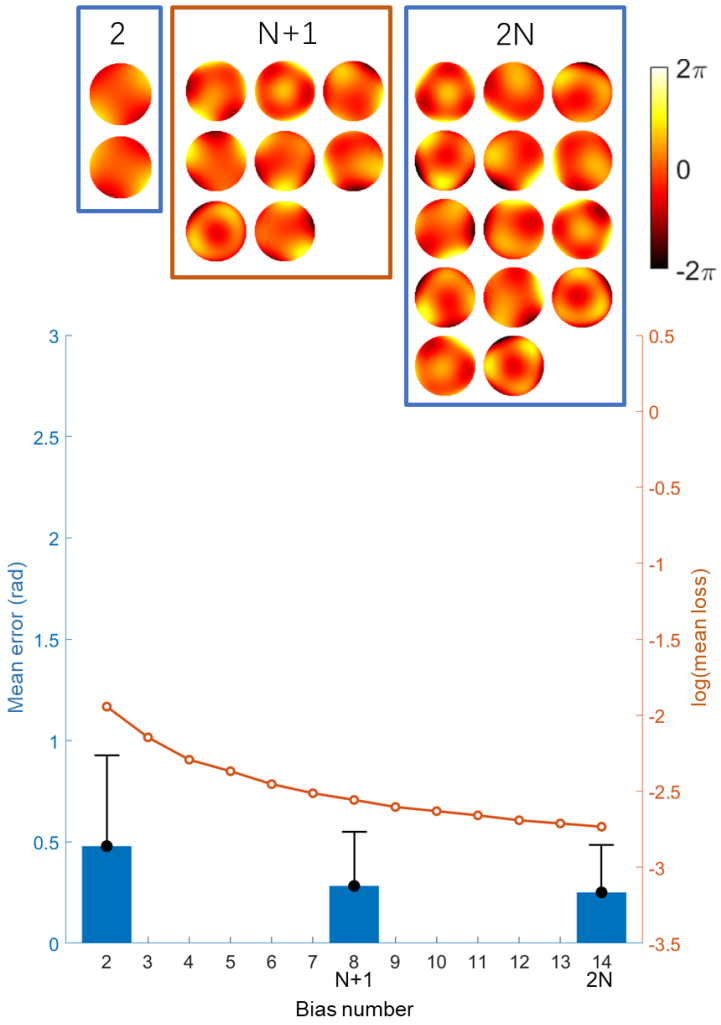


Figure 4-10: The logarithm of the mean loss and the mean error of $\hat{\phi}_A$ in the RMS against the bias number when the photon budget was high. The error bars represented the standard deviations of the errors in the RMS. The insets displayed the phases of the optimised biases in rad when the bias number was 2, $N+1$ and $2N$ respectively.

Figure 4-10 showed the plot of the mean loss (on a logarithm scale) changing with

the bias number and the mean errors calculated when the bias number was 2, $N+1$ and $2N$ respectively with the error bars representing the standard deviations of the errors. It was observed that as the bias number grew, the mean loss dropped consistently but at a reducing rate, which matched with the trend shown in the three mean errors as well as their corresponding standard deviations. This suggested that the performance was further improved as the biases were brought closer to the optimal state by the optimisation with more degrees of freedom, and that such improvement started to saturate as the degree of freedom kept growing. Potential reasons for the saturation of improvement included limitations from the aberration representation, the estimation algorithm and the optimisation implementation. In addition, the phases of the optimised biases when there were 2, $N+1$ and $2N$ biases were provided as insets of Figure 4-10.

4.7.2 Medium photon budget case

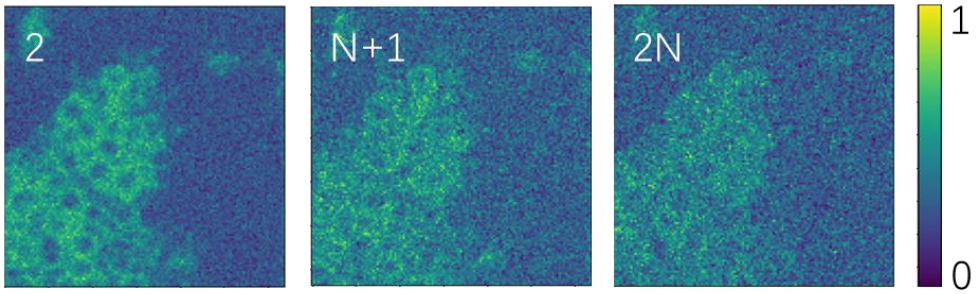


Figure 4-11: Image examples obtained with medium photon budget when there was no aberration and the bias number was 2, $N+1$ and $2N$ respectively. Each image was normalised to its own maximum.

Figure 4-11 showed the three diffraction-limited image examples obtained with medium photon budget when there were 2, $N+1$ and $2N$ biases. Compared to the

images in Figure 4-9, these images were apparently of lower SNR.

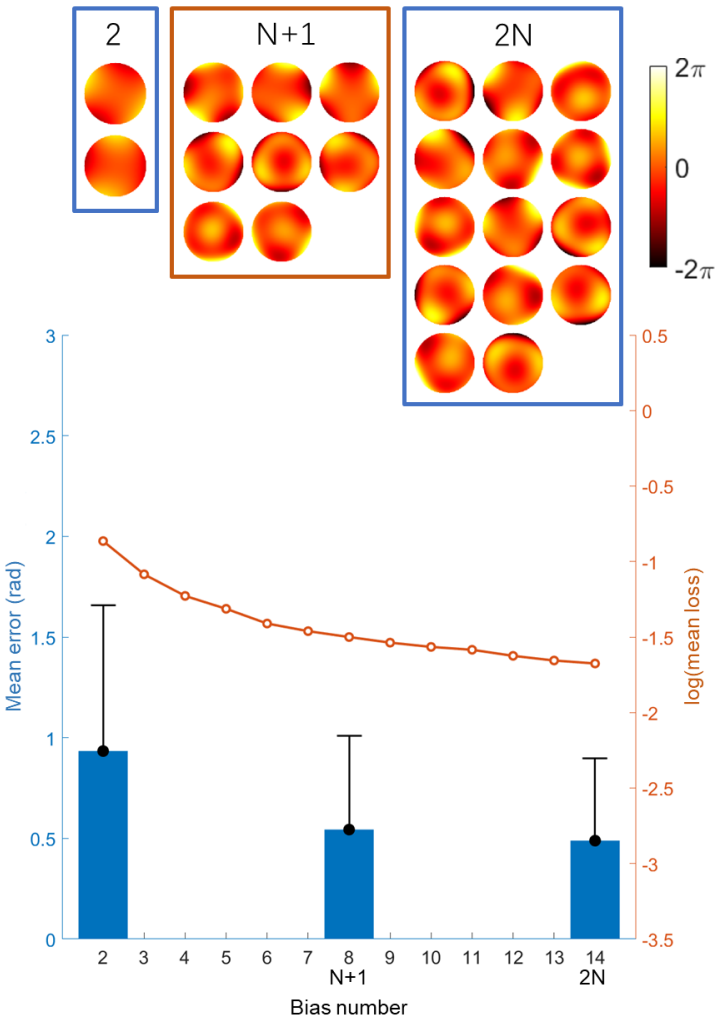


Figure 4-12: The logarithm of the mean loss and the mean error of $\hat{\phi}_A$ in the RMS against the bias number when the photon budget was medium. The error bars represented the standard deviations of the errors in the RMS. The insets displayed the phases of the optimised biases in rad when the bias number was 2, $N+1$ and $2N$ respectively.

As a result, overall poorer performance could be seen from Figure 4-12 showing higher mean losses and higher mean errors with larger standard deviations compared to the results in Figure 4-10, while similar declines could still be observed in the mean loss, the mean error and the standard deviation as the bias number increased. This means that similar improvement was brought to the performance by the optimisation with more

degrees of freedom regardless of the change in photon budget. Again, the phases of optimised biases of the three representative bias numbers (2, $N+1$ and $2N$) were displayed as insets of Figure 4-12.

4.7.3 Low photon budget case

When the photon budget was low, it could be seen from Figure 4-13 that the three diffraction-limited image examples became so noisy that the object structures were barely distinguishable, especially when the total signal was separated into more images. Accordingly, the mean errors shown in Figure 4-14 got close to the mean size of the initial aberration (1.5 rad), which indicated very weak performance. Such performance was also well predicted by the highest mean losses in Figure 4-14 compared to those in Figure 4-10 and Figure 4-12. Despite this, the improvement in performance could still be observed as the bias number increased thereby allowing better optimised biases, which demonstrated the strong robustness of the optimisation to noise. As in Figure 4-10 and Figure 4-12, the phases of the optimised biases of the three representative bias numbers (2, $N+1$ and $2N$) could be found in the insets of Figure 4-14.

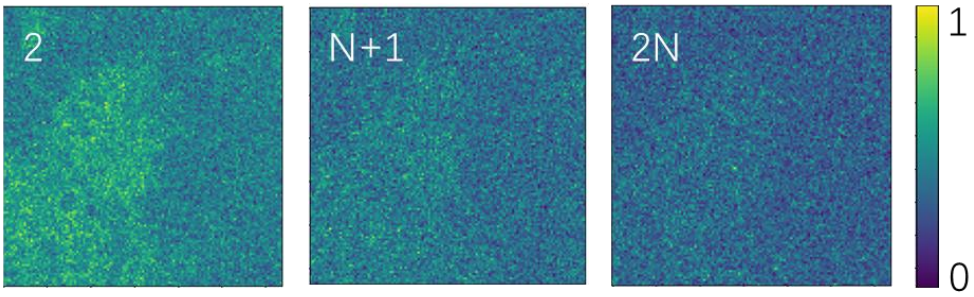


Figure 4-13: Image examples obtained with low photon budget when there was no aberration and the bias number was 2, $N+1$ and $2N$ respectively. Each image was normalised to its own maximum.

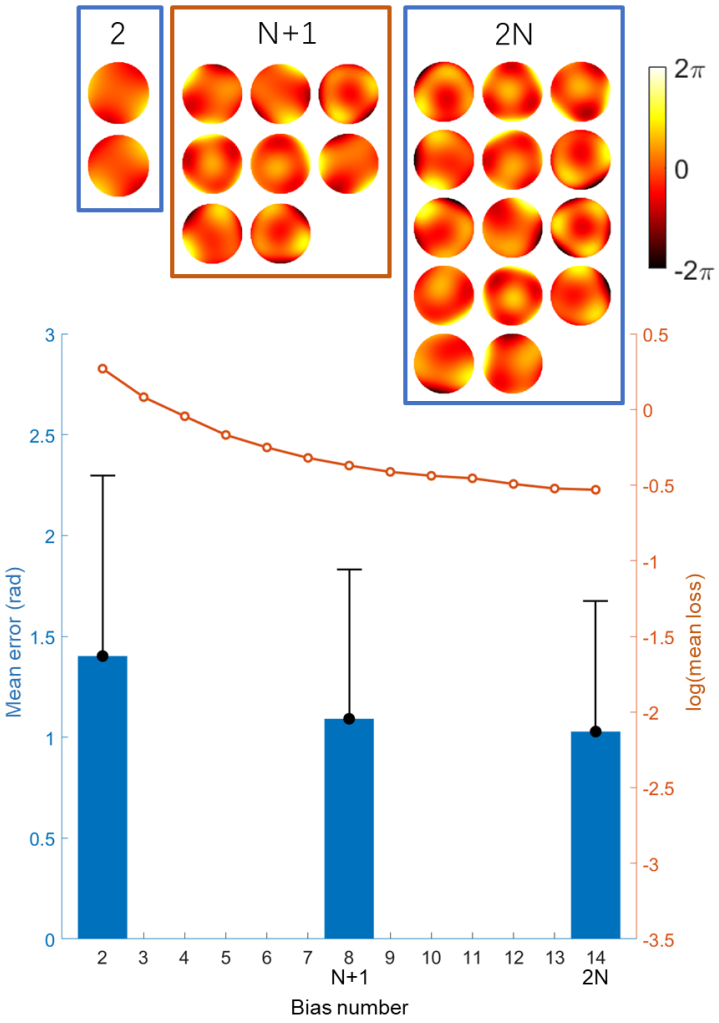


Figure 4-14: The logarithm of the mean loss and the mean error of $\hat{\phi}_A$ in the RMS against the bias number when the photon budget was low. The error bars represented the standard deviations of the errors in the RMS. The insets displayed the phases of the optimised biases in rad when the bias number was 2, $N+1$ and $2N$ respectively.

4.7.4 Optimised biases characteristics

At last, the optimised biases shown in the insets of Figure 4-10, Figure 4-12 and Figure 4-14 were collectively assessed to reveal their characteristics in different aspects. The RMS phase of these optimised biases was calculated and displayed in Figure 4-15. It was observed that the optimised biases tended to be of larger RMS phase as the bias number increased.

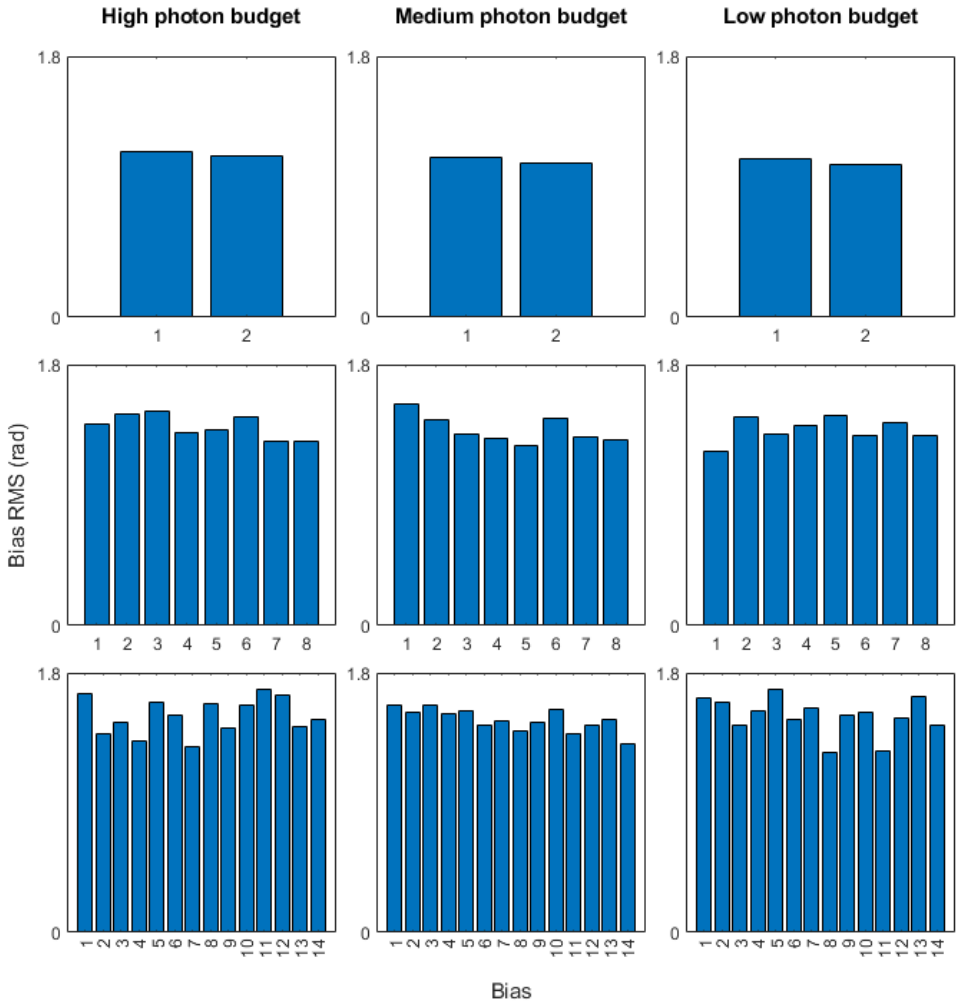


Figure 4-15: The RMS phase of the optimised biases obtained with high / medium / low photon budget when the bias number was 2 (the first row) / $N+1$ (the second row) / $2N$ (the third row).

To assess the composition of the $N = 7$ Zernike modes in the optimised biases, modes of the same type were grouped together such as Z_5 and Z_6 for astigmatism, Z_7 and Z_8 for coma, Z_9 and Z_{10} for trefoil, and Z_{11} on its own for spherical. This is because modes in the same group only differ in orientation and the optimisation generated an arbitrary combination of them. The percentage of each group of modes in the optimised biases was calculated in terms of its contribution to the mean square phase (equivalently the phase variance) by:

$$\%\ of\ each\ group = \frac{\sum_{j\ in\ the\ group} {b_j}^2}{{RMS_{\phi_B}}^2} \tag{4-7}$$

where RMS_{ϕ_B} denotes the RMS phase of each bias ϕ_B and b_j is the coefficient of Z_j in ϕ_B.

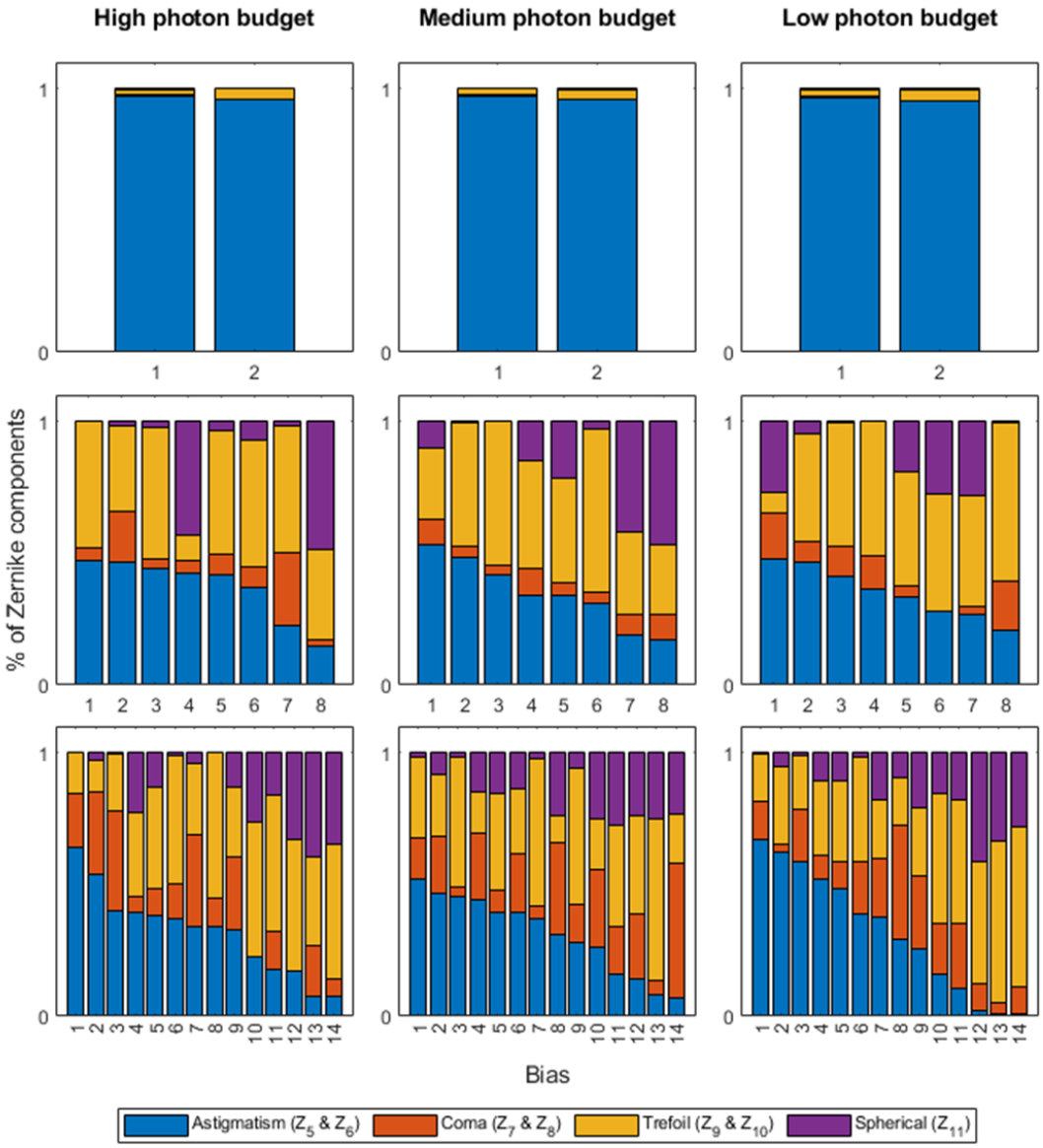


Figure 4-16: The Zernike mode composition of the optimised biases obtained with high / medium / low photon budget when the bias number was 2 (the first row) / $N + 1$ (the second row) / $2N$ (the third row).

In Figure 4-16, it could be seen that astigmatism was the dominant component in the optimised biases when the bias number was 2. When the bias number grew to $N +$

1, the percentage of trefoil became close to that of astigmatism and spherical also became a large contributor in some of the optimised biases, while there was only a small increase in the percentage of coma. For $2N$ biases, no dominant component could be observed and each component featured significant contribution.

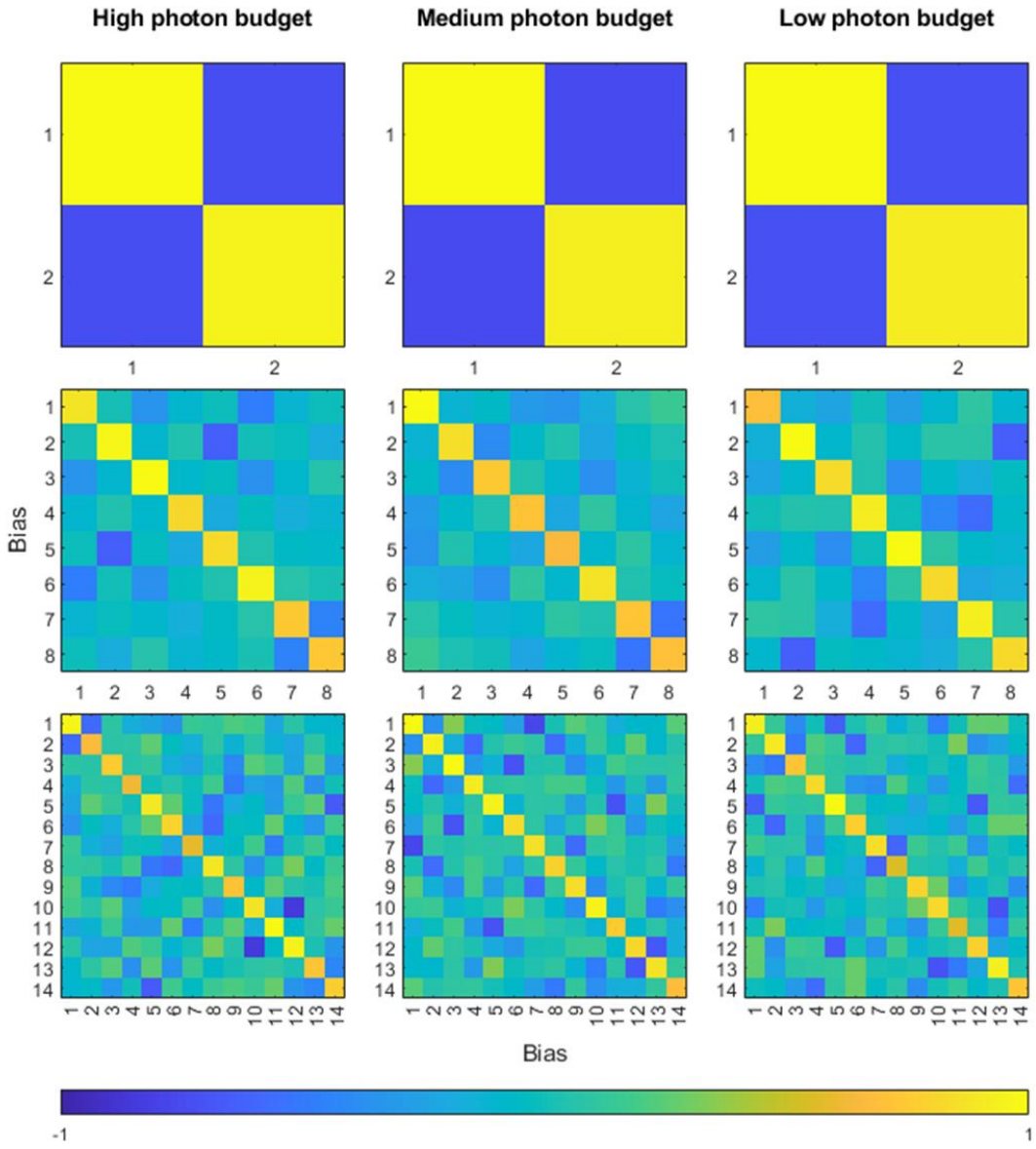


Figure 4-17: The inner product matrices of the optimised biases obtained with high / medium / low photon budget when the bias number was 2 (the first row) / $N+1$ (the second row) / $2N$ (the third row). The inner products were normalised to the range from -1 to 1.

Moreover, the inner product between each pair of biases in the optimised biases was calculated to show how the optimised biases were distributed in the aberration space. The inner product between two biases ϕ_B and ϕ'_B was calculated by:

$$\langle \phi_B, \phi'_B \rangle = \sum_j^N b_j b'_j \quad (4\text{-}8)$$

where b_j and b'_j are the Zernike coefficients of ϕ_B and ϕ'_B respectively. Figure 4-17 presented the matrices of inner products normalised to the range between -1 and 1. For 2 biases, the opposite diagonal and off-diagonal elements in the matrices indicated that the optimised biases were equal in size but opposite in direction. For $N+1$ biases, most off-diagonal matrix elements were close to zero which means that the optimised biases were equally spread out thereby providing good spanning of the aberration space. For $2N$ biases, a significant off-diagonal negative element was observed in most rows of the matrices while other off-diagonal elements were close to zero. This suggested that the optimised biases roughly consisted of pairs of opposite biases with each pair orthogonal to the other pairs, which again resulted in a close to even distribution in the aberration space.

4.8 Summary and discussion

In this chapter, a framework was introduced to analyse the information encoded for sensor-less AO by calculating the Fisher information with a corresponding numerical model and further derive a CRLB based loss function to optimise the biases for better information encoding and thus higher sensor-less AO performance. The framework was demonstrated with different sensor-less AO methods of both a conventional algorithm

and a learning-based algorithm working under different imaging scenarios in a multiphoton fluorescence microscope such as 2-P and 3-P imaging of a point-like or extended volumetric object with three different levels of photon budget. The results showed that the optimised biases resulted in the lowest mean error of the aberration estimator and the mean loss well predicted the mean error. With more degrees of freedom allowed in the biases, the optimisation was able to steer the biases closer to the optimal state and bring further improvement in the sensor-less AO performance though the further improvement became smaller as the degrees of freedom increased. To consider such results with respect to optimisation of the number of biases, it was demonstrated that a larger number of biases would lead to better performance when the total photon budget was controlled the same. Also, the characteristics of the optimised biases were assessed in the demonstration.

Admittedly, the framework was limited in several aspects. An image formation model was required for the information analysis, which hindered the application of the framework when such a model was not available. Fortunately, the image formation process is well understood in most sensor-less AO tasks, so the framework can still be able to widely adopted. Another problem lies in the fact that the CRLB based loss was based on the assumption that the aberration estimator was unbiased while this is not necessarily true in many sensor-less AO methods. To address this problem, features leading to a biased aberration estimator such as the processing to the measured images and the usage of prior knowledge were considered in the information analysis to adjust

the calculation of the Fisher information so that the derived loss was tailored to each different method. When it is difficult to fully model the features, the information analysis can usually apply approximations without affecting too much the final result. A similar strategy may also be suitable for the case when the whole Fisher information matrix is difficult to calculate. Moreover, since the loss corresponded to a lower bound of the estimation error, it might deviate from the actual value when the estimator was not fully efficient or based on inaccurate assumptions. However, as long as the general trend of the loss matches with that of the actual error which is mostly the case, the optimisation can still work effectively. It is also noted that the Fisher information was computed locally around a point in the parameter space so that the information analysis was less sensitive to the cases where very different parameters resulting in similar images. To avoid this issue affecting the optimisation, the parameters can be constrained in the optimisation to rule out such cases or an additional non-local term can be designed and added to the loss function. On top of this, the optimisation could fail to converge to the global optimum if there were multiple local optimums. Feasible solutions include applying extra constraints to the optimisation, introducing regularisation terms to the loss function and using a more robust optimiser.

Overall, the framework provided a highly versatile information-guided approach to design the biasing scheme of different sensor-less AO methods for different imaging scenarios. Although the demonstration was limited to a multiphoton fluorescence microscope here, the framework can also be applied to other kinds of imaging systems

with different imaging models used in implementation. Compared to conventional bias design based on empirical experience, the framework optimises the biases objectively to achieve the best sensor-less AO performance. This can be very useful when direct comparison in resultant performance among different choices of biases are too computationally expensive to be practical, such as for most state-of-the-art methods of a DL based algorithm, while the framework enables a reliable prediction of the performance by the loss function derived from the information analysis and an automatic optimisation of the biases. Hence, this framework is expected to help these methods release their full potential and realise more efficient sensor-less AO.

Chapter 5

Single image sensor-less adaptive optics for multiphoton fluorescence microscopy

An important reason why sensor-less AO can be time-consuming is the need for sequential measurement of multiple images. Therefore, sensor-less AO can work much faster if a way can be found to parallelise the sequential measurement. In many realistic imaging scenarios of a multiphoton fluorescence microscope, the object has some sparsity in its structures, which means that most signal is concentrated within small areas of the image while the remaining areas contain mostly background and noise. In this case, considerable proportions of each image measured for sensor-less AO do not contribute usefully but do cause disturbance to the aberration estimation. Based on this fact, a new sensor-less AO method for multiphoton fluorescence microscopy was proposed in this chapter. In this method, the excitation light was focused into an array of focal spots for imaging with each focal spot containing a different bias, so the image became the superposition of laterally offset biased sub-images. As a result, the parallel image measurement was realised by making use of the originally wasted area to concentrate the information that used to be separated in multiple images into one single image. The much more concentrated information was extracted by a DL based algorithm where a pre-trained NN estimated the aberration directly from the

superimposed images. In this way, the method could finish in milliseconds with even a single image measurement. The demonstration of the method was through experiments on a 2-P microscope equipped with a SLM imaging under different scenarios. The method could also be applied to the recent work of using pseudo-PSF in [121] for better performance.

5.1 Imaging with a focal spot array

Single focal spot imaging in a conventional sensor-less AO method for multiphoton fluorescence microscopy has been described in Section 3.1. Based on that, when the excitation light is focused instead into an array of T focal spots for imaging, the light intensity of the focal spot array can be written as:

$$I_e = \sum_t^T I_e^t \tag{5-1}$$

where I_e^t represents the light intensity of the t-th focal spot in the array. According to the equation (2-38), the image signal detected by scanning the focal spot array can be described as:

$$I_D = O \otimes_{3-D} (\sum_t^T I_e^t)^\kappa \tag{5-2}$$

with the same notation system as in Section 3.1. If the focal spots in the array are sufficiently separated with negligible overlapping, namely $I_e^t I_e^{t'} \approx 0$ when $t \neq t'$, the equation (5-2) becomes:

$$I_D = \sum_t^T O \otimes_{3-D} (I_e^t)^\kappa = \sum_t^T I_D^t \tag{5-3}$$

where I_D^t represents the image signal detected by individually scanning the t-th focal spot in the array. The image formed by the focal spot array is hence equivalent to the

summation of the separate images formed by each focal spot in the array when the focal spots are displaced from one another. Based on the equation (2-39), the actual image detected with a limited SNR and a potential background can be described as:

$$D = R[\xi(\sum_t^T I_D^t) + \beta] \tag{5-4}$$

Here, assuming that the total excitation intensity is equally divided into each focal spot, the total signal in D is about $1/T^{\kappa-1}$ of its counterpart when imaging with a single focal spot and the same signal amplification ξ. For a 2-P microscope, $\kappa = 2$ so the total signal in D drops to $1/T$ of its original value when changing from a single focal spot to an array of T focal spots. To maintain the signal level as the spot number changes, the total excitation intensity can be increased to $\sqrt{T}$ times, which equals to ξ times T as the 2-P fluorescence signal is proportional to the square of the excitation intensity.

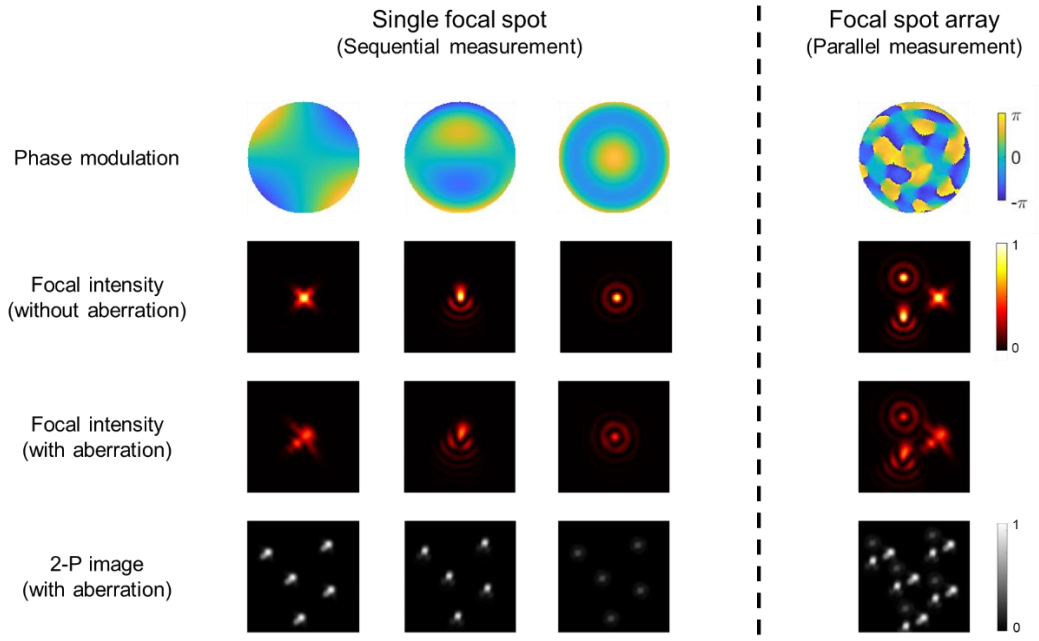


Figure 5-1: The comparison between the sequential image measurement with a single focal spot modulated with a different bias at each time and the

parallel image measurement with a focal spot array in which each spot contains a different bias. The phase modulations are in rad and all the other graphs were normalised to the overall maximum of graphs in the same type.

As shown in Figure 5-1, if the excitation light is modulated to form such a focal spot array that each focal spot contains a different bias, the image will be the superposition of laterally offset biased images. In this way, the biased images that in a conventional sensor-less AO method would have been measured sequentially are measured in parallel.

5.2 Holography and Gerchberg-Saxton algorithm

The focal spot array can be generated by holography. Holography is a widely adopted technique for 3-D imaging [16]. In principle, it allows light field to be recorded as a hologram and reconstructed afterwards. Traditionally, a hologram is generated by capturing an interference pattern on a photosensitive material. The interference pattern results from the light from the object interfering with another coherent light as the reference. Later, the hologram is illuminated by the reference light and the object light can be recreated by diffraction.

As an alternative, a hologram can also be generated by a computer algorithm to show a virtual object [155]. The computer-generated hologram can be not only printed onto a mask or film, but also directly displayed on a dynamic holographic display like a SLM. For reconstruction, a coherent light source is modulated by the hologram and diffracted to form the holographic image. Normally, a pure phase or pure amplitude hologram is generated for easier implementation of modulation. One of the most

prevalent algorithms used to produce computer-generated holograms is the Gerchberg-Saxton (GS) algorithm [156]. It is an iterative algorithm computing a phase-only hologram based on the light intensity on the hologram plane and the diffraction plane, between which the light propagation is generally defined by Fourier transform [157-159].

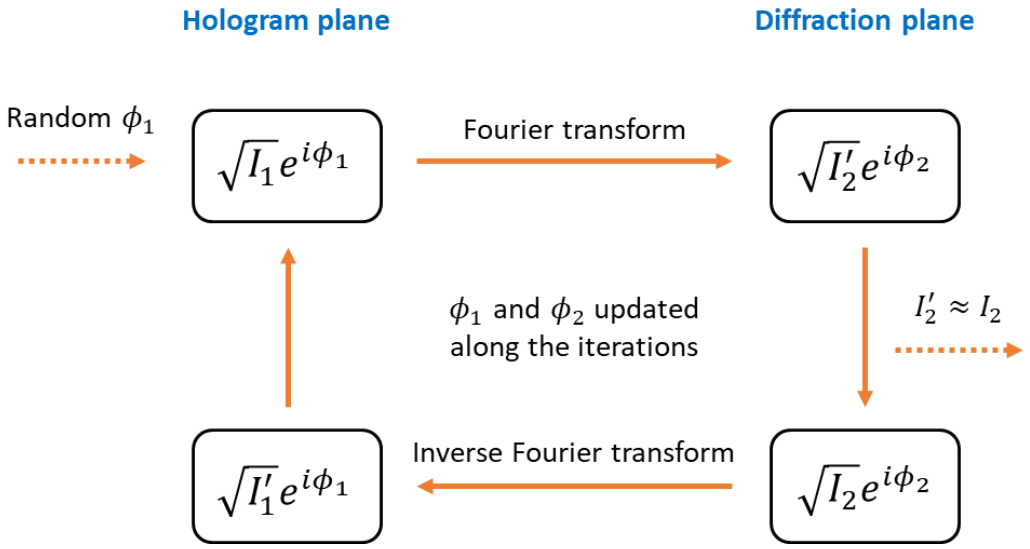


Figure 5-2: An implementation of the GS algorithm. I_1, the correct light intensity on the hologram plane; I_2, the target light intensity on the diffraction plane; I_1', the calculated light intensity on the hologram plane, replaced by I_1 in each iteration; I_2', the calculated light intensity on the diffraction plane, replaced by I_2 in each iteration until the error criterion is met which ends the iterations; ϕ_1, the phase on the hologram plane, randomly initialised and updated in each iteration; ϕ_2, the phase on the diffraction plane, updated in each iteration.

An implementation of the GS algorithm is illustrated in Figure 5-2. At the start of the algorithm, the phase ϕ_1 on the hologram plane is randomly initialised. The light field on the hologram plane can be built by $\sqrt{I_1}e^{i\phi_1}$ with I_1 the light intensity on the hologram plane. Then, the light field of the diffraction plane can be calculated with Fourier transform to be $\sqrt{I_2'}e^{i\phi_2}$. If the error between I_2' and the target light intensity

I_2 on the diffraction plane fails to meet the chosen criterion, the light field of the diffraction plane is replaced to be $\sqrt{I_2}e^{i\phi_2}$. With inverse Fourier transform, a new light field on the hologram plane can be calculated and the new phase is used to update ϕ_1. The aforementioned procedures are iterated until the error criterion is satisfied and ϕ_1 will gradually converge to the correct hologram along the iterations.

5.3 Workflow

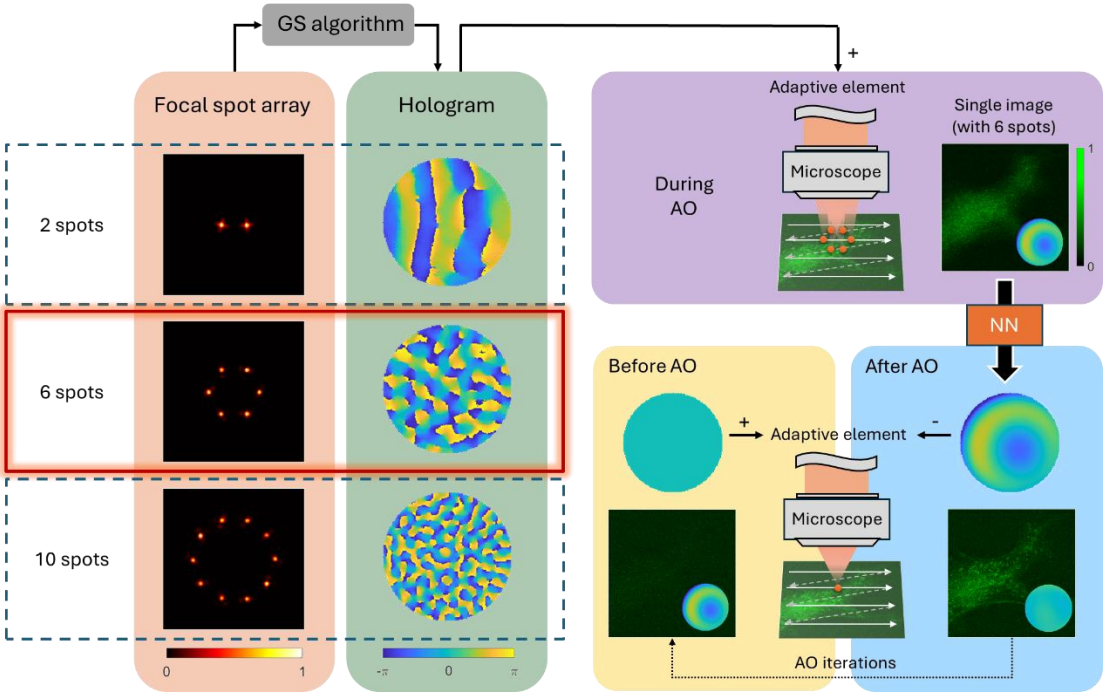


Figure 5-3: A diagram of the method workflow. The focal spot array was first designed to have 2, 6 or 10 spots and then the corresponding phase-only hologram was computed by GS algorithm. During AO, the hologram was added on the adaptive element to generate the focal spot array, which was scanned to measure a single image frame as the input of a pre-trained NN to estimate the aberration for compensation. After AO, the microscope imaged normally with a single focal spot but with better image quality. Further AO iterations could be implemented by option. In this diagram, the phases are in rad and the other images were normalised to their own maximum.

The workflow of the new method was illustrated as in Figure 5-3. The first step was to

design the focal spot array. For the proof-of-principle demonstration in this chapter, the total aberration in the microscope was assumed mainly consisting of $N = 5$ commonly observed Zernike modes including astigmatism (Z_5 and Z_6), coma (Z_7 and Z_8) and spherical (Z_{11}). Accordingly, the focal spot array was designed to have 2, $N + 1 = 6$ or $2N = 10$ spots with each spot containing a different bias, so there were 2, $N + 1 = 6$ or $2N = 10$ biases applied. The three numbers were selected following the same rationale mentioned in Section 4.7: 2 biases are the minimum required for the sensor-less AO method [116]; $N + 1 = 6$ biases are the minimum required to span the aberration space of N modes [99,103]; $2N = 10$ biases allow symmetrical biasing for each aberration mode [104], so testing on such three numbers was enough to deduce the cases of other possible numbers. Based on the design of the focal spot array, the corresponding phase-only hologram was calculated by GS algorithm. During the AO aberration estimation process, the hologram was added on the adaptive element to modulate the phase of the excitation light at the pupil so that the focal spot array was generated at the focal plane as designed if no aberration ϕ_A existed to be estimated or further distorted by ϕ_A. By scanning the focal spot array across a limited field of view (FOV), a single image frame was measured with multiple biased sub-images contained within it. The single image was normalised to a range from 0 to 1 and fed into a pre-trained NN as the input to estimate ϕ_A as $\hat{\phi}_A$ in the form of a vector of Zernike coefficients. After the estimation, the hologram was removed from the adaptive element so that the microscope would image normally with a single focal spot and a conjugate

aberration equivalent to $-\hat{\phi}_A$ was added on the adaptive element to compensate for ϕ_A. As with other sensor-less AO methods, further AO iterations could be conducted to correct any residual aberration if necessary.

5.4 Focal spot array design

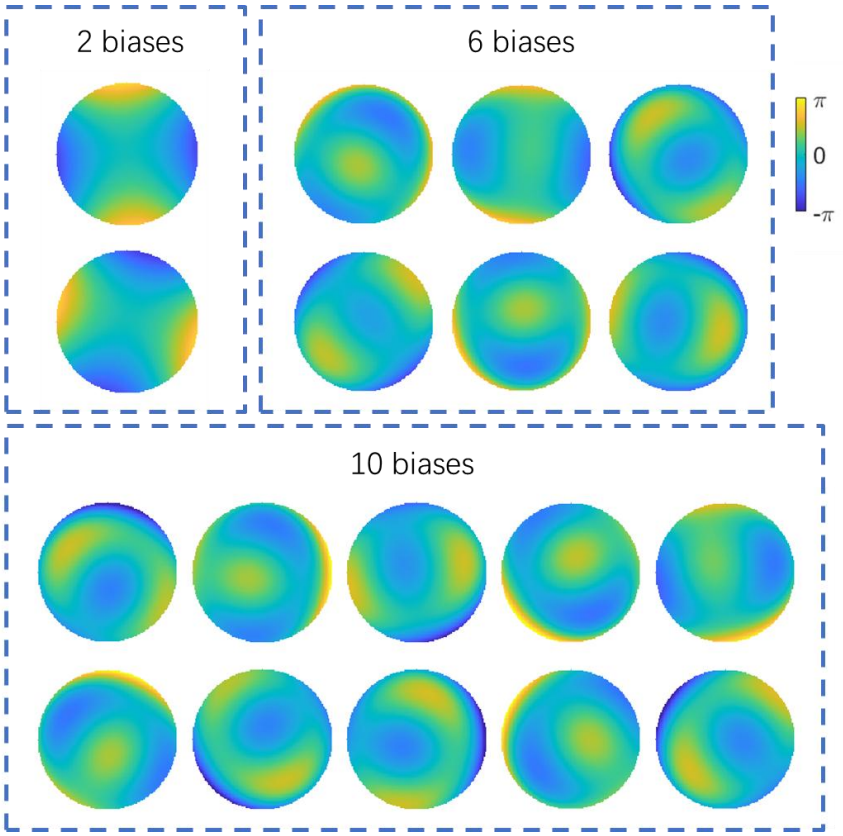


Figure 5-4: The phases in rad of the biases optimised for 2, 6 and 10 focal spots.

The design of the focal spot array involved the biases applied to the spots and the spatial arrangement of the spots. The biases were determined using the framework introduced in Chapter 4 to perform information guided optimisation. To implement the framework specifically for the demonstration in this chapter, the optimised biases obtained with a sequential method using a single focal spot was assumed to be still applicable to the parallelised method using a focal spot array. Thus, the observation X was defined by

the corresponding biased images measured sequentially with a single focal spot. Although there were other unknown parameters contributing to X in this demonstration, Θ was defined solely by the $N = 5$ Zernike coefficients of ϕ_A to ease the Fisher information calculation. The Fisher information $\mathfrak{I}_X(\Theta)$ was calculated based on the same numerical model as in Section 4.1 and the loss was then calculated with $\hat{\phi}_A$ consisting of $M = N$ Zernike coefficient estimators. For the optimisation, the range of scenarios was defined in multiple aspects: the imaging was on a 2-P microscope; the RMS of ϕ_A was limited within 2.236 rad; O was an extended volumetric object with arbitrary structures; the photon budget ξ of each biased image was in a range from 20,000 to 1,000,000; the background β of each biased image was 1. The optimisable parameters were all the Zernike coefficients representing the biases. Each bias coefficient was initialised by a random value drawn from a uniform distribution within 1.0 rad. The batch size was set at 16 when there were 2 or 6 biases, and reduced to 8 when there were 10 biases. In each example scenario, ϕ_A was drawn from the surface of an n-sphere as in [153] with the n-sphere radius corresponded to an RMS value drawn from a uniform distribution in the limited RMS range of ϕ_A; O had a half chance being generated by randomly combining geometric structures of different numbers, sizes and shapes, and the other half chance being generated by drawing a stack of consecutive images from a pool of 350 images taken by a high-resolution 2-P microscope imaging various biological samples, down-sampling the drawn image stack and finally applying a random rotation; ξ was drawn from a uniform distribution. The

biases were optimised for 2000 steps with the learning rate starting from 0.05 and reducing by half every 400 steps. Figure 5-4 showed the phases of the biases optimised for 2, 6 and 10 focal spots.

On the other hand, the spatial arrangement was chosen empirically as a balance between maintaining a compact focal spot array while minimising overlap between adjacent spots. As a suitable solution for the demonstration in this chapter, the spots were placed evenly on the focal plane to form a circle centred at the axis with the distance between adjacent spots at 8 times the FWHM of the diffraction-limited 2-P PSF on the focal plane. Technically, the spatial arrangement could also be optimised like the biases. However, such optimisation was not practicable as there were many different arrangements resulting in almost identical performance. Hence, a fixed spatial arrangement was chosen. By default, each focal spot was designed to have an equal share of the total intensity.

5.5 Hologram calculation

When the phase-only hologram was calculated by GS algorithm, the hologram plane was defined by the pupil plane while rather than defining the focal plane as the only diffraction plane, two diffraction planes were defined by two defocused planes on either side of the focal plane. The use of two diffraction planes was to resolve the ambiguity mentioned in Section 4.6.2 that two different aberrations result in the same light intensity on the focal plane when they are opposite in even radial order modes and the same in odd radial order modes.

Table 5-1. The pseudo-code of the hologram calculation by GS algorithm

Input:

The correct light intensity on the pupil plane: I_1

The phase shift at the pupil corresponding to 0.5 rad defocus: ζ

The target light intensity on the +0.5 rad defocused plane: I_{2+}

The target light intensity on the -0.5 rad defocused plane: I_{2-}

Procedure:

Initialise the phase on the pupil plane ϕ_1 with random values in $(-\pi, \pi]$

Normalise I_{2+} and I_{2-} to their respective maximum

Set $error = 2;\ count = 0$

While True

Build the light field $E_{1+} = \sqrt{I_1}e^{i(\phi_1+\zeta)},\ E_{1-} = \sqrt{I_1}e^{i(\phi_1-\zeta)}$

Calculate the light field $E_{2+} = \mathcal{F}(E_{1+}),\ E_{2-} = \mathcal{F}(E_{1-})$

Get I'_{2+} and ϕ_{2+} as the intensity and phase of E_{2+}

Get I'_{2-} and ϕ_{2-} as the intensity and phase of E_{2-}

Normalise I'_{2+} and I'_{2-} to their respective maximum

Calculate $error_+$ as the mean absolute error between I'_{2+} and I_{2+}

Calculate $error_-$ as the mean absolute error between I'_{2-} and I_{2-}

If $error < error_+ + error_-$

$count = count + 1$

Else

$count = 0$

End If

If $count = 0$

Set the hologram ϕ_H as ϕ_1

Update $error$ by $error_+ + error_-$

End If

If $count > 10$

Break the While loop

End If

Set the light field $E_{2+} = \sqrt{I_{2+}}e^{i\phi_{2+}},\ E_{2-} = \sqrt{I_{2-}}e^{i\phi_{2-}}$

Set the light field $E_{1+} = \mathcal{F}^{-1}(E_{2+}),\ E_{1-} = \mathcal{F}^{-1}(E_{2-})$

Get ϕ_{1+} by the phase of E_{1+} minus ζ

Get ϕ_{1-} by the phase of E_{1+} plus ζ

Calculate $\Delta\phi_1 = \phi_{1+} - \phi_{1-}$ and then wrap $\Delta\phi_1$ to $(-\pi, \pi]$

Update ϕ_1 by $\phi_{1+} - \Delta\phi_1/2$ and then wrap ϕ_1 to $(-\pi, \pi]$

End While

End Procedure

Output:

The hologram ϕ_H

The hologram pupil illumination was assumed to be uniform. The target light intensity on either of the diffraction plane was derived from the focal spot array design plus a global amount of defocus. The propagation between the hologram plane and either of the diffraction plane was modelled by the equation (3-2). The GS algorithm was implemented in two parallel branches, one for each of the two diffraction planes, and the hologram was obtained by combining the calculation results from both branches. The pseudo-code of this process for this demonstration was provided in Table 5-1.

5.6 Neural network implementation

In this demonstration, the NN adopted the ResNet-18 model architecture provided in Section 2.5.4. The single image measured with the focal spot array was normalised to a range from 0 to 1 and served as the NN input. The output of the NN was a vector of $M = N$ Zernike coefficient estimators representing $\hat{\phi}_A$. For the NN training, a dataset of 100,000 examples was prepared and an extra smaller dataset of 1000 examples was also prepared for the validation. Each example consisted of an NN input labelled with its corresponding Zernike coefficients of ϕ_A. To generate each example, ϕ_A was randomly drawn from the surface of an n-sphere as in [153] with the radius corresponding to an RMS value drawn from a uniform distribution ranging from 0 to 2.236 rad. The corresponding image was obtained by numerical simulation of imaging with the focal spot array formed by applying the calculated hologram. The pixel dimension of the image was either 128×128 or 256×256. The pixel size of the image was set as a random value drawn from a uniform distribution between 0.4 and

0.6 times of the FWHM of the diffraction-limited 2-P PSF on the focal plane (which was roughly between 0.11 and 0.16 μm). In the numerical simulation, the object O was generated in the same way as in the bias optimisation in Section 5.4; the 3-D convolution was calculated by sampling seven axial planes with the central plane placed on the focal plane and the spacing between any two adjacent planes as a half of the FWHM of the diffraction-limited 2-P PSF along the optical axis (which was about 0.40 μm); the signal amplification ξ was defined as the photon budget of each sub-image contained in the image and was set randomly from a uniform distribution between 20,000 and 1,000,000; the background β of the whole image was fixed at 1; the distribution R was modelled as a random mix of Poisson distributed shot noise, Gaussian distributed white noise and noise from electrical interference. The noise from electrical interference was modelled as an additive sinusoidal pattern of arbitrary frequency and orientation, which was generated by taking the real part of the Fourier transform of a random one-hot image. Hence, the training data covered a wide range of scenarios which ensured the NN would generalise well to the actual scenarios that could appear in this demonstration after training.

The NN was separately trained for different cases of focal spot numbers and image pixel dimensions in this demonstration. During the training, the NN parameters were initialised by Glorot uniform initialisation [154]; the batch size was 32; the loss was defined by the mean error of $\hat{\phi}_A$ in the RMS; the Adam gradient descent optimiser trained the NN for 50 epochs [151]; the learning rate started from 10^{-3} and decayed

to 1/3 of its previous value when the decrease of the validation loss was less than 10^{-3} for two epochs in a row until the learning rate finally reached 10^{-5}; both L1 and L2 regularisation were applied to prevent overfitting. After training, the NN typically completed the estimation in 8ms. The NN implementation was programmed with TensorFlow in Python and accelerated by a GPU (Nvidia GeForce RTX 3070). It has been reported in [116] that a similarly implemented NN could only work on images of point-like objects. However, with a much larger proportion of the image showing effective signal when the focal spot array was used, the NN capacity got better employed so that the NN could also be used for extended objects.

5.7 Experimental setup

The method was experimentally demonstrated on a 2-P microscope with a SLM as the adaptive element. A schematic diagram of the experimental setup was presented in Figure 5-5. In this setup, a femtosecond pulsed laser beam of 850 nm wavelength was adjusted on power and polarisation by a half wave plate (HWP) coupled with a polarised beam splitter (PBS). The beam was expanded by cascaded 4f systems until it reached the SLM for phase wavefront modulation. A blazed grating was formed on the SLM to separate the modulated light in the first order from unmodulated light in the zero order which was blocked out by a spatial filter after the SLM. The SLM was conjugated to the pupil plane of the objective lens with a series of relay lenses in 4f configuration and a pair of galvanometer mirrors for raster scanning. The modulated beam was focused by the objective lens into the sample for 2-P excitation. The objective lens was mounted

on a piezo stage, which enabled axial translation of the focus within the sample. The emitted 2-P fluorescence was collected by the objective lens and reflected by a dichroic filter into a PMT for detection. The setup was calibrated by replacing the objective lens with a power meter to measure the excitation power while changing the HWP angle so that the HWP angle was mapped to the excitation power for precise control.

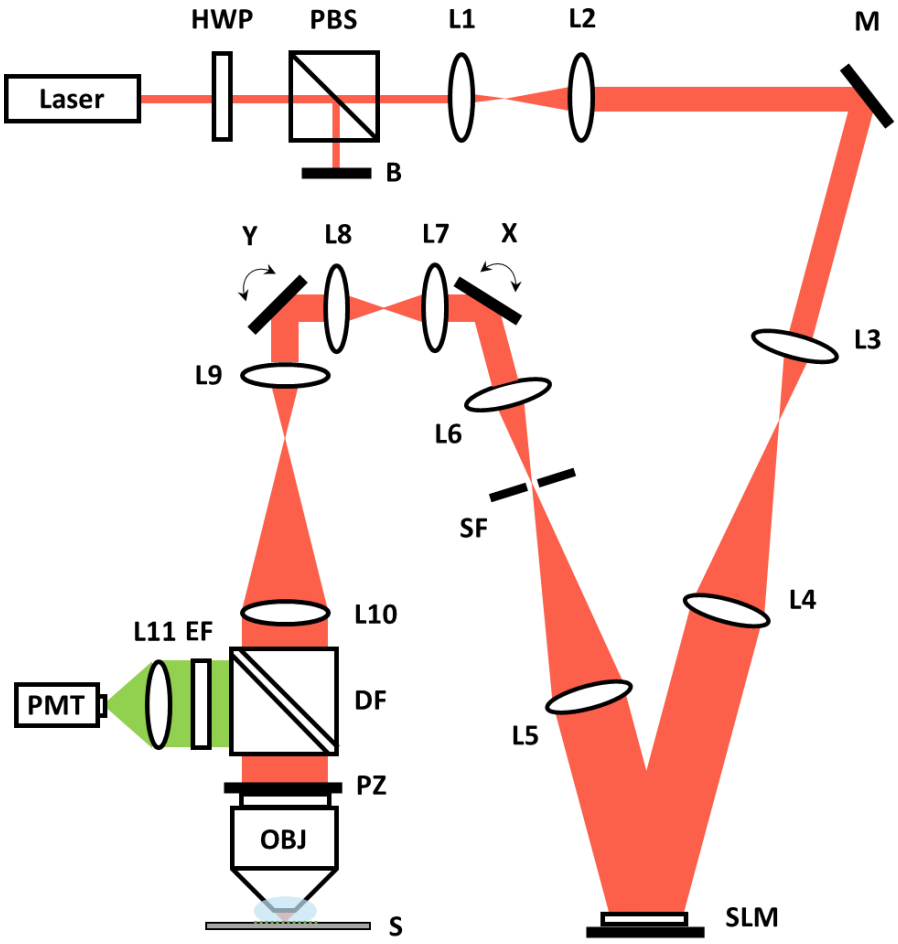


Figure 5-5: A schematic diagram of the experimental setup. Laser: femtosecond pulsed laser at 850 nm wavelength; HWP: half wave plate; PBS: polarized beam splitter; B: block; L1-11: lenses; M: mirror; SLM: spatial light modulator; SF: spatial filter; X&Y: X&Y Galvanometer scanning mirrors; DF: dichroic filter; EF: emission filter; PMT: photomultiplier tube; PZ: Z-piezo translation stage; OBJ: objective (40x, 1.15NA, water immersion); S: sample stage.

5.8 Experiments imaging fluorescent beads

The method was firstly demonstrated by imaging experiments using fluorescent beads. For the imaging, each image was measured from a $20 \times 20\ \mu m$ FOV in 128×128 pixels with the dwell time of each pixel at 10 μs and the signal level of each image was controlled by adjusting the excitation power. The method performance was tested in different cases of spot numbers, signal levels, aberration sizes, bead sizes and sparsity levels. As fluorescent beads are much more robust to photodamage compared to biological samples, multiple experiments were conducted for each tested case to obtain statistical results. In each experiment, the unknown system aberration was corrected in advance by the 5N parabolic fitting method in [104], while a known aberration was introduced as ϕ_A by the SLM to be estimated and corrected. For the result assessment, an image quality metric y_I was defined based on the image intensity so that the method performance could be quantified by the improvement of y_I after each AO iteration. The pseudo-code of y_I calculation was provided in Table 5-2.

Table 5-2. The pseudo-code of y_I calculation

Input: the image D
Procedure:
Get ΔD as the mean of the smallest 5% values in D
Shift the value range of D to positive by $D = D - \Delta D$
Get y_I as the sum of the largest 350 values in D
End Procedure
Output: y_I

5.8.1 Imaging fluorescent beads with different signal levels

In this section, each image was measured from a fixed FOV containing fluorescent

beads of $2\mu m$ diameter. The total signal in each image was controlled at a low, medium or high level corresponding to the excitation power at 5.8, 11.6 or 23.2 mW when imaging with a single focal spot and $\sqrt{T}$ times of the power when imaging with an array of $T = 2, 6 \text{ or } 10$ focal spots. Figure 5-6 presented image examples of the three different signal levels with different numbers of focal spots used for imaging when no ϕ_A existed. The influence from noise became more and more obvious in these images as the peak signal dropped. This included some electrical signal interference picked up by the photodetectors, which was visible as parallel bands in these images. As expected, images taken with multiple focal spots contained laterally offset sub-images and images taken with more focal spots showed lower peak signal as the same total signal was spread over more pixels.

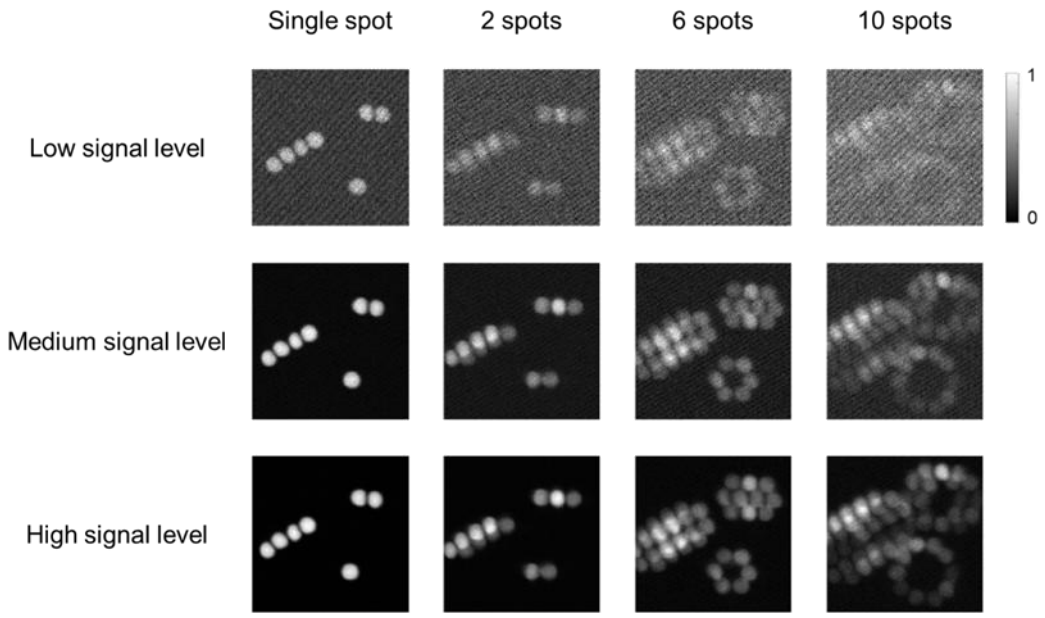


Figure 5-6: Image examples of a fixed $20 \times 20\ \mu m$ FOV containing fluorescent beads with the low, medium and high signal levels and a single, 2, 6 and 10 focal spot(s) used for imaging when no ϕ_A existed. Each image example was normalised to a range between 0 and 1.

The method performance was tested independently for the three different signal levels with 2, 6 or 10 spots used. For each combination of signal level and spot number, 20 experiments were conducted. In each experiment, ϕ_A was randomly drawn from the surface of an n-sphere as in [153] with the radius corresponding to an RMS smaller than 2.236 rad and corrected for five AO iterations. Apart from the single image measured for the aberration estimation in each AO iteration, an image was taken with a single focal spot before AO and after each AO iteration to calculate y_I for performance assessment.

Figure 5-7 presented the change of y_I over the five AO iterations in each group of 20 experiments. Here, y_I was normalised such that its mean value before AO equalled 1, and then mean y_I was plotted against the number of images measured for AO, with the shaded area representing the standard deviation of y_I as a performance consistency indicator. In the low signal level case, the method gradually improved y_I to an intermediate level over the five AO iterations when using 6 spots and failed to bring any improvement when using 2 or 10 spots. In the medium signal level case, y_I was improved for all three spot numbers with most improvement in the first AO iteration; for 6 and 10 spots, similar performance was observed with y_I converging at a much higher level after five AO iterations with better consistency than the low signal level case, while the method performed less well with 2 spots. In the high signal level case, the largest and fastest y_I improvement was observed for all the three spot numbers. Except for the more challenging low signal case, most of the correction generally

happened in the first iteration, hence sensor-less AO based on a single image was made possible with the method.

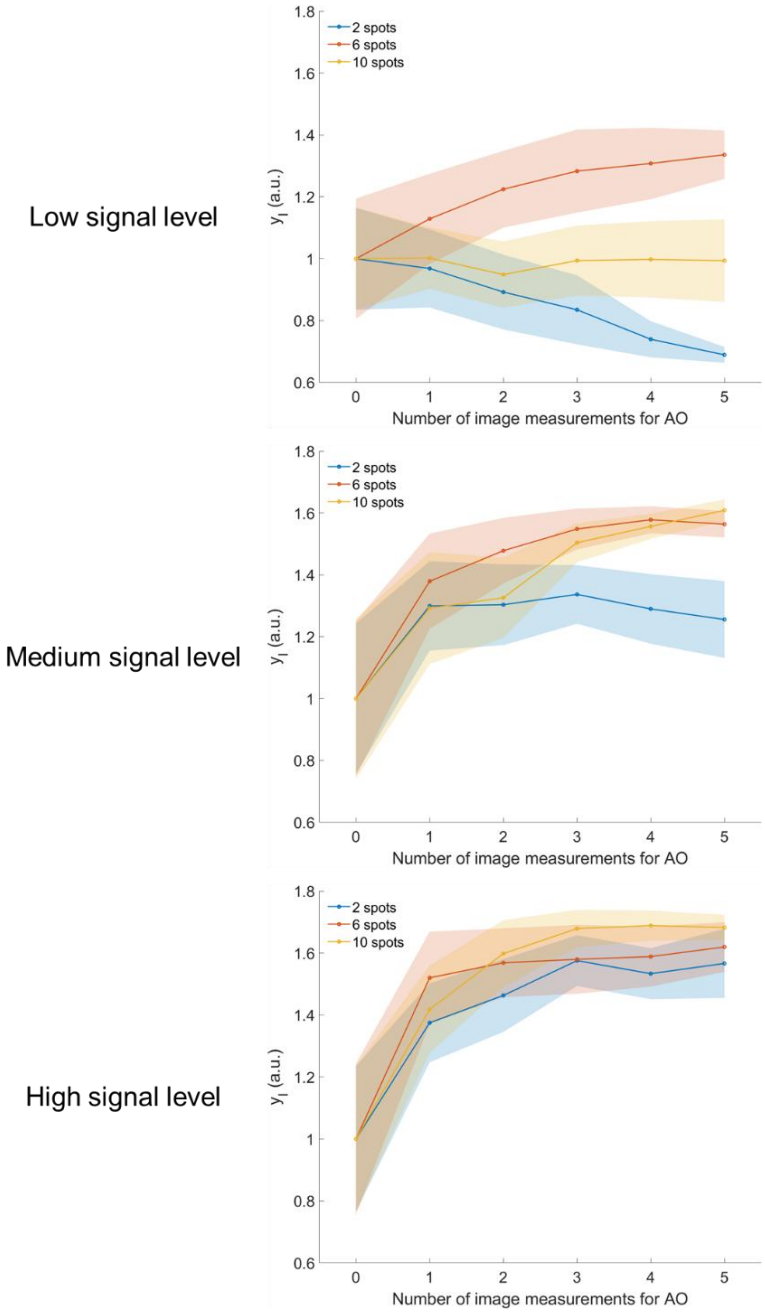


Figure 5-7: The plots of mean y_I against the number of images measured with 2, 6 and 10 focal spots for AO in the experiments of imaging a fixed FOV containing fluorescent beads with the low, medium and high signal levels when the RMS of ϕ_A was within 2.236 rad. y_I was normalised such that its mean value before AO equalled 1. The shaded areas represented the standard deviations of y_I.

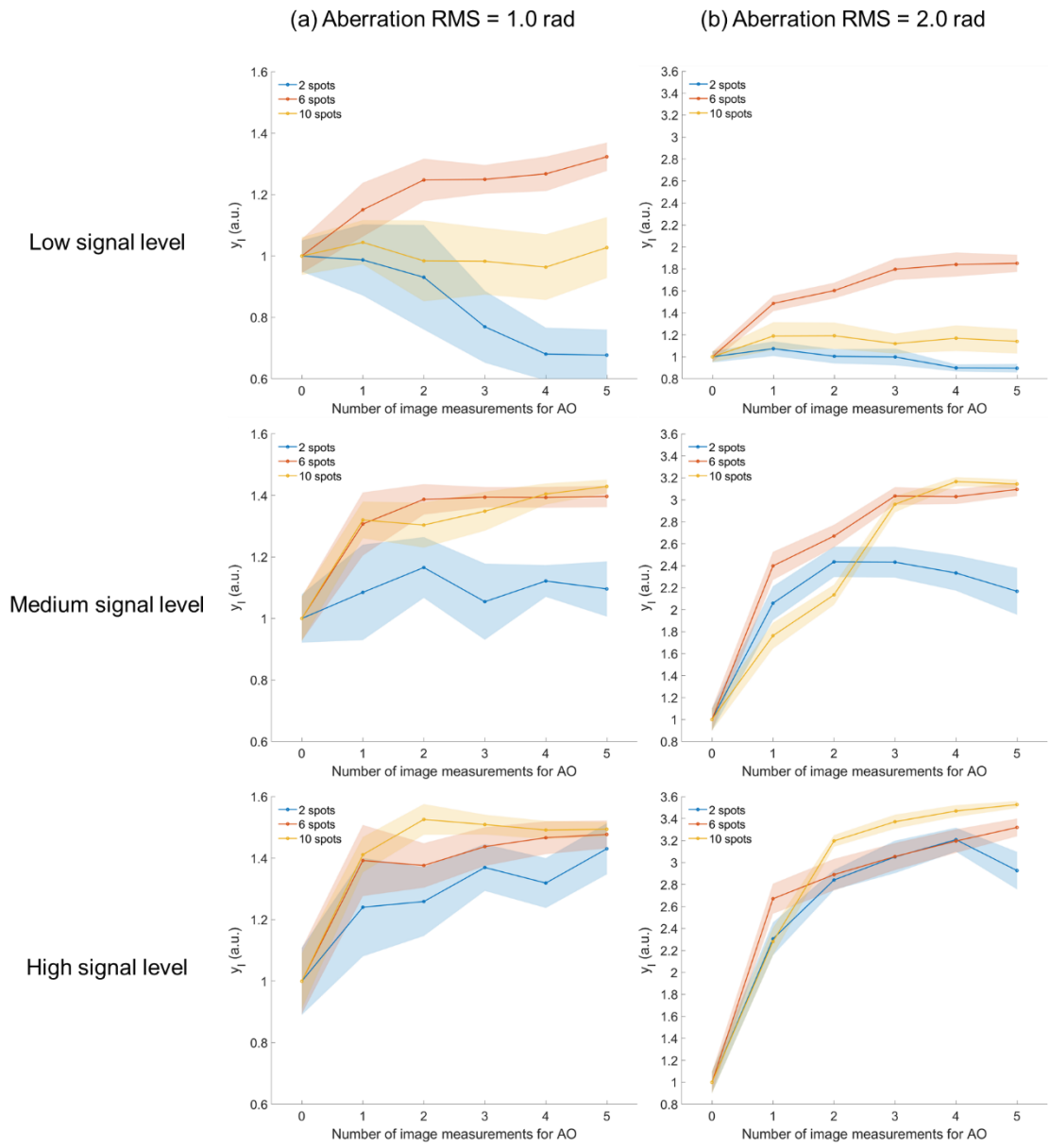


Figure 5-8: The plots of mean y_I against the number of images measured with 2, 6 and 10 focal spots for AO in the experiments of imaging a fixed FOV containing fluorescent beads with the low, medium and high signal levels when the RMS of ϕ_A was 1.0 rad (a) and 2.0 rad (b). y_I was normalised such that its mean value before AO equalled 1. The shaded areas represented the standard deviations of y_I.

To check if different sizes of ϕ_A would cause different test results, the experiments were repeated with the RMS of the ϕ_A drawn in each experiment controlled at 1.0 and 2.0 rad respectively. The corresponding results were presented in Figure 5-8(a) and 5-8(b). Compared to Figure 5-7, similar trends were shown in Figure

5-8, which suggested that the change of method performance with either the signal level or the spot number was independent from the size of ϕ_A. On the other hand, Figure 5-8(a) showed in general smaller y_I improvement than Figure 5-7, while much larger relative y_I improvement was generally observed in Figure 5-8(b). This is mainly because y_I before AO was lower for larger ϕ_A.

5.8.2 Imaging fluorescent beads of different sizes

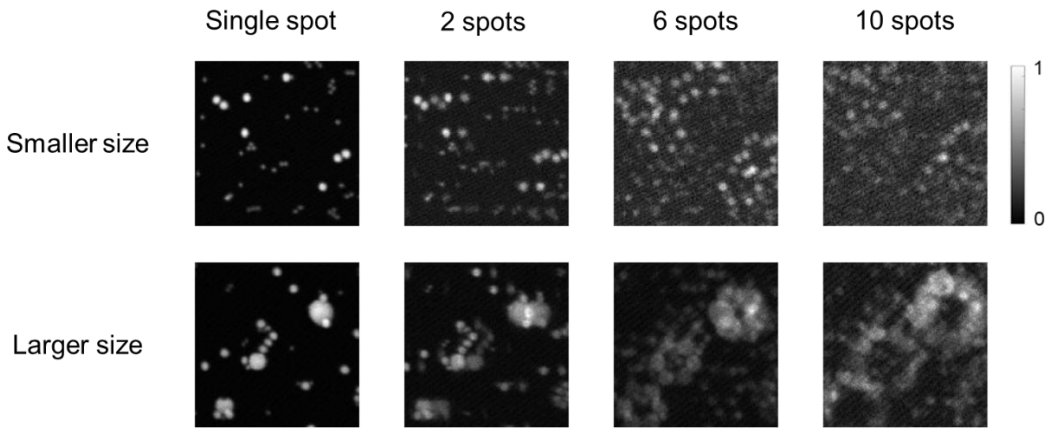


Figure 5-9: Image examples of the two selected $20 \times 20\ \mu m$ FOVs containing fluorescent beads of smaller and larger sizes with a single, 2, 6 and 10 focal spot(s) used for imaging when no ϕ_A existed. Each image example was normalised to a range between 0 and 1.

The method performance was further tested with object structures of different sizes in this section. For the tests, fluorescent beads of 0.5, 1.0 and 2.0 μm diameters were mixed together and two FOVs were selected for imaging with one FOV containing more smaller beads and the other FOV containing more larger beads. The signal level of each image was controlled by setting the excitation power at 15.6 mW for a single focal spot and increasing the excitation power by $\sqrt{T}$ times that value when changing to $T = 2, 6$ or 10 focal spots. Figure 5-9 displayed image examples from the two

FOVs with different numbers of focal spots used for imaging when there was no ϕ_A.

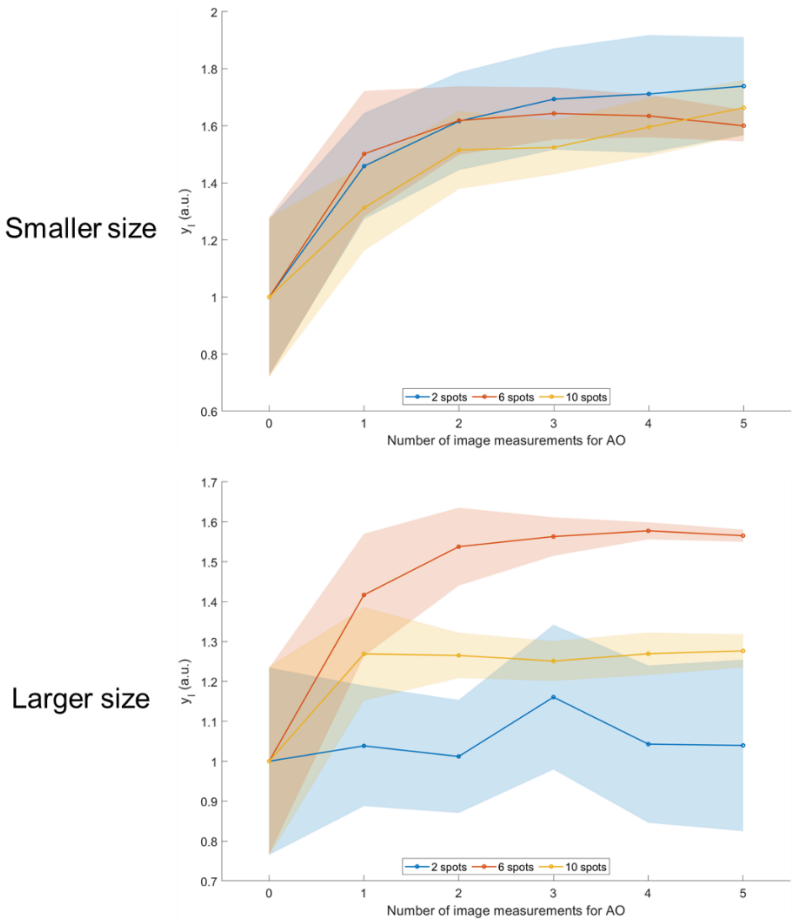


Figure 5-10: The plots of mean y_I against the number of images measured with 2, 6 and 10 focal spots for AO in the experiments of imaging two selected FOVs containing fluorescent beads of smaller and larger sizes when the ϕ_A RMS was within 2.236 rad. y_I was normalised such that its mean value before AO equalled 1. The shaded areas represented the standard deviations of y_I.

The experimental results were obtained and presented in the same way as in the last section. Figure 5-10 provided the results when the RMS of the ϕ_A drawn in each experiment was within 2.236 rad. With more smaller beads in the FOV, y_I was improved significantly in the first AO iteration and then converged at a similar level for all the three spot numbers. Comparatively, for the FOV containing more larger beads, a slight decrease was observed in the y_I improvement using 6 spots, while y_I was

only improved to a much lower level using 10 spots and barely improved using 2 spots. When the RMS of the ϕ_A drawn in each experiment controlled at 1.0 and 2.0 rad respectively, the results were presented in Figure 5-11(a) and 5-11(b). Again, similar trends as in Figure 5-10 were observed in Figure 5-11, thus the effect of object structures size had a more significant influence on performance than the size of ϕ_A.

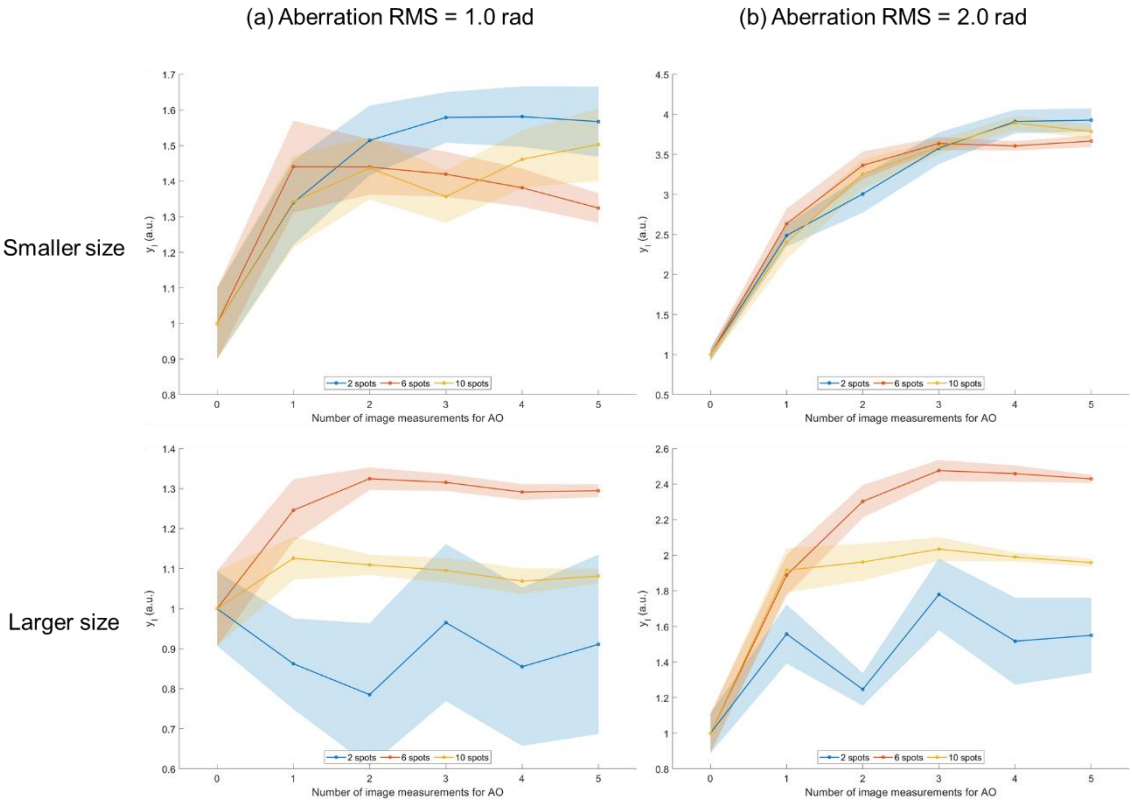


Figure 5-11: The plots of mean y_I against the number of images measured with 2, 6 and 10 focal spots for AO in the experiments of imaging two selected FOVs containing fluorescent beads of smaller and larger sizes when the ϕ_A RMS was 1.0 rad (a) and 2.0 rad (b). y_I was normalised such that its mean value before AO equalled 1. The shaded areas represented the standard deviations of y_I.

5.8.3 Imaging fluorescent beads of different sparsity levels

As the method relied on a degree of sparsity of object structures, different sparsity levels were created by including different numbers of 2.0μm diameter fluorescent beads in

the fixed size FOV to test the method performance. Specifically, three FOVs containing 3, 11 and 27 beads were selected to represent high, medium and low sparsity levels. It could be concluded from previous results that 6 spots provided the most robust performance, so here the tests were conducted only for the case of 6 spots. To control the signal level, the excitation power was set at 22.8 mW for a single spot and 55.8 mW for 6 spots. With no ϕ_A, image examples from the three FOVs using a single spot and 6 spots were presented in Figure 5-12.

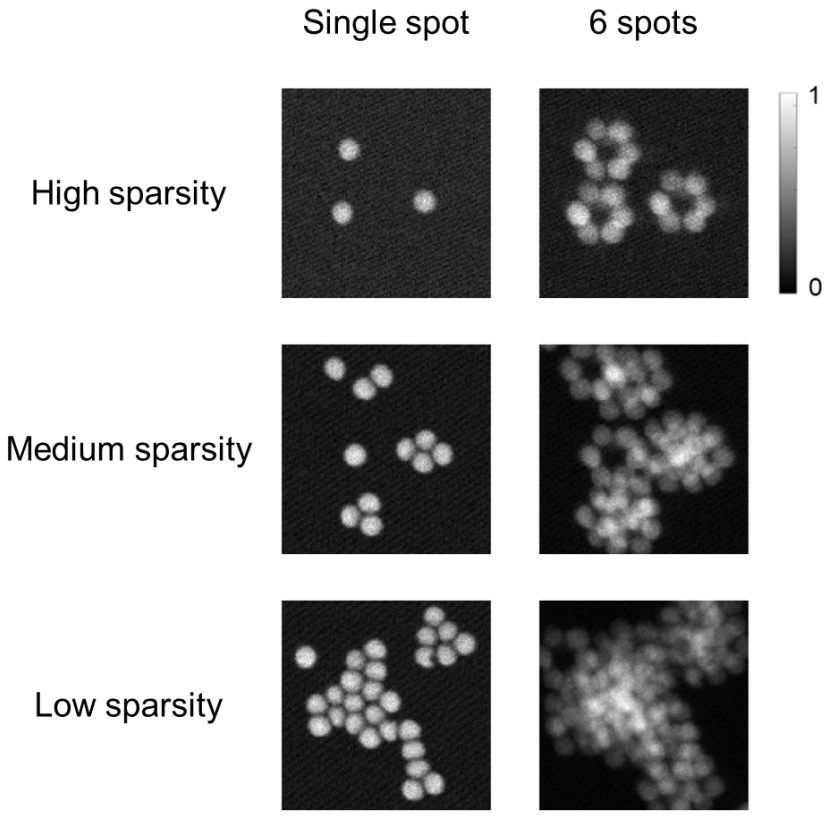


Figure 5-12: Image examples of the three selected $20 \times 20\ \mu m$ FOVs containing fluorescent beads of high, medium and low sparsity levels with a single spot and 6 spots used for imaging when no ϕ_A existed. Each image example was normalised to a range between 0 and 1.

For each sparsity level, 20 experiments were similarly conducted with ϕ_A randomly drawn from the surface of an n-sphere as in [153] with the radius

corresponding to an RMS smaller than 2.236 rad and corrected for five AO iterations in each experiment. As in the previous two sections, y_I was calculated, normalised and plotted in Figure 5-13. The largest and fastest y_I improvement was observed in the medium sparsity case. With the high sparsity, y_I improved gradually over the five AO iterations, while the method was unable to provide steady correction in the low sparsity case.

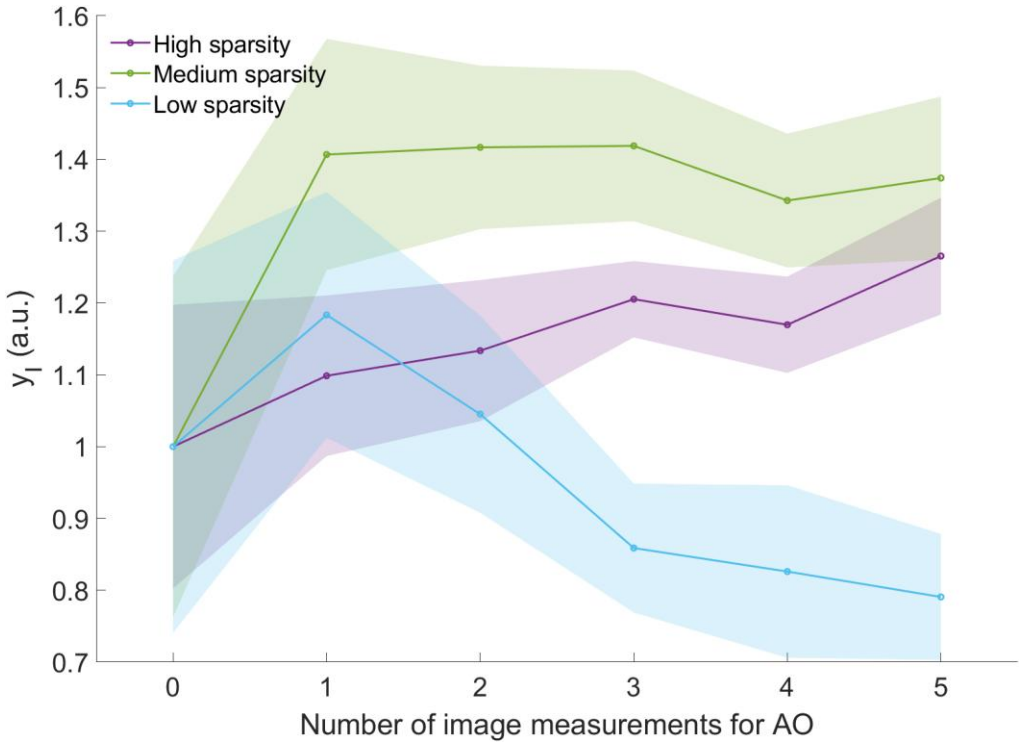


Figure 5-13: The plots of mean y_I against the number of images measured with 6 focal spots for AO in the experiments of imaging the three selected FOVs containing fluorescent beads of high, medium and low sparsity levels when the ϕ_A RMS was within 2.236 rad. y_I was normalised such that its mean value before AO equalled 1. The shaded areas represented the standard deviations of y_I.

The experiments were however conducted in an alternative way to test how the method would perform with different sparsity levels when correcting ϕ_A of different sizes. Here, the RMS of ϕ_A was controlled at 0.5, 1.0, 1.5 and 2.0 rad. For each case

of the RMS and sparsity level, 50 experiments were performed with ϕ_A still randomly drawn from the surface of an n-sphere as in [153] but only corrected for one AO iteration based on only a single image measurement in each experiment. Instead of y_I, the residual aberration RMS after AO was calculated to compare with the RMS of ϕ_A introduced before AO.

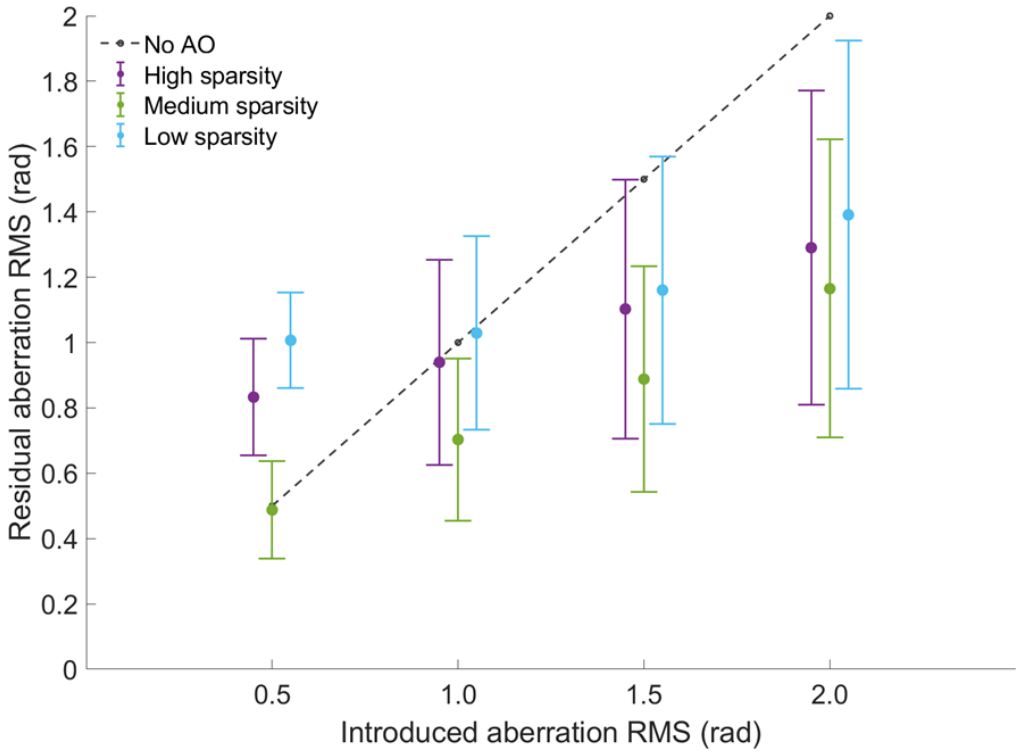


Figure 5-14: The plots of the mean residual aberration RMS after one AO iteration based on a single image measured with 6 focal spots against the introduced ϕ_A RMS in the experiments of imaging the three selected FOVs containing fluorescent beads of high, medium and low sparsity levels when the ϕ_A RMS was controlled at 0.5, 1.0, 1.5 and 2.0 rad. The error bars represented the standard deviation of the residual aberration RMS. The dashed line showed the case without AO.

The comparison results were presented in Figure 5-14 where the mean residual aberration RMS was plotted with the error bar representing the standard deviations of the residual aberration RMS and the case without AO was shown by the dashed line for

reference. No matter how much the size of introduced ϕ_A was, the residual aberration after AO was always the smallest in the medium sparsity case and the largest in the low sparsity case, which agreed with the results in Figure 5-13. For all the three sparsity levels, the residual aberration became larger with larger ϕ_A whereas more aberration was corrected. It was also observed that the correction only took effect when the RMS of ϕ_A was larger than 0.5 rad for the medium sparsity and 1.0 rad for the other two sparsity levels.

5.9 Experiments imaging biological samples

Instead of fluorescent beads, the method was then demonstrated by experiments of imaging biological structures. To match with the extended scale of object structures, each image was expanded to 256×256 pixels corresponding to a $40 \times 40\ \mu m$ FOV. For the most robust performance, the method used 6 spots according to the results obtained with fluorescent beads. The signal level of each image was controlled by setting the excitation power at 15.6 mW for a single spot and 72.0 mW for 6 spots. Since biological samples are much more sensitive to photodamage, a single experiment was performed for each FOV while multiple different FOVs were selected from different biological samples to test the method performance. In each experiment, the system aberration was corrected in advance with fluorescent beads by the 5N parabolic fitting method in [104], while ϕ_A came from either a known aberration introduced by the SLM for a small imaging depth or an unknown aberration induced by the sample for a large imaging depth. ϕ_A was corrected for five AO iterations. To evaluate the

method performance, an image was taken before AO and after each AO iteration with the image quality metric y_I calculated and normalised such that its value before AO equals 1. When correcting a known ϕ_A introduced by the SLM, an aberration free image was also taken for reference before ϕ_A was introduced and the residual aberration was calculated after each AO iteration.

5.9.1 Imaging Bovine Pulmonary Artery Endothelial cells

In this section, different FOVs were selected from a sample (FluoCells™ Prepared Slide #1) containing fixed Bovine Pulmonary Artery Endothelial (BPAE) cells to image fluorescent labelled microtubules of various structures. As the sample was thin, a known aberration of 2.0 rad RMS was randomly drawn from the n-sphere surface as ϕ_A in each experiment [153].

For two selected FOVs, the first one containing dim structures and the second one containing bright structures, the corresponding experimental results were presented in Figure 5-15(a) and 5-15(b). Here, with the aberration free images as references, the images taken before AO and after each AO iteration were shown for comparison. The inset on each image showed the corresponding residual aberration and normalised y_I was plotted against the number of images measured for AO. Due to the dim structures, the images in Figure 5-15(a) generally displayed a low signal level with distinct noise, while in Figure 5-15(b) the images of bright structures showed a high signal level with little noise. For both FOVs, the method achieved effective correction after five AO iterations with y_I improved, residual aberration reduced and image details restored,

however much better and faster correction was observed in Figure 5-15(b) showing the higher signal level compared to Figure 5-15(a) showing the lower signal level, which coincided with the previous results of different signal levels.

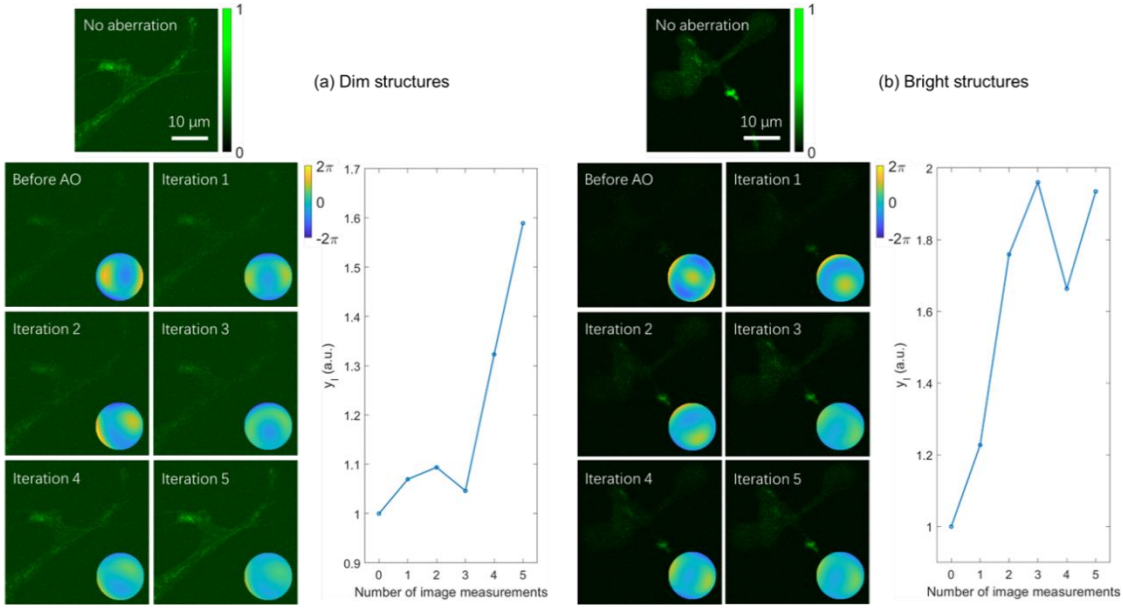


Figure 5-15: The experimental results of imaging two $40 \times 40\ \mu m$ FOVs containing dim (a) and bright (b) microtubule structures selected from a sample of fixed BPAE cells with a known introduced ϕ_A corrected by the method using 6 spots for five AO iterations. With the aberration free images as references, the images taken before AO and after each AO iteration were shown for comparison. The inset on each image showed the corresponding residual aberration in rad. y_I was normalised such that its value before AO equalled 1 and plotted against the number of images measured for AO.

Figure 5-16 presented the experimental results for another two selected FOVs. In Figure 5-16(a), the images showed many small granular structures in one FOV of the two, while there were mainly large continuous structures in the images of the other FOV shown in Figure 5-16(b). Again, the image was improved in quality with most aberration corrected after five AO iterations in both cases. Meanwhile, despite a fluctuation in the middle of the AO process, the method achieved better correction for

the small structures than the large structures, which was similar to the results with fluorescent beads.

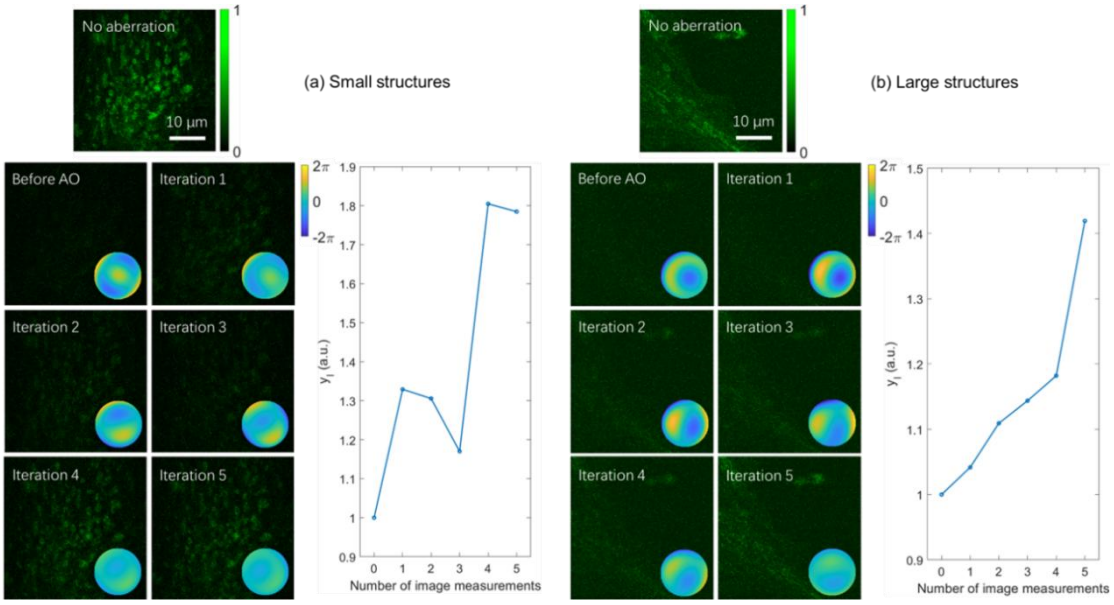


Figure 5-16: The experimental results of imaging two $40 \times 40\ \mu m$ FOVs containing small (a) and large (b) microtubule structures selected from a sample of fixed BPAE cells with a known introduced ϕ_A corrected by the method using 6 spots for five AO iterations. With the aberration free images as references, the images taken before AO and after each AO iteration were shown for comparison. The inset on each image showed the corresponding residual aberration in rad. y_I was normalised such that its value before AO equalled 1 and plotted against the number of images measured for AO.

In addition, a group of three FOVs containing structures of high, medium and low sparsity levels were selected and the experimental results were provided in Figure 5-17. As shown in Figure 5-17(b), the image quality was improved drastically to a level close to the one without aberration in the first two AO iterations when the structures were of medium sparsity. In contrast, the improvement in the image quality was less significant after five AO iterations in the other two cases shown in Figure 5-17(a) and 5-17(c). For the high sparsity level, the correction was also much slower. These observations

confirmed again that the method performed the best in the case of medium sparsity.

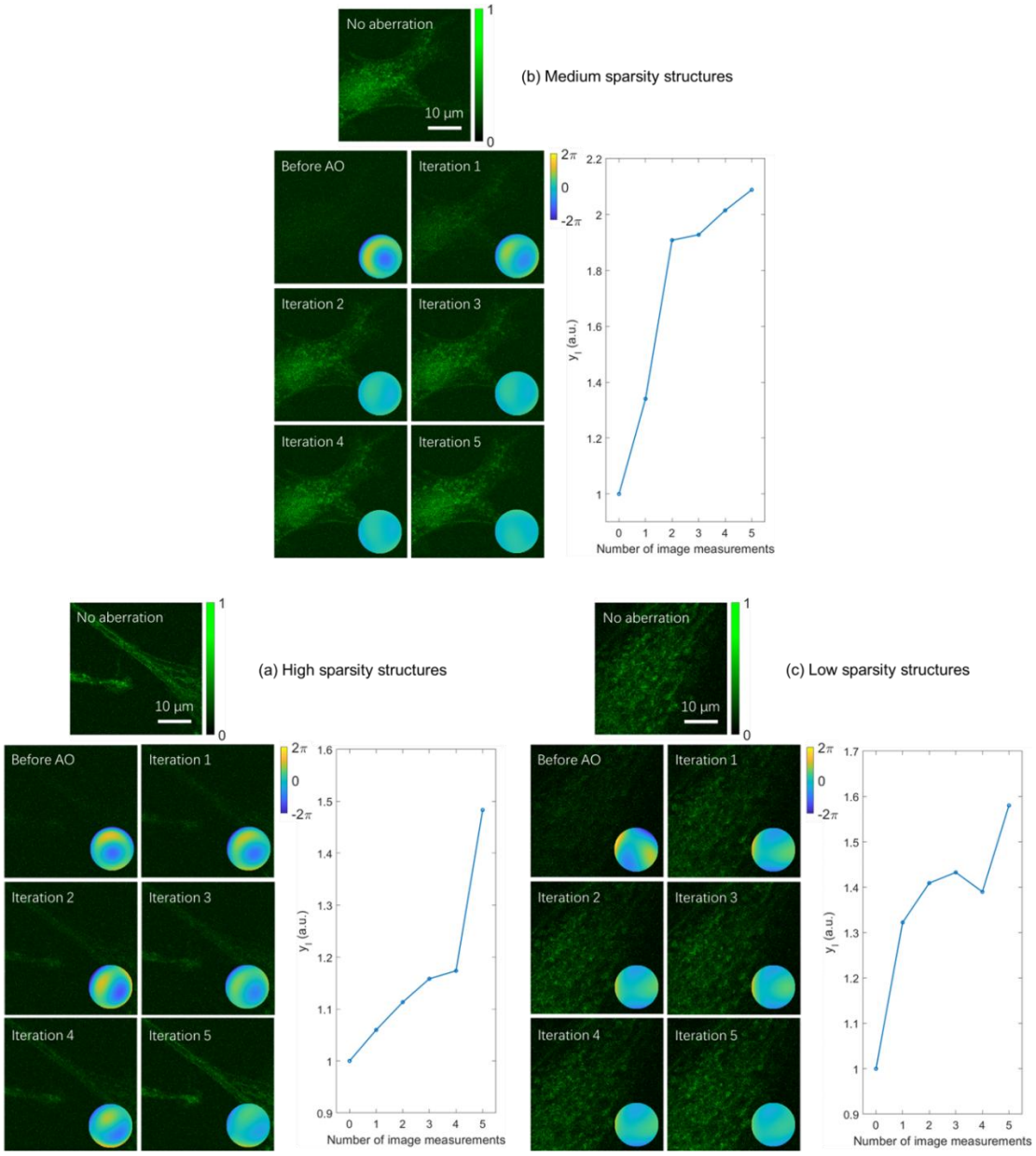


Figure 5-17: The experimental results of imaging three $40 \times 40\ \mu m$ FOVs containing high (a), medium (b) and low (c) sparsity microtubule structures selected from a sample of fixed BPAE cells with a known introduced ϕ_A corrected by the method using 6 spots for five AO iterations. With the aberration free images as references, the images taken before AO and after each AO iteration were shown for comparison. The inset on each image showed the corresponding residual aberration in rad. y_I was normalised such that its value before AO equalled 1 and plotted against the number of images measured for AO.

5.9.2 Imaging Thy1-YFP mouse brain tissue

The method performance was further tested with a sample of 250 μm thick fixed Thy1-YFP mouse brain tissue where two FOVs were selected separately at 100 and 200 μm depths to image fluorescent labelled neural structures. Here, the aberration from the sample itself acted as ϕ_A in each experiment. For each selected FOV, the images taken before AO and after each AO iteration were compared in Figure 5-18 with normalised y_I plotted against the number of images measured for AO. Still, the image quality was considerably improved after AO for both cases, even though more distinct noise was observed in Figure 5-18(b) as a result of larger imaging depth.

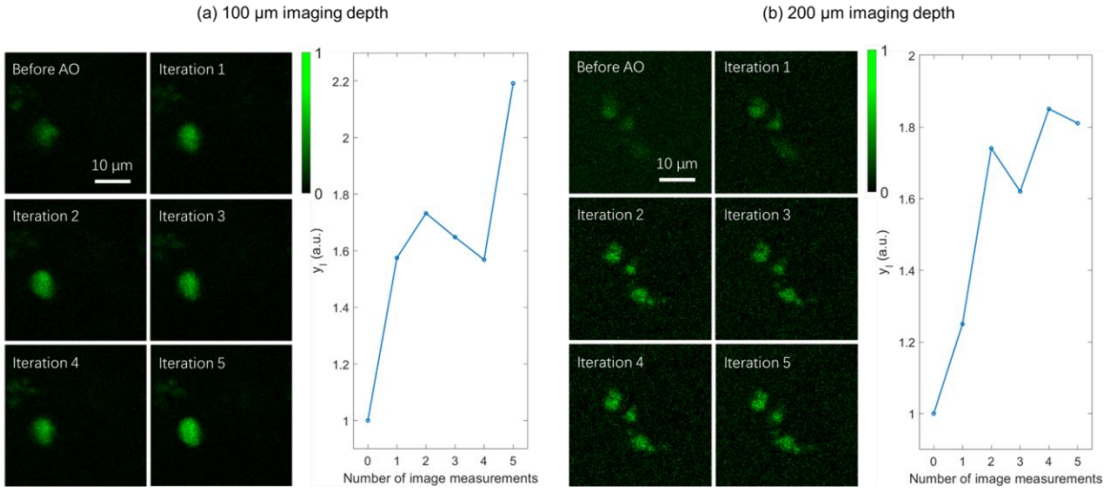


Figure 5-18: The experimental results of imaging two $40 \times 40\ \mu m$ FOVs respectively at 100 and 200 μm deep into fixed Thy1-YFP mouse brain tissue with an unknown sample induced ϕ_A corrected by the method using 6 spots for five AO iterations. The images taken before AO and after each AO iterations were shown for comparison. y_I was normalised such that its value before AO equalled 1 and plotted against the number of images measured for AO.

5.10 Summary and discussion

In this chapter, a new sensor-less AO method for multiphoton fluorescence microscopy

was proposed to parallelise the sequential measurement of multiple biased images so that correction was made possible with even a single image measurement. In this method, a focal spot array was designed such as each spot contained a different bias and generated by holography for imaging during AO. The resultant image was a superposition of laterally offset biased sub-images and served as the input to a pre-trained NN to estimate the aberration for correction. By assuming the total aberration could be represented by $N = 5$ Zernike modes, the method was implemented with 2, $N + 1 = 6$ and $2N = 10$ spots and experimentally demonstrated on a 2-P microscope with a SLM imaging both fluorescent beads and biological structures with different signal levels, aberration sizes, structure sizes, structure sparsity levels and imaging depths. The results showed that the method worked in most scenarios though with different performance.

Among the three spot numbers, $N + 1 = 6$ spots brought the most robust performance across a range of scenarios. This can be interpreted in terms of the amount of aberration information in the image measured with the focal spot array, which depended on the biases that determined information encoding and the overlaps among sub-images that caused information loss. As demonstrated in Section 4.7, the information encoding of $N + 1$ optimised biases was much better than that of 2 optimised biases and only slightly worse than that of $2N$ optimised biases. In the meantime, the potential rate of overlaps increased proportionally with the spot number. Thus, $N + 1$ spots generally resulted in the most aberration information. When N

becomes large, it can be difficult to generate $N+1$ spots. In this case, dominant modes can be used to approximately represent the aberration with a smaller N, or the $N+1$ spots can be separated into multiple spot arrays to be used sequentially for image measurement at a cost of AO speed.

Like most sensor-less AO methods, the method performance declined when the image SNR ratio decreased. However, as the total excitation intensity was divided into multiple spots in the method, for multiphoton fluorescence microscopy the overall signal level dropped as the spot number increased. To compensate for the signal drop and maintain the method performance, one can increase the excitation power, the scanning pixel dwell time or the gain of the photodetector. When there is a limit in signal amplification, a denoising step can be applied to the measured image.

Also, the method performed better with smaller object structures than larger ones. This is mainly due to the limited capacity of the specific NN in extracting aberration from images of extended objects. For better performance with large object structures, the NN can be implemented differently, for example by introducing a pre-processing step of calculating pseudo-PSFs from measured images [121], to reduce the influence from object structures.

The method performance was sensitive to the sparsity level of object structures as well. A high sparsity level meant a low signal level in the measured image as there were only a small number of structures in the FOV, while a low sparsity level tended to cause great information loss from overlaps among sub-images. Thus, a medium sparsity level

provided the best trade-off between the signal level and the chance of overlaps, which naturally led to the best performance.

In general, the method could correct a wide range of aberrations to less than 1 rad. For larger aberrations, although more aberration tended to be corrected in one AO iteration, more iterations were typically required to complete the correction. Apart from the factors discussed above, such performance also highly depended on the discrepancies between the focal spot arrays designed and actually generated in experiments, which could originate from the error in calculating the hologram, the display error of the SLM and the misalignment of the experimental system. Another important influencing factor was how well the synthetic NN training data represented the real practice, which was largely decided by the accuracy of the image formation model used for numerical simulation and the coverage of possible scenarios in experiments.

Although the method could not always achieve its best performance, it still has great potential for much faster sensor-less AO, which is particularly important for the situations where real-time correction is required to deal with fast-changing aberrations. Moreover, the main idea of the method can be easily integrated into many other AO methods to accelerate the process and extended to imaging modalities other than multiphoton fluorescence microscopy.

Chapter 6

Conclusion and outlook

6.1 Conclusion

In this thesis, it was explained how more efficient sensor-less AO could be achieved by optimising the information contained in the measured images to better estimate and correct the aberrations, thus increasing imaging performance of multiphoton fluorescence microscopy.

First, a framework was introduced to optimise the biases so that more information about the aberration was encoded in the images. In the framework, an analysis of the information contained in the images was realised by calculating the Fisher information with a numerical model; Based on the Fisher information, a CRLB based loss function was defined to guide the bias optimisation objectively for more aberration information encoded in the images; An automatic approach inspired by NN training was provided to implement the bias optimisation. The framework was demonstrated to work with different sensor-less AO methods and imaging scenarios. The results showed that the optimised biases always led to more accurate aberration estimation compared to biases conventionally defined based on empirical experience and the loss function served as a reliable predictor of the estimation accuracy. The demonstration also showed that the bias optimisation brought more improvement to the estimation when more degrees of freedom were allowed in the biases. In addition, useful features of the optimised biases

were assessed. The framework has the potential to allow existing sensor-less AO methods that used heuristically determined biases to estimate the aberration more efficiently with biases optimised. Also, the framework and the optimised biases presented in the demonstration can potentially guide the design of the biases in new sensor-less AO methods to come.

Second, a new sensor-less AO method was proposed to parallelise the sequential image measurement so that information used to separate in multiple images was concentrated into one image. In this method, a focal spot array was designed such that each spot contained a different bias and the corresponding hologram was calculated by GS algorithm; The hologram was added on the adaptive element to generate the focal spot array for imaging during AO so that each measured image was the superposition of multiple laterally offset biased sub-images; A pre-trained NN was applied to estimate the aberration from a single image measurement for correction. The method was experimentally demonstrated on a 2-P microscope with a SLM. The results showed that the method enabled aberration correction based on even a single image measurement and remained effective in a wide range of scenarios. For correction of N aberration modes, it was demonstrated that $N+1$ spots generally brought the most robust performance. Also, the change of method performance was demonstrated with respect to the image signal level, aberration size, object structure size, object structure sparsity levels and imaging depths. The method has great potential to significantly increase the speed of sensor-less AO, which is important to correct fast-changing aberrations.

Inspirations drawn from the method can also help to accelerate other sensor-less AO methods.

Despite the fact that the optimised sensor-less AO was developed and demonstrated for multiphoton fluorescence microscopy, it can be easily extended to other imaging modalities and applications (such as those in free-space laser communications) for wider benefits.

6.2 Outlook

This research is expected to be further developed in the future. In the framework for bias optimisation, the Fisher information calculation relied on an image formation model. This limited the application of the framework when the image formation process was not well understood and caused computational difficulties when there were a large number of parameters in the model. To improve on this, a generative NN similar to the one in [117] can be trained to replace the image formation model so the full understanding of the image formation process will be no longer required. When the Fisher information is calculated based on the generative NN, the partial derivatives can be conveniently calculated by backward propagation. Also, a much more concise parameterisation can be obtained by restricting the number of parameters in the NN input, which will make sure the full Fisher information matrix can be calculated for more accurate information analysis.

Another plan is to test how well the biases optimised by the framework can be generalised to scenarios that the optimisation did not take into account, and even used

for a different sensor-less AO method. This test is meaningful because the scenario can be hard to predict and model accurately in real practice and it can be difficult to apply the framework to certain methods without strong approximations. The test results will then suggest whether the optimised biases can still lead to good performance even if there is a potential mismatch between the actual scenarios and the scenarios defined for the bias optimisation and how confidently the optimised biases can be shared among different methods.

In the new sensor-less AO method, a classic NN implementation was adopted while the superimposed input image posed a new challenge for the NN to estimate the aberration. Hence, the NN implementation will be updated to better address the challenge. A good way is to specially train the initial convolutional layers of the NN to separate the sub-images into different feature maps. To achieve this, dilated convolution should be used in these layers to expand the receptive field of each neuron [160,161]. Also, the focal spot array can be fed into the NN as an extra input to guide the training. The separation of the sub-images will further allow the pseudo-PSFs to be calculated from each pair of separated sub-images, which can help the NN overcome the influence from object structures in the following aberration estimation [121].

Moreover, the new sensor-less AO method was demonstrated by correcting low-order aberrations in five Zernike modes, which was sufficient to prove the principle but did not extend to showing the method performance in typical biological imaging scenarios where more complex aberrations may exist. Therefore, follow-up

demonstrations will be carried out with the method correcting aberrations in a larger number of modes. As more high-order aberrations are involved, Zernike modes may no longer serve as a suitable choice of aberration representation. The method is thus expected to try different aberration representations in the future demonstrations. To demonstrate the advantage of the method in speed, the time consumed for AO will be compared between the method and other methods. The performance of correcting fast-changing aberrations when imaging living tissue or in vivo will also be tested for comparison.

In the long run, both the framework for bias optimisation and the new sensor-less AO method will be adapted to work with other types of imaging systems and even other types of applications that benefit from sensor-less AO. To enable a wider and more convenient use of the framework for bias optimisation, a toolkit is expected to be developed based on the framework and shared so that other researchers can try out and take advantage of this powerful tool. The hardware requirement on the user side can be minimised by linking the toolkit to Google Colab, which is a Python development environment that runs in the browser using Google Cloud and offers free access to their GPU usage. The toolkit will also allow the optimised biases to be shared among users for quick reference.

Bibliography

[1] M. W. Davidson and M. Abramowitz. “Optical microscopy.” Encyclopedia of Imaging Science and Technology 2(120): 1106-1141, 2002.

[2] B. Herman and J. J. Lemasters. Optical microscopy: emerging methods and applications (Eds.). Academic Press, New York, 2012.

[3] J. Mertz. Introduction to optical microscopy. Cambridge University Press, 2019.

[4] M Renz. “Fluorescence microscopy-A historical and technical perspective.” Cytometry Part A, 83(9):767-779, 2003.

[5] J. W. Lichtman and J. A. Conchello. “Fluorescence microscopy.” Nature Methods, 2(12): 910-919, 2005.

[6] E. C. Jensen. “Types of imaging, Part 2: An overview of fluorescence microscopy.” The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, 295(10):1621-1627, 2012.

[7] F. Helmchen and W. Denk. “New developments in multiphoton microscopy.” Current Opinion in Neurobiology, 12(5):593-601, 2002.

[8] W. R. Zipfel et al. “Nonlinear magic: multiphoton microscopy in the biosciences.” Nature Biotechnology 21(11): 1369-1377, 2003.

[9] M. J. Booth. “Adaptive optics in microscopy.” Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365(1861): 2829-2843, 2007.

[10] M. J. Booth. “Adaptive optical microscopy: the ongoing quest for a perfect image.” Light: Science & Applications 3(4): e165, 2014.

Booth, Martin J. "Adaptive optical microscopy: the ongoing quest for a perfect image." Light: Science & Applications 3.4 (2014): e165-e165.

[11] N. Ji. “Adaptive optical fluorescence microscopy.” Nature Methods 14(4): 374-380, 2017.

[12] K. M. Hampson et al. “Adaptive optics for high-resolution imaging.” Nature Reviews Methods Primers 1(1): 68, 2021.

[13] Q. Zhang et al. “Adaptive optics for optical microscopy.” Biomedical Optics Express 14(4): 1732-1756, 2023.

[14] Q. Hu et al. “A universal framework for microscope sensorless adaptive optics: Generalized aberration representations.” APL Photonics 5(10), 2020.

[15] M. Born and E. Wolf. Principles of optics: electromagnetic theory of propagation, interference and diffraction of light. Cambridge University Press, 1999.

[16] E. Hecht. Optics. Pearson, 2017.

[17] M. O. Scully and M. S. Zubairy. Quantum optics. Cambridge University Press, 1997.

[18] “File:EM spectrum.svg.” Wikimedia Commons, 2025.

[19] P. A. Buser and M. Imbert. Vision. MIT Press, 1992.

[20] J. W. Goodman. Introduction to Fourier optics. Roberts and Company Publishers, 2005.

[21] M. Gu. Principles of three dimensional imaging in confocal microscopes. World Scientific, 1996.

[22] M. J. Booth. Microscope Resolution Estimation and Normalised Coordinates. Zenodo, 2020.

[23] K. Uluç, G. C. Kujoth and M. K. Başkaya. “Operating microscopes: past, present, and future.” Neurosurgical Focus 27(3): E4, 2009.

[24] C. R. Burch and J. P. P. Stock. “Phase-contrast microscopy.” Journal of Scientific Instruments 19(5): 71, 1942.

[25] A. Barty et al. “Quantitative optical phase microscopy.” Optics Letters 23(11): 817-819, 1998.

[26] C. J. Mann et al. “High-resolution quantitative phase-contrast microscopy by digital holography.” Optics Express 13(22): 8693-8698, 2005.

[27] G. Popescu et al. “Diffraction phase microscopy for quantifying cell structure and dynamics.” Optics Letters 31(6): 775-777, 2006.

[28] W. Choi et al. “Tomographic phase microscopy.” Nature Methods 4(9): 717-719, 2007.

[29] M. Chen, L. Tian and L. Walker. “3D differential phase contrast microscopy.” Biomedical Optics Express 7(10): 3940-3950, 2016.

[30] S. S. Sahl, S. W. hell and S. Jakobs. “Fluorescence nanoscopy in cell biology.” Nature Reviews Molecular Cell Biology 18(11): 685-701, 2017.

[31] P. A. Franken et al. “Generation of optical harmonics.” Physical Review Letters

7(4): 118-119, 1961.

[32] J. A. Squier, et al. “Third harmonic generation microscopy.” Optics Express 3(9): 315-324, 1998.

[33] P. Friedl et al. “Biological second and third harmonic generation microscopy.” Current Protocols in Cell Biology 34(1): 4-15, 2007.

[34] X. Chen et al. “Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure.” Nature Protocols 7(4): 654-669, 2012.

[35] C. J. R. Sheppard et al. “Signal-to-noise ratio in confocal microscopes.” Handbook of biological confocal microscopy (pp. 442-452), Boston, MA: Springer US, 2006.

[36] G. Y. Wiederschain. “The Molecular Probes handbook. A guide to fluorescent probes and labeling technologies.” Biochemistry (Moscow) 76(11): 1276-1277, 2011.

[37] K. Aswani, T. Jinadasa and C. M. Brown. “Fluorescence microscopy light sources.” Microscopy Today 20(4): 22-28, 2012.

[38] V. Magidson and A. Khodjakov. “Circumventing photodamage in live-cell microscopy.” Methods in Cell Biology 114: 545-560, 2013.

[39] A. Diaspro et al. “Photobleaching.” Handbook of biological confocal microscopy (pp. 690-702), Boston, MA: Springer US, 2006.

[40] J. Icha et al. “Phototoxicity in live fluorescence microscopy, and how to avoid it.” BioEssays 39(8): 1700003, 2017.

[41] M. Minsky. “Memoir on inventing the confocal scanning microscope.” Scanning 10(4):128-138, 1988.

[42] J. Pawley (Ed.) Confocal microscopy (vol. 426). London: Academic press, 1990.

[43] J. A. Conchello and J. W. Lichtman. “Optical sectioning microscopy.” Nature Methods 2(12): 920-931, 2005.

[44] J. Mertz. “Nonlinear microscopy: new techniques and applications.” Current Opinion in Neurobiology 14(5): 610-616, 2004.

[45] R. W. Boyd et al. “Nonlinear optics.” Springer Handbook of Atomic, Molecular, and Optical Physics (pp. 1097-1110). Cham: Springer International Publishing, 2008.

[46] W. Denk, J. H. Strickler and W. W. Webb. “Two-photon laser scanning fluorescence microscopy.” Science 248(4951): 73-76, 1990.

[47] F. Helmchen and W. Denk. “Deep tissue two-photon microscopy.” Nature Methods, 2(12):932-940, 2005.

[48] K. Svoboda and R. Yasuda. “Principles of two-photon excitation microscopy and its applications to neuroscience.” Neuron 50(6): 823-839, 2006.

[49] C. Xu et al. “Multiphoton fluorescence excitation: new spectral windows for biological nonlinear microscopy.” Proceedings of the National Academy of Sciences 93(20): 10763-10768, 1996.

[50] N. G. Horton et al. “In vivo three-photon microscopy of subcortical structures within an intact mouse brain.” Nature Photonics 7(3): 205-209, 2013.

[51] T. Wang and C. Xu. “Three-photon neuronal imaging in deep mouse brain.” Optica 7(8): 947-960, 2020.

[52] M. Oheim et al. “Two-photon microscopy in brain tissue: parameters influencing

the imaging depth." Journal of Neuroscience Methods 111(1): 29-37, 2001.

[53] D. Kobat et al. "Deep tissue multiphoton microscopy using longer wavelength excitation." Optics Express 17(16): 13354-13364, 2009.

[54] K. Svoboda and S. M. Block. "Biological applications of optical forces." Annual Review of Biophysics and Biomolecular Structure 23(1): 247-285, 1994.

[55] J. M. Squirrell et al. "Long-term two-photon fluorescence imaging of mammalian embryos without compromising viability." Nature Biotechnology 17(8): 763-767, 1999.

[56] B. Wang and M. J. Booth. "Optimum deformable mirror modes for sensorless adaptive optics." Optics Communications 282(23): 4467-4474, 2009.

[57] M. Schwertner et al. "Measurement of specimen-induced aberrations of biological samples using phase stepping interferometry." Journal of Microscopy 213(1): 11-19, 2004.

[58] R. J. Noll. "Zernike polynomials and atmospheric turbulence." Journal of the Optical Society of America 66(3): 207-211, 1976.

[59] M. Schwertner et al. "Characterizing specimen induced aberrations for high NA adaptive optical microscopy." Optics Express 12(26): 6540-6552, 2004.

[60] R. K. Tyson and B. W. Frazier. Principles of adaptive optics. CRC press, 2022.

[61] M. A. Ealey and J. Wellman. "Fundamentals of deformable mirror design and analysis." Precision Engineering and Optomechanics (vol. 1167). SPIE, 1989.

[62] D. J. Dagel et al. "Large-stroke MEMS deformable mirrors for adaptive optics." Journal of Microelectromechanical Systems 15(3): 572-583, 2006.

[63] T. Bifano. “MEMS deformable mirrors.” Nature Photonics 5(1): 21-23, 2011.

[64] M. Booth et al. “Methods for the characterization of deformable membrane mirrors.” Applied Optics 44(24): 5131-5139, 2005.

[65] Z. Zhang et al. “Fundamentals of phase-only liquid crystal on silicon (LCOS) devices.” Light: Science & Applications 3(10): e213, 2014.

[66] C. Maurer et al. “What spatial light modulators can do for optical microscopy.” Laser & Photonics Reviews 5(1): 81-101, 2011.

[67] B. C. Platt and R. Shack. “History and principles of Shack-Hartmann wavefront sensing.” Journal of Refractive Surgery 17(5): S573-577, 2001.

[68] J. W. Cha, J. Ballesta and P. T. C. So. “Shack-Hartmann wavefront-sensor-based adaptive optics system for multiphoton microscopy.” Journal of Biomedical Optics 15(4): 046022-046022, 2010.

[69] X. Tao et al. “Adaptive optics microscopy with direct wavefront sensing using fluorescent protein guide stars.” Optics Letters, 36(17):3389-3391, 2011.

[70] K. Wang et al. “Rapid adaptive optical recovery of optimal resolution over large volumes.” Nature Methods 11(6): 625-628, 2014.

[71] M. R. Teague. “Deterministic phase retrieval: a Green’s function solution.” Journal of the Optical Society of America, 73(11): 1434-1441, 1983.

[72] R. A. Gonsalves. “Phase retrieval and diversity in adaptive optics.” Optical Engineering 21(5): 829-832, 1982.

[73] R. G. Paxman, T. J. Schulz and J. R. Fienup. “Joint estimation of object and

aberrations by using phase diversity." Journal of the Optical Society of America A 9(7): 1072-1085, 1992.

[74] T. M. Cover and J. A. Thomas. Elements of Information Theory (2nd Edition). Wiley-Interscience, 2006.

[75] C. E. Shannon. "A mathematical theory of communication." The Bell System Technical Journal 27(3): 379-423, 1948.

[76] C. M. Bishop. Pattern recognition and machine learning. Springer, 2006.

[77] Y. Bengio, I. Goodfellow and A. Courville. Deep learning. MIT press, 2017.

[78] K. Arulkumaran et al. "Deep reinforcement learning: A brief survey." IEEE Signal Processing Magazine 34(6): 26-38, 2017.

[79] M. D. Zeiler and R. Fergus. "Visualizing and understanding convolutional networks." European Conference on Computer Vision (pp. 818-833), Cham: Springer International Publishing, 2014.

[80] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition." 3rd International Conference on Learning Representations, Computational and Biological Learning Society, 2015.

[81] R. K. Srivastava, K. Greff, and J. Schmidhuber. "Training very deep networks." Advances in Neural Information Processing Systems 28, 2015.

[82] Y. Bengio, P. Simard and P. Frasconi. "Learning long-term dependencies with gradient descent is difficult." IEEE Transactions on Neural Networks 5(2): 157-166, 1994.

[83] X. Glorot, and Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks.” Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (pp. 249-256), JMLR Workshop and Conference Proceedings, 2010.

[84] K. He et al. “Deep residual learning for image recognition.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[85] V. Nair and G. E. Hinton. “Rectified linear units improve restricted boltzmann machines.” Proceedings of the 27th International Conference on Machine Learning (pp. 807-814), 2010.

[86] S. Ioffe and C. Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift.” International Conference on Machine Learning (pp. 448-456), 2015.

[87] P. N. Marsh, D. Burns and J. M. Girkin. “Practical implementation of adaptive optics in multiphoton microscopy.” Optics Express 11(10): 1123-1130, 2003.

[88] A. J. Wright et al. “Exploration of the optimisation algorithms used in the implementation of adaptive optics in confocal and multiphoton microscopy.” Microscopy Research and Technique 67(1): 36-44, 2005.

[89] S. P. Poland, A. J. Wright and J. M. Girkin. “Evaluation of fitness parameters used in an iterative approach to aberration correction in optical sectioning microscopy.” Applied Optics 47(6): 731-736, 2008.

[90] G. Vdovin, P. M. Sarro and S. Middelhoek. “Technology and applications of

micromachined adaptive mirrors." Journal of Micromechanics and Microengineering 9(2): R8, 1999.

[91] L. Huang, C. Rao and W. Jiang. "Modified Gaussian influence function of deformable mirror actuators." Optics Express 16(1): 108-114, 2008.

[92] M. B. Roopashree, A. Vyas and B. R. Prasad. "A novel model of influence function: calibration of a continuous membrane deformable mirror." International Journal on Control System and Instrumentation 3(2): 10-14, 2012.

[93] L. Sherman et al. "Adaptive correction of depth - induced aberrations in multiphoton scanning microscopy using a deformable mirror." Journal of Microscopy 206(1): 65-71, 2002.

[94] L. Zhu et al. "Wave-front generation of Zernike polynomial modes with a micromachined membrane deformable mirror." Applied Optics 38(28): 6019-6026, 1999.

[95] M. J. Booth et al. "Adaptive aberration correction in a confocal microscope." Proceedings of the National Academy of Sciences 99(9): 5788-5792, 2002.

[96] T. Ota et al. "Dynamic axial-position control of a laser-trapped particle by wave-front modification." Optics Letters 28(6): 465-467, 2003.

[97] M. J. Booth, M. A. A. Neil and T. Wilson. "New modal wave-front sensor: application to adaptive confocal fluorescence microscopy and two-photon excitation fluorescence microscopy." Journal of the Optical Society of America A 19(10): 2112-2120, 2002.

[98] M. J. Booth. “Direct measurement of Zernike aberration modes with a modal wavefront sensor.” Advanced Wavefront Control: Methods, Devices, and Applications (vol. 5162). SPIE, 2003.

[99] M. J. Booth. “Wave front sensor-less adaptive optics: a model-based approach using sphere packings.” Optics Express 14(4): 1339-1352, 2006.

[100] D. Débarre et al. “Image-based adaptive optics for two-photon microscopy.” Optics Letters 34(16): 2495-2497, 2009.

[101] M. Warber et al. “Combination of scene-based and stochastic measurement for wide-field aberration correction in microscopic imaging.” Applied Optics 49(28): 5474-5479, 2010.

[102] W. M. Lee and S. H. Yun. “Adaptive aberration correction of GRIN lenses for confocal endomicroscopy.” Optics Letters 36(23): 4608-4610, 2011.

[103] J. Antonello et al. “Semidefinite programming for model-based sensorless adaptive optics.” Journal of the Optical Society of America A 29(11), 2428-2438, 2012.

[104] A. Facomprez, E. Beaurepaire and D. Débarre. “Accuracy of correction in modal sensorless adaptive optics.” Optics Express 20(3): 2598-2612, 2012.

[105] P. Kner. “Phase diversity for three-dimensional imaging.” Journal of the Optical Society of America A 30(10): 1980-1987, 2013.

[106] P. T. Galwaduge et al. “Simple wavefront correction framework for two-photon microscopy of in-vivo brain.” Biomedical Optics Express 6(8): 2997-3013, 2015.

[107] M. Skorsetz, P. Artal and J. M. Bueno. “Performance evaluation of a sensorless

adaptive optics multiphoton microscope." Journal of Microscopy 261(3): 249-258, 2016.

[108] Y. Jin et al. "Machine learning guided rapid focusing with sensor-less aberration corrections." Optics Express 26(23): 30162-30171, 2018.

[109] S. W. Paine and J. R. Fienup. "Machine learning for improved image-based wavefront sensing." Optics Letters 43(6): 1235-1238, 2018.

[110] Y. Nishizaki et al. "Deep learning wavefront sensing." Optics Express 27(1): 240-251, 2019.

[111] L. Möckl, P. N. Petrov and W. E. Moerner. "Accurate phase retrieval of complex 3D point spread functions with deep residual neural networks." Applied Physics Letters 115(25), 2019.

[112] B. P. Cumming and Min Gu. "Direct determination of aberration functions in microscopy by an artificial neural network." Optics Express 28(10): 14511-14521, 2020.

[113] I. Vishniakou and J. D. Seelig. "Wavefront correction for adaptive optics with reflected light and deep neural networks." Optics Express 28(10): 15459-15471, 2020.

[114] J. Antonello et al. "Multi-scale sensorless adaptive optics: application to stimulated emission depletion microscopy." Optics Express 28(11): 16749-16763, 2020.

[115] G. Allan et al. "Deep residual learning for low-order wavefront sensing in high-contrast imaging systems." Optics Express 28(18): 26267-26283, 2020.

[116] D. Saha et al. "Practical sensorless aberration estimation for 3D microscopy with deep learning." Optics Express 28(20): 29044-29053, 2020.

[117] E. Bostan et al. “Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network.” Optica 7(6): 559-562, 2020.

[118] L. Streich et al. “High-resolution structural and functional deep brain imaging using adaptive optics three-photon microscopy.” Nature Methods 18(10): 1253-1258, 2021.

[119] E. Durech et al. “Wavefront sensor-less adaptive optics using deep reinforcement learning.” Biomedical Optics Express 12(9): 5423-5438, 2021.

[120] P. A. Khorin et al. “Neural networks application to determine the types and magnitude of aberrations from the pattern of the point spread function out of the focal plane.” Journal of Physics: Conference Series (vol. 2086, no. 1), IOP Publishing, 2021.

[121] Q. Hu et al. “Universal adaptive optics for microscopy through embedded neural network control.” Light: Science & Applications 12(1): 270, 2023.

[122] G. Xu et al. “Image metric-based multi-observation single-step deep deterministic policy gradient for sensorless adaptive optics.” Biomedical Optics Express 15(8): 4795-4814, 2024.

[123] C. Johnson et al. “Phase-diversity-based wavefront sensing for fluorescence microscopy.” Optica 11(6): 806-820, 2024.

[124] I. Kang et al. “Coordinate-based neural representations for computational adaptive optics in widefield microscopy.” Nature Machine Intelligence 6(6): 714-725, 2024.

[125] Y. E. Kok et al. “Practical aberration correction using deep transfer learning with

limited experimental data." Optics Express 33(6): 14431-14444, 2025.

[126] J. Wang et al. "Deep learning-driven adaptive optics for laser wavefront correction." Applied Optics 64(29): 8625-8632, 2025.

[127] F. Bortoletto et al. "Multiphoton fluorescence microscopy with GRIN objective aberration correction by low order adaptive optics." PloS One 6(7): e22321, 2011.

[128] M. J. Booth. "Wavefront sensorless adaptive optics for large aberrations." Optics Letters 32(1): 5-7, 2006.

[129] D. Débarre, M. J. Booth and T. Wilson. "Image based adaptive optics through optimisation of low spatial frequencies." Optics Express 15(13): 8176-8190, 2007.

[130] F. Wang. "Wavefront sensing through measurements of binary aberration modes." Applied Optics 48(15): 2865-2870, 2009.

[131] J. Tang, R. N. Germain and M. Cui. "Superpenetration optical microscopy by iterative multiphoton adaptive compensation technique." Proceedings of the National Academy of Sciences 109(22): 8434-8439, 2012.

[132] D. B. Conkey, A. M. Caravaca-Aguirre and R. Piestun. "High-speed scattering medium characterization with application to focusing light through turbid media." Optics Express 20(2): 1733-1740, 2012.

[133] C. Stockbridge et al. "Focusing through dynamic scattering media." Optics Express 20(14): 15086-15092, 2012.

[134] R. Fiolka, K. Si and M. Cui. "Complex wavefront corrections for deep tissue focusing using low coherence backscattered light." Optics Express 20(15): 16532-

16543, 2012.

[135] O. Katz et al. “Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers.” Optica 1(3): 170-174, 2014.

[136] L. Kong and M. Cui. “In vivo neuroimaging through the highly scattering tissue via iterative multi-photon adaptive compensation technique.” Optics Express 23(5): 6145-6150, 2015.

[137] X. Tao et al. “High-speed scanning interferometric focusing by fast measurement of binary transmission matrix for channel demixing.” Optics Express 23(11): 14168-14187, 2015.

[138] D. Sinefeld et al. “Adaptive optics in multiphoton microscopy: comparison of two, three and four photon fluorescence.” Optics Express 23(24): 31472-31483, 2015.

[139] J. H. Park, W. Sun and M. Cui. “High-resolution in vivo imaging of mouse brain through the intact skull.” Proceedings of the National Academy of Sciences, 112(30): 9236-9241, 2015.

[140] X. Tao et al. “Three-dimensional focusing through scattering media using conjugate adaptive optics with remote focusing (CAORF).” Optics Express 25(9): 10368-10383, 2017.

[141] Q. Hu et al. “The Lattice Geometry of Walsh-Function-Based Adaptive Optics.” Photonics 9(8), MDPI, 2022.

[142] N. Ji, D. E. Milkie and E. Betzig. “Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues.” Nature Methods 7(2): 141-147, 2010.

[143] D. E. Milkie, N. Ji and E. Betzig. “Pupil-segmentation-based adaptive optical microscopy with full-pupil illumination.” Optics Letters 36(21): 4206-4208, 2011.

[144] N. Ji, T. R. Sato and E. Betzig. “Characterization and adaptive optical correction of aberrations during in vivo imaging in the mouse cortex.” Proceedings of the National Academy of Sciences 109(1): 22-27, 2012.

[145] C. Wang and N. Ji. “Pupil-segmentation-based adaptive optical correction of a high-numerical-aperture gradient refractive index lens for two-photon fluorescence endoscopy.” Optics Letters 37(11): 2001-2003, 2012.

[146] C. Wang and N. Ji. “Characterization and improvement of three-dimensional imaging performance of GRIN-lens-based two-photon fluorescence endomicroscopes with adaptive optics.” Optics Express 21(22): 27142-27154, 2013.

[147] R. Liu et al. “Direct phase measurement in zonal wavefront reconstruction using multidither coherent optical adaptive technique.” Optics Express 22(2): 1619-1628, 2014.

[148] C. Wang et al. “Multiplexed aberration measurement for deep tissue imaging in vivo.” Nature Methods 11(10): 1037-1040, 2014.

[149] C. Rodríguez et al. “An adaptive optics module for deep tissue multiphoton imaging in vivo.” Nature Methods 18(10): 1259-1264, 2021.

[150] O. Soloviev. “Alias-free basis for modal sensorless adaptive optics using the second moment of intensity.” International Journal of Wavelets, Multiresolution and Information Processing 20(3): 2040008, 2022.

[151] D. Kinga and J. Ba. “Adam: A method for stochastic optimization.” International Conference on Learning Representations (vol. 5, no. 6), 2015.

[152] R. Von Mises. Mathematical theory of probability and statistics. Academic Press, 2014.

[153] G. Marsaglia. “Choosing a point from the surface of a sphere.” The Annals of Mathematical Statistics 43(2): 645-646, 1972.

[154] X. Glorot and Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks.” Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, 2010.

[155] E. Sahin et al. “Computer-generated holograms for 3D imaging: a survey.” ACM Computing Surveys (CSUR) 53(2): 1-35, 2020.

[156] R. W. Gerchberg and W. O. Saxton. “A practical algorithm for the determination of the phase from image and diffraction plane pictures.” Optik 35(2): 237–246, 1972.

[157] P. Memmolo et al. “Investigation on specific solutions of Gerchberg–Saxton algorithm.” Optics and Lasers in Engineering 52: 206-211, 2014.

[158] S Liu and Y. Takaki. “Optimization of phase-only computer-generated holograms based on the gradient descent method.” Applied Sciences 10(12): 4283, 2020.

[159] G. Huang et al. “Generalizing the Gerchberg–Saxton algorithm for retrieving complex optical transmission matrices.” Photonics Research 9(1): 34-42, 2021.

[160] F. Yu and V Koltun. “Multi-scale context aggregation by dilated convolutions.”

arXiv preprint arXiv:1511.07122, 2015.

[161] L. C. Chen et al. “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs.” IEEE Transactions on Pattern Analysis and Machine Intelligence 40(4): 834-848, 2017.

List of publications

1. Biwei Zhang, Qi Hu, Jingyu Wang and Martin J. Booth. “Single image sensorless adaptive optics for microscopy” arXiv preprint arXiv:2509.17869, 2025.
2. Biwei Zhang, Martin J. Booth and Qi Hu. “Information-guided optimization of image-based sensorless adaptive optics methods” arXiv preprint arXiv:2506.07482, 2025.

Acknowledgement

The completion of this thesis and my DPhil studies is credited to not only my own efforts but also the support and love that I have received from many people throughout the journey.

First and foremost, I would like to thank my supervisor Prof. Martin Booth for all his professional guidance, great patience and quick responses between his lots of duties. I would also like to thank Dr. Qi Hu for all the constructive suggestions and valuable experience from those many long discussions, and Dr. Jingyu Wang for his technical expertise in microscopy.

My thanks also go to all my friends from the DOP group and SMP group. To Dr. Linpei Xue and Dr. Yuyao Xiao, who were also my wonderful flatmates, I will always remember how you tolerated my flaws and brightened up my life. And to Dr. Bohan Chen, Prof. Andong Wang, Dr. Mengmeng Li, Dr. Mohan Wang, Dr. Zipei Song, Dr. Eleanor Warrington, Dr. Xuke Qiu and Dr. Tongyu Liu, thanks for your kindness and assistance.

I am also grateful to all my lovely college buddies from Mansfield College. Dr. Liz Wan, it is such an amazing surprise to forge this deep bond between us and have you there by my side, and I have truly learnt a lot from you. Ji Ma, I really miss all the days and nights hanging out together. And to Ellen Yang, Dr. Alex Xu, Karim Alaa El-Din, Advait Tambe, Dr. Jonathan Young, Jonathan Seiz de Filippi, Adam Bernstein, Yiping

Miao, Ming Sheng, Dr. Juan Rafael Alvarez, Alexandra Roden, Dr. Amirmohammad Farzaneh, Ryan Chia, Jaka Pelaic, Joshua Grimond, Dr. Alexander Cliffe, Samo Novak, Jude Smith, Dr. Christopher Williams, Constance Hardesty, Dr. Robin Scales, Eunju Choe, Johanna Barop and many more, thanks for being friendly to me and all the good memory we share.

I am also lucky to meet my other friends in Oxford. They are Dr. Fengning Yang, for all those gym and tennis sessions, Dr. Feng Liu, all start from the London trip, Dr. Volkan Kumtepeli and Dr. Zihao Zhou, from lab demonstrators to more, Dr. Andy Fletcher and Dr. Edward Shellard, love from Keble+, Dr. Tongshan Liu and Xiaohang Cai, boardgame pals, Dr. Justin Limkaichong, so many gym meetings, Dr. Zhaoyi Yan and Graham Higgins, the rain and the rowing, Dr. Li Wang, days in Castle Mill. Thanks for making my Oxford days colourful.

I would also like to thank Dr. Zehua Sheng. I am amazed how many years have passed and we are still in touch even though we have been apart for long. Thank you for keeping some light from old days for me when most things dated from then have faded away. Auld Lang Syne, my friend, hope we will keep it longer and longer.

Last but not least, my deepest love and gratitude go to my parents Yongsheng Wu and Zhile Zhang. I could have never started and got through this journey without your financial and emotional support. I want to especially send my immense love and gratitude to my mother Yongsheng Wu, who is always the closest person to me and always ready to support me selflessly in every possible way. All the best wishes and

blessings to you and to us three!