Neural Networks For Phase Demodulation In Optical Interferometry

Jacob A. Black

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Electrical Engineering

Yizheng Zhu, Chair

Yunhui Zhu

Jia-Bin Huang

T. C. Poon

Ahmad Safaai-Jazi

July 22, 2019

Blacksburg, Virginia

Keywords: Phase imaging, Neural Networks, Machine Learning, Biological Imaging

Neural Networks For Phase Demodulation In Optical Interferometry

Jacob A. Black

(ABSTRACT)

Neural Networks (NNs) (or 'deep' neural networks (DNNs)) have found great success in many applications across all fields of engineering, and in particular have found recent success in the field of Photonics. In this work we discuss the application of NNs to optical interferometry for the purpose of quantitative phase imaging (QPI). We show that NNs are capable of quantifying the optical pathlength difference in an interferogram with sensitivities that achieve the fundamental limit given by the Cramér-Rao bound (CRB). As an application, we consider a particular QPI technique known as wavelength shifting interferometry (WSI) which obtains the OPL by acquiring multiple interferograms at different, evenly spaced wavenumbers. Traditional phase demodulation algorithms for WSI fail to reach the theoretical OPL sensitivity limit set by the CRB. We have designed NNs which are capable of achieving this bound across a wide range of OPL differences. The NNs are trained on simulated data, and then applied to experimental data. In both simulation and experiment, the NNs outperform the existing analytical demodulation techniques and provide highly sensitive signal demodulation in cases where the analytical approach fails. Thus, NNs provide better performance and more flexibility in the design and use of a WSI system. We expect that the techniques developed in this work can be extended to other two-beam interference based QPI system.

Neural Networks For Phase Demodulation In Optical Interferometry

Jacob A. Black

(GENERAL AUDIENCE ABSTRACT)

Neural Networks (NNs) (or 'deep' neural networks (DNNs)) have found great success in many applications across all fields of engineering, and in particular have found recent success in the field of Photonics. In this work we discuss the application of NNs to making so-called 'phase' images of biological cells and tissues (e.g. red blood cells, sperm cells). This is necessary for many biological samples which are transparent under traditional bright field microscopy. We show that NNs are capable of quantifying the phase of these samples to produce images with higher contrast than possible in a typical microscope image. As an example, we introduce a particular phase microscopy system and study the application of NNs to this system. We show that the NNs are capable of providing solutions for this phase in situations where existing analytical techniques fail. The NNs are also capable of making more precise calculations of the phase than the traditional algorithms in many situations where either technique could be used. Therefore, NNs can provide simultaneously higher performance and more flexibility when designing phase microscopy systems.

Dedication

To my parents, brothers Joshua and Jonathan, my sister Katelyn, and my fiancée Jessica.

Acknowledgments

I would like to acknowledge my advisor, Dr. Yizheng Zhu, for welcoming me into his group and for the guidance, support, and insight he has provided since I joined. With his help, we have written two high quality research papers based on this work (with a third being written), far exceeding my own expectations. I would also like to thank my lab mates Dr. Shichao Chen, and Joseph Thomas who both helped me get up to speed after joining the group, and directly supported me through discussions and constructive feedback as I developed the techniques described in this work. Much of what I have learned during my time with the group has been through them, and I cannot thank them enough. I would also like to thank my entire committee and Dr. Yunhui Zhu in particular for the fantastic feedback and discussions we've had about my research over the last year or so. I look forward to more insightful discussions and collaborations with my committee.

Finally, I would like to acknowledge and thank the Bradley Department of Electrical and Computer Engineering and the Bradley Endowment for the extremely generous fellowship which has provided funding for me during the last two years. With the extra support and freedom provided by this fellowship, I was able to find and carry out research I am truly passionate about.

Contents

List of Figures

Chapter 1

Introduction

Quantitative phase imaging (QPI) has found many applications in recent years in microscopy, and new techniques are constantly being developed [3, 4, 5, 6, 7]. QPI is distinct from other traditional microscopy techniques such as Zernike phase contrast microscopy, which is considered a *qualitative* technique. Qualitative imaging techniques such as these generally convert phase variations into intensity variations, typically via some form of wave interference. This is a popular modality of most commercial microscopes. Therefore, the output of a phase contrast microscope is an intensity image with enhanced contrast for transparent samples due to the phase variations being encoded into the intensity. In contrast, QPI is capable of directly calculating the phase. This phase, or optical pathlength (OPL), is then plotted to form the image itself. Figure 1.1 shows an example of a phase contrast image alongside a similar image produced by a QPI system. We are interested in the application of neural networks (NNs) to QPI, and whether or not NNs can provide enhanced image quality and sensitivity [8]. Prior to this analysis, we provide the necessary background and some examples to illustrate the goals of a QPI system and some of the ways in which we can evaluate a given system's performance. This includes some background on Wave Optics, Statistical Optics, and the particular QPI system we will apply our NNs to. We will also cover some of the basics of machine learning and our particular implementation.

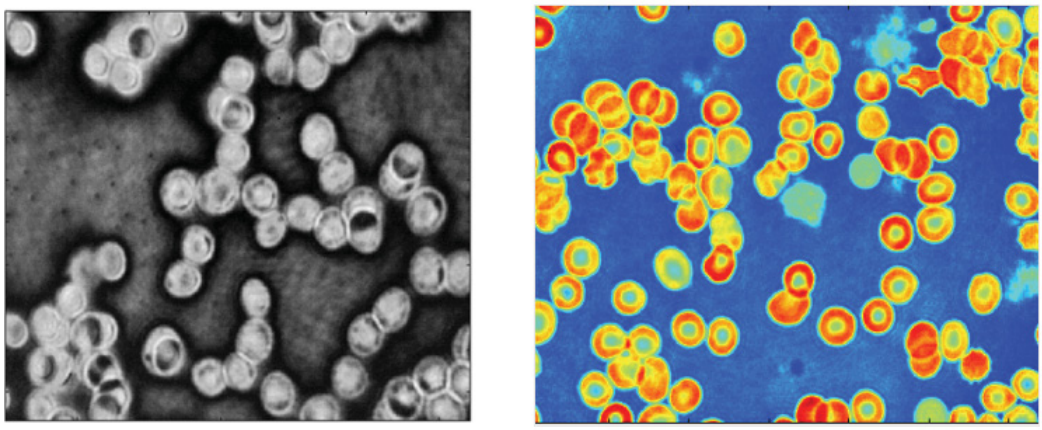

Figure 1.1: Left: A traditional phase contrast image of a red blood cell sample . Right: A QPI image of a similar red blood cell sample. The phase is plotted directly as a two dimensional map.

1.1 Electromagnetics and Wave Optics

1.1.1 Free Space Plane Waves

We begin by introducing the crown jewel of QPI: ***phase***. More specifically in QPI, we seek to solve for (quantify) the ***phase differences*** (more on this when we discuss wave optics) and represent them as an image of a given object (sample). We derive the concept of phase in the simplest case by deriving the plane wave solution of Maxwell's equations, and then demonstrate its importance in wave optics and how changes in phase can be plotted to represent an image. Although in essence the phase of interest is not fundamentally different from the phase ϕ of any general phasor $Ae^{j\phi}$, it is important to know what ϕ represents in

the context of QPI. The natural starting point is Maxwell's equations [9]:

$$\begin{aligned} \nabla \cdot (\epsilon \vec{\mathcal{E}}) &= \rho \\ \nabla \cdot (\mu \vec{\mathcal{H}}) &= 0 \\ \nabla \times \vec{\mathcal{E}} &= -\frac{\partial(\mu \vec{\mathcal{H}})}{\partial t} \\ \nabla \times \vec{\mathcal{H}} &= \frac{\partial(\epsilon \vec{\mathcal{E}})}{\partial t} + \vec{\mathcal{J}} \end{aligned} \tag{1.1}$$

where $\vec{\mathcal{J}}$ is the total electric current density in the region of interest, ϵ is the electric permittivity, μ is the magnetic permeability, and ρ is the charge density. For the purpose of illustrating the importance of phase, we can consider an isotropic, homogeneous (scalar, constant ϵ and μ) source free region of free space. This amounts to setting $\rho = 0$ and $\vec{\mathcal{J}} = 0$ in the above set of equations. This gives:

$$\begin{aligned} \nabla \cdot \vec{\mathcal{E}} &= 0 \\ \nabla \cdot \vec{\mathcal{H}} &= 0 \\ \nabla \times \vec{\mathcal{E}} &= -\mu_0 \frac{\partial \vec{\mathcal{H}}}{\partial t} \\ \nabla \times \vec{\mathcal{H}} &= \epsilon_0 \mu_0 \frac{\partial \vec{\mathcal{E}}}{\partial t} \end{aligned} \tag{1.2}$$

where we have simply divided out ϵ in Gauss's law. If we consider time-harmonic solutions of the form $\vec{\mathcal{X}}(\vec{r}, t) = \vec{X}_0(\vec{r}) \cos[\omega t + \alpha(\vec{r})] = \mathrm{Re}[\vec{X}(\vec{r}) e^{j\omega t}]$, where $\vec{\mathcal{X}}$ represents any of the vector quantities in Eqs. (1.1), $\vec{r} = (x, y, z)$ is a position vector, and $\vec{X}(\vec{r}) = \vec{X}_0 e^{j\alpha(\vec{r})}$ [10]. We can rewrite Eqs. (1.2) using the identity $\frac{\partial}{\partial t}(\vec{X}(\vec{r})) = j\omega \vec{X}(\vec{r})$ as:

$$\begin{aligned} \nabla \cdot \vec{\mathcal{E}} &= 0 \\ \nabla \cdot \vec{\mathcal{H}} &= 0 \\ \nabla \times \vec{\mathcal{E}} &= -j\omega\mu_0 \vec{\mathcal{H}} \\ \nabla \times \vec{\mathcal{H}} &= j\omega\epsilon_0 \vec{\mathcal{E}} \end{aligned} \tag{1.3}$$

Taking the curl of $\nabla \times \vec{\mathcal{E}}$, we obtain:

$$\nabla \times (\nabla \times \vec{\mathcal{E}}) = \nabla(\nabla \cdot \vec{\mathcal{E}}) + \nabla^2 \vec{\mathcal{E}} = -j\omega\mu_0 \nabla \times \vec{\mathcal{H}}$$

After substitution of $\nabla \times \vec{\mathcal{H}}$, we obtain:

$$\nabla^2 \vec{\mathcal{E}} + k^2 \vec{\mathcal{E}} = 0 \tag{1.4}$$

where $k^2 = \omega^2 \mu_0 \epsilon_0 = w^2/c^2$ are the freespace wavenumbers as a function of ω. Equation (1.4) is known as the vector Helmholtz equation, and an equivalent expression can also be obtained for $\vec{\mathcal{H}}$ (or $\vec{\mathcal{B}} = \mu\vec{\mathcal{H}}$) by instead beginning by taking the curl of Ampere's law and then making use of Faraday's law. One advantage of considering a source-free, free space solutions to Maxwell's equations for the purpose of illustration is that the Helmholtz equation becomes ***separable***. This means we can apply a technique known as ***separation of variables*** to each component of $\vec{\mathcal{E}}$ to obtain the solution. For example, for the x-component of the electric field, we have [10]:

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} + k^2 E_x = 0$$

in seperation of variables, we assume that E_x will be the product of three functions, each of which only depend on one spatial variable i.e. E_x will have the form:

$$E_x = X(x)Y(y)Z(z)$$

which gives:

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} + \frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} + k^2 = 0 \tag{1.5}$$

each of the partial derivative terms, being independent of one another, must be equal to some constant in order for their sum with k^2 to be zero e.g. :

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2} = -k_x^2, \quad \frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} = -k_y^2, \quad \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2} = -k_z^2$$

where $k_x^2 + k_y^2 + k_z^2 = k^2$. Each of these equations has a solution of the form $A_i e^{\pm k_i i}$, where $i = x, y, z$. We can write down an expression for E_x:

$$E_x = Ae^{\pm k_x x \pm k_y y \pm k_z z}$$

This process can be repeated to obtain each component of $\vec{\mathcal{E}}$, giving us the final solution for the complex amplitude $\vec{E(\vec{r})}$ for a linearly polarized monochromatic (single frequency or wavelength) plane wave in free space:

$$\vec{E}(\vec{r}) = \vec{E_0} e^{\pm \vec{k} \cdot \vec{r}}$$

where $\vec{k} = (k_x, k_y, k_z)$ is the ***wavevector***, $\vec{E_0} = (E_{x_0}, E_{y_0}, E_{z_0})$ are amplitudes of the individual components. Also note that the + sign in the above about refers to a wave traveling in the $-\vec{r}$ direction, whereas the - sign specifies a wave traveling in the $\vec{r}$ direction. Noting that, in general, $\vec{k}$ is a complex vector and can be written as $\vec{k} = \vec{k}_{re} + j\vec{k}_{im}$. The full field $\vec{\mathcal{E}}$ is found to be:

$$\vec{\mathcal{E}}(\vec{r},t) = \mathrm{Re}[\vec{E}_0 e^{j\omega t \pm (\vec{k}_{re} + j\vec{k}_{im})\cdot\vec{r}}] = \vec{E}_0 e^{\pm\vec{k}_{re}\cdot\vec{r}} \cos(\omega t \pm \vec{k}_{im}\cdot\vec{r})$$

we define the phase of a plane wave traveling in free space as:

$$\phi(\vec{r}) = \pm\vec{k}_{im}\cdot\vec{r} \tag{1.6}$$

where the units for ϕ are radians, and the wavevector necessarily has units of rad/m. We proceed by discussing phase information similar to that defined above can be determined via experiments based on wave optics.

1.1.2 Wave Optics

Wave optics is the study of light waves, their optical properties, propagation, and how optical components (lenses, transparencies, etc.) transform them. This is in contrast to geometrical optics where light is assumed to travel in straight lines between points [11]. For wave optics, light travels as an electromagnetic wave, which can be analyzed in a similar fashion to the approach presented in 1.1.1. In the free space case, light simply travels at the speed of light, $c_o \approx 2.99792458 \cdot 10^8$ m/s. When light enters a transparent medium with refractive index $n > 1$, the speed of light becomes:

$$c = \frac{c_o}{n}$$

More importantly, light waves obey the well known ***wave equation***, where if we use $\psi(\vec{r},t)$ to represent the 'wavefunction', then this function obeys the following differential equation:

$$\nabla^2\psi = \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} \tag{1.7}$$

This relation is essentially a scalar Helmholtz equation (see Eq. (1.4)), but with an explicit time-dependence. This allows us to study the wave nature of light from an electromagnetic perspective, which is especially relevant in the context of phase imaging where the objects being imaged (e.g. cells, tissues, organisms, etc.) possess a length scale on the order of the wavelength of the illuminating light. Of particular interest in QPI is the phenomenon of ***wave interference***, in which the sum of N waves depends not only on the intensity (amplitude) of the waves, but also their relative phases and coherence properties. To see this, we first consider one of the most important (and simplest) cases of interference: two monochromatic (frequency ω) plane waves occupying the same region of space. Let $\psi_1(\vec{r})$ and $\psi_2(\vec{r})$ represent the complex amplitudes (with phases ϕ_1, ϕ_2 respectively) of the two waves, then the sum is another with frequency ω:

$$\psi(\vec{r}) = \psi_1(\vec{r}) + \psi_2(\vec{r})$$

If we define the ***optical intensity*** as $I(\vec{r}) = \psi\psi^* = |\psi(\vec{r})|^2$, then the total intensity is:

$$I(\vec{r}) = \psi\psi^* = |\psi_1|^2 + |\psi_2|^2 + \psi_1^*\psi_2 + \psi_1\psi_2^*$$

If we express the two waves as $\psi_1 = \sqrt{I_1}e^{j\phi_1}$ and $\psi_2 = \sqrt{I_2}e^{j\phi_2}$ then this intensity can be written as:

$$I(\vec{r}) = I_1(\vec{r}) + I_2(\vec{r}) + (\psi_1^*\psi_2) + (\psi_1^*\psi_2)^*$$

$$I(\vec{r}) = I_1(\vec{r}) + I_2(\vec{r}) + 2\sqrt{I_1 I_2}\cos(\phi_2 - \phi_1) \tag{1.8}$$

Equation (1.8) is known as an ***interferogram***, and in this case it represents the interference of two beams (waves). This is the fundamental equation which we will expand on and eventually use as the basis for our training data. Technically, due to extremely high frequency

of optical waves (224 THz for 840 nm light) any measurement of intensity is necessarily a time average e.g. the definition in terms of the field associated with the optical wave is $I(\vec{r}) = 2\langle\Psi^2(\vec{r},t)\rangle$ [11]. We will return to this definition when discussing statistical optics.

Because the intensity given in Eq. (1.8) depends on the phase difference between the two waves, we can use an interferogram as a proxy for the phase information $\phi_2 - \phi_1$. This is accomplished through use of an ***optical interferometer***, a broad class of devices which are capable of measuring phase differences with extremely high spatial and temporal sensitivity (often sub-nanometer - see for example [1], [3], and [12]). We will first consider one of the most fundamental types of interferometers, the Michelson Interferometer. Figure 1.2 shows a schematic of a basic Michelson interferometer. Light from a (typically monochromatic or quasi-monochromatic) point source S is collimated by a lens and directed towards a beam splitter 'BS' (typically a 50/50 beam splitter). Two waves are generated, with ψ_1 directed towards mirror number 1 (M_1), the other, ψ_2 towards mirror 2 (M_2), and then reflected back towards the beam splitter. The portion of ψ_1 which is transmitted through the beam splitter is combined with the portion of ψ_2 reflected by the back of the beam splitter towards the camera. The two beams then combine to form the total field at the detector (the camera), ψ, and the resulting interferogram can be described by Eq. (1.8).

We can analyze the resulting interferogram in more detail by specifying some parameters as well as the forms of ψ_1 and ψ_2. First, we assume that we have an idealized monochromatic (single λ) source, and that we are far enough from S to safely assume that all of the waves can be represented reasonably well as plane waves. Second, we assume that the wave travelling towards the camera only travels in the positive z direction, meaning we have $\vec{k} = k_z\hat{z}$, where

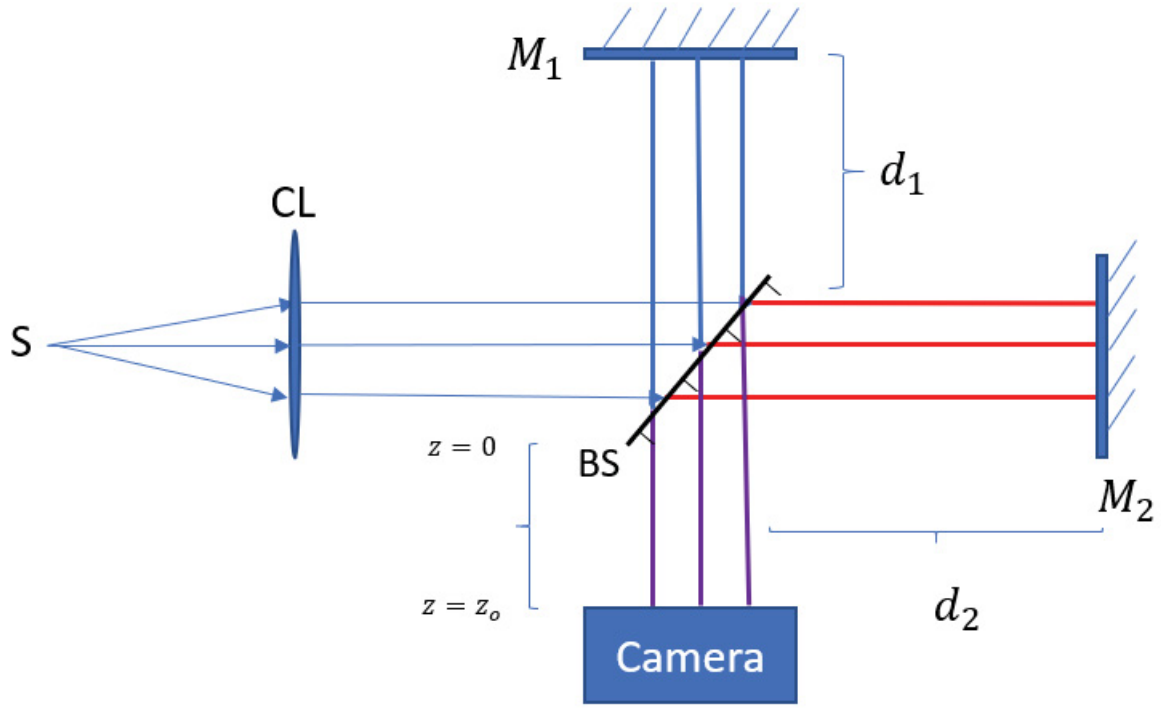

Figure 1.2: Basic diagram of a traditional Michelson interferometer. Light from a point source is collimated by a lens (CL) into a beam splitter. The two beams then travel distances $2d_1$ and $2d_2$, respectively, before being combined and the resulting interferogram is captured on the camera (or any other suitable photo-detector).

$k_z = 2\pi/\lambda$. Thus, for the region $z \in [0, z_0]$ we can write:

$$\begin{aligned} \psi_1 &= \sqrt{I_1} e^{-j(\vec{k}\cdot\vec{r})-2kd_1-kd_0}, \quad \psi_2 = \sqrt{I_2} e^{-j(\vec{k}\cdot\vec{r})-2kd_2-kd_0} \\ \psi_1(z) &= \sqrt{I_1} e^{-jk_z z-2k_z d_1-k_z d_0}, \quad \psi_2(z) = \sqrt{I_2} e^{-jk_z z-2k_z d_2-k_z d_0} \end{aligned} \tag{1.9}$$

where the additional phase shifts $-2kd_1$ and $-2kd_2$ account for the phase that each wave has travelled through distances $2d_1$ and $2d_2$ before recombination, respectively. Additionally, there is a phase shift common to both waves, $-kd_0$ which accounts for the entire distance the wave emitted from the sourced travels before reaching the beam splitter. However, this is simply a constant phase shift that will drop out in the interferogram. The total intensity at the detector ($z = z_0$) $I(z = z_0) = |\psi|^2 = |\psi_1(z = z_0) + \psi_2(z = z_0)|^2$ is found to be:

$$I(z = z_0) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos[(k_z z_0 + 2k_z d_2 + k_z d_0) - (k_z z_0 + 2k_z d_1 + k_z d_0)]$$

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos[2k_z(d_2 - d_1)]$$

We can simplify the above relation by defining the ***DC component*** of the interferogram as $\alpha \equiv I_1 + I_2$. We define the ***visibility*** of the interferogram as $V \equiv 2\sqrt{I_1 I_2}/(I_1 + I_2)$. Additionally, If we write $d = 2(d_2 - d_1)$, and write $k_z = nk_0$, where k_0 is the free space wavenumber and n is the refractive index of the medium the light in traveling in, we can write the interference equation in terms of the quantity $L = nd$ which is known as the Optical Pathlength (OPL):

$$I = \alpha[1 + V\cos k_0 L], \quad k_0 = \frac{2\pi}{\lambda_0} \tag{1.10}$$

In Eq. (1.10), α is a positive quantity, and the visibility ranges from $[0, 1]$. A normalized (I/α) plot of this pattern with $\lambda_0 = 800$ nm and $V = 1$ is shown in Fig. 1.3. Note that

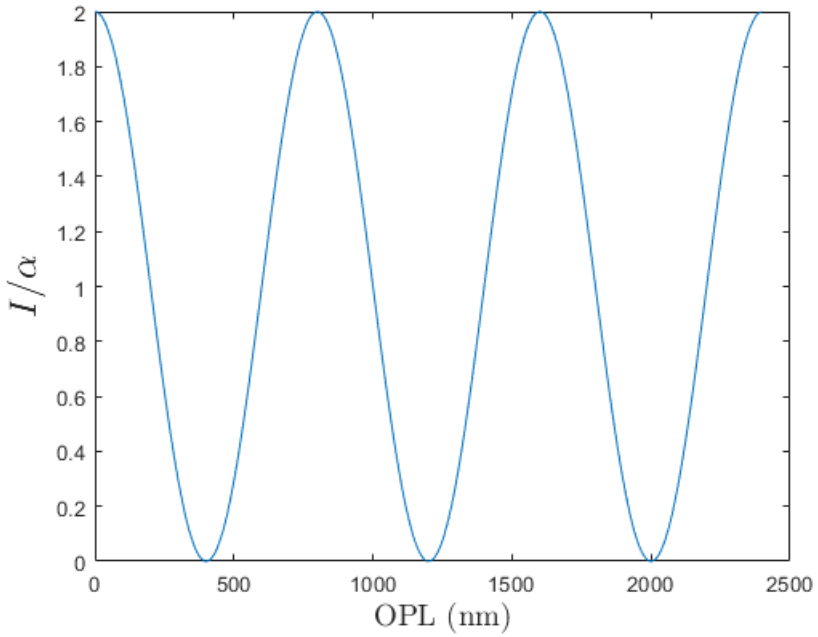

Figure 1.3: Normalized plot of Eq. (1.10) with $V = 1$ and $\lambda_0 = 800$ nm.

the interference pattern of two monochromatic waves has period λ_0. We will be focused on calculating the OPL directly through the use of NNs. We continue by discussing statistical optics, which will allow us to analyze the type of interferogram just introduced in a more practical context. The statistical optics point of view of Eq. (1.10) also provides a good context to discuss the notion of the ***coherence*** of optical waves.

1.2 Statistical Optics

To begin, let us return to situation in Fig. 1.2, but with a few twists. First, we define $d_2 = d_1 + h$, where h is some small difference in the two arms. This is done to simplify some of the resulting expressions. We will also drop the explicit $\vec{r}$ dependence for the phasor representation of ψ, as the interference will be found to depend on it's auto-correlation function. We also will re-interpret ψ to represent an analytic signal representation of the

complex amplitude. Essentially, this amounts to considering ψ to be a sample function of a complex random process Ψ. See [13] for more details. Additionally, we will no longer assume the light source to be monochromatic. Rather, as is the case for real laser (coherent) sources, there will naturally be some finite spectral bandwidth, $\Delta\lambda$. We will further assume that S represents a narrowband source in the sense that $\Delta\lambda \ll \lambda_0$. This will have a drastic effect on when we are able to observe an interference pattern such as the one in Fig. 1.3. To account for these effects, we use the definition of optical intensity (at the detector):

$$I(\tau) = \left\langle |\xi_1\psi(t) + \xi_2\psi(t+\tau)|^2 \right\rangle, \quad \tau = \frac{L}{c_0}$$

where ξ_1 and ξ_2 are constants representing any differences between the two arms which could affect the amplitude of the two waves. It is again assumed that the waves (and thus $\vec{k} = k_z\hat{z}$) travel in the positive z direction. The phase shift (time delay) that arises due to the distance traveled along the z-axis in Fig. 1.2 cancels, as before. Note that we have written the time that the light arriving from the second arm of the interferometer is delayed by in terms of the OPL. That is, in our original notation, $\psi_1 = \psi(t)$, $\psi_2 = \psi(t+\tau)$ where $\tau = L/c_0$, and $L = n(2h)$. Also note that I is a function of τ. Expanding Eq. (1.2) yields:

$$I(\tau) = \xi_1^2\langle|\psi(t)|^2\rangle + \xi_2^2\langle|\psi(t+\tau)|^2\rangle + \xi_1\xi_2\langle|\psi(t+\tau)\psi^*(t)|\rangle + \xi_1\xi_2\langle|\psi(t)\psi^*(t+\tau)|\rangle \quad (1.11)$$

by noting that the time average of an optical signal will not depend on the time delay τ, we can write: $I_0 \equiv \langle|\psi(t)|^2\rangle = \langle|\psi(t+\tau)|^2\rangle$ Furthermore, we can restore the form of Eq. 1.10 by defining: $I_1 = \xi_1^2 I_0, \quad I_2 = \xi_2^2 I_0$ Finally, we define the auto-correlation function for as [14]:

$$\Gamma(\tau) = \langle \psi(t+\tau)\psi^*(t) \rangle \tag{1.12}$$

Substituting these relations into Eq. (1.11) and noting that $\xi_1 = \sqrt{I_1/I_0}$, and $\xi_2 = \sqrt{I_2/I_0}$ we obtain:

$$I(\tau) = I_1 + I_2 + \tfrac{\sqrt{I_1 I_2}}{I_0}\Gamma(\tau) + \tfrac{\sqrt{I_1 I_2}}{I_0}\Gamma^*(\tau)$$
$$I(\tau) = I_1 + I_2 + \tfrac{\sqrt{I_1 I_2}}{I_0}(\Gamma(\tau) + \Gamma^*(\tau))$$
$$\Gamma(\tau) + \Gamma^*(\tau) = 2\,\mathrm{Re}\{\Gamma(\tau)\}$$

Note that $\Gamma(0) = \langle |\psi(t)|^2 \rangle = I_0$. We can therefore define a normalized auto-correlation function $\gamma(\tau) = \Gamma(\tau)/\Gamma(0)$. This function is also known as the "complex degree of coherence" [13]. Since $\gamma(\tau)$ is complex, it can be written as:

$$\gamma(\tau) = |\gamma(\tau)| e^{-j(\omega\tau - \kappa(\tau))}$$

where $\kappa(\tau)$ represents an inherent delay-dependent phase shift from the source S, and is defined as $\kappa(\tau) = \angle\gamma(\tau) + \omega_0\tau$ and $\omega_0 = 2\pi c/\lambda_0$. The final expression for $I(\tau)$ (equivalently $I(L)$), is:

$$I(\tau) = I_1 + I_2 + 2\sqrt{I_1 I_2}|\gamma(\tau)| \cos(\omega_0\tau - \kappa(\tau)) \tag{1.13}$$

If we follow the approach used in the case of purely monochromatic light, we can again define $\alpha = I_1 + I_2$, but the visibility is now given by:

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}|\gamma(\tau)| = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}\left|\gamma\left(\frac{L}{nc}\right)\right| \tag{1.14}$$

the visibility is found to depend on the delay (OPL) introduced by the different lengths of the arms in the interferometer. This can have a significant impact depending on the shape of the power spectral density (PSD) of the source S. To see why, we refer to the Wiener-Khinchin theorem which states that the PSD and the auto-correlation function (an even function), $\Gamma(\tau)$, are a Fourier transform pair [14]:

$$\Gamma(\tau) = \frac{1}{\pi}\int_0^\infty S(\omega)e^{j\omega\tau}d\omega \tag{1.15}$$

where $S(\omega)$ is the PSD of the source light (with complex analytic signal $\psi(\vec{r},t)$). This is an important property when analyzing an optical system and/or light source, as $S(\omega)$ is readily obtained (for example, from a spectrometer) but $\Gamma(\tau)$ often is not. If a normalized spectrum $\hat{S}$ is used, the complex degree of coherence $\gamma(\tau)$ is obtained directly.

Gaussian PSD

In this section we perform a simple example to explore the effects of the complex degree of coherence on the interference pattern detected by the Michelson interferometer. Perhaps the most common distribution encountered in practice is the Gaussian PSD:

$$\hat{S}(\omega) = \frac{4\sqrt{\pi}\sqrt{\ln 2}}{\Delta\omega}\exp\left[-\left(2\sqrt{\ln 2}\frac{\omega-\omega_0}{\Delta\omega}\right)^2\right]$$

Inverse Fourier transforming the above expression gives $\gamma(\tau)$ as another Gaussian [15]:

$$\gamma(\tau) = e^{-j\omega_0\tau}\exp\left[-\left(\frac{\Delta\omega\tau}{4\sqrt{\ln 2}}\right)^2\right]$$

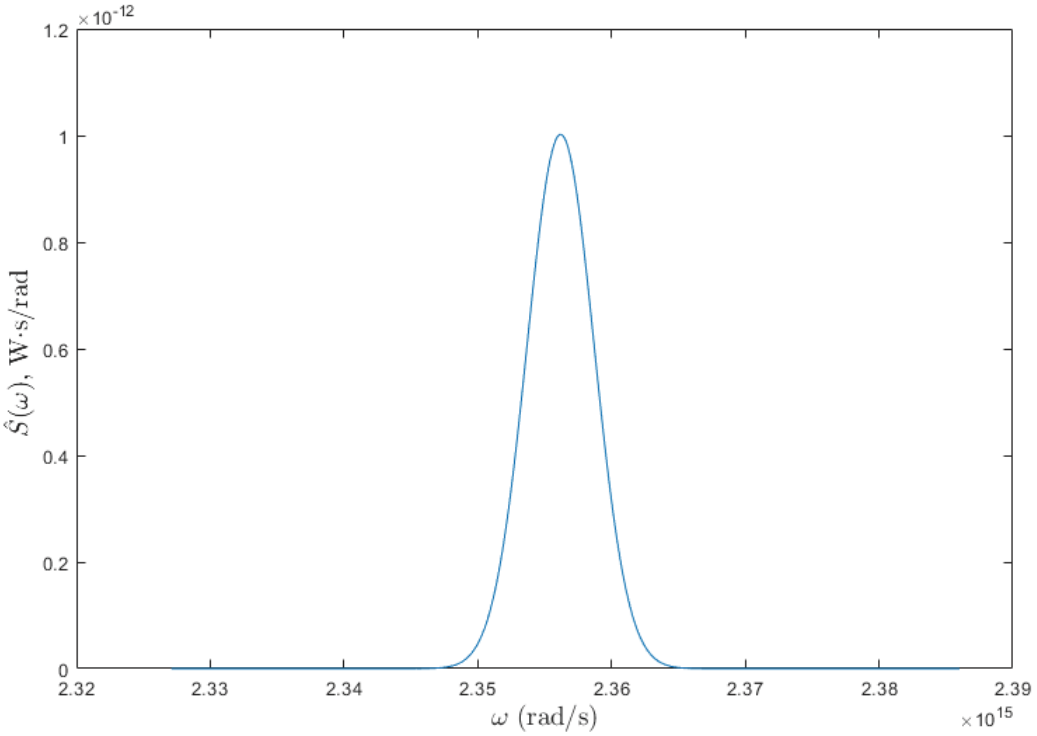

Figure 1.4: Plot of a Gaussian Power Spectral Density for $\Delta\omega \approx 5.89 * 10^{12}$ rad/s, $\omega_0 = 2\pi c/\lambda_0 \approx 2.356 * 10^{15}$ rad/s.

Evidently, $\kappa(\tau) = 0$ in this case. However, based on Eq. (1.14) and the expression above, we can see that the visibility of the interferogram decreases as τ increases or equivalently as L increases. Figure 1.5 shows a similar plot to Fig. 1.3, but based on Eq. 1.13 and the expression above for $\gamma(\tau)$. We have used the same parameters as before ($\lambda_0 = 800$ nm, $I_1 = I_2$ so that $V(L) = \gamma(L)$, $n = 1$), and $\Delta\lambda = 2$ nm which gives $\Delta\omega = 2\pi c\Delta\lambda/\lambda_0^2$ to produce Fig. 1.5. The PSD is also plotted in Fig. 1.4 for the same parameters.

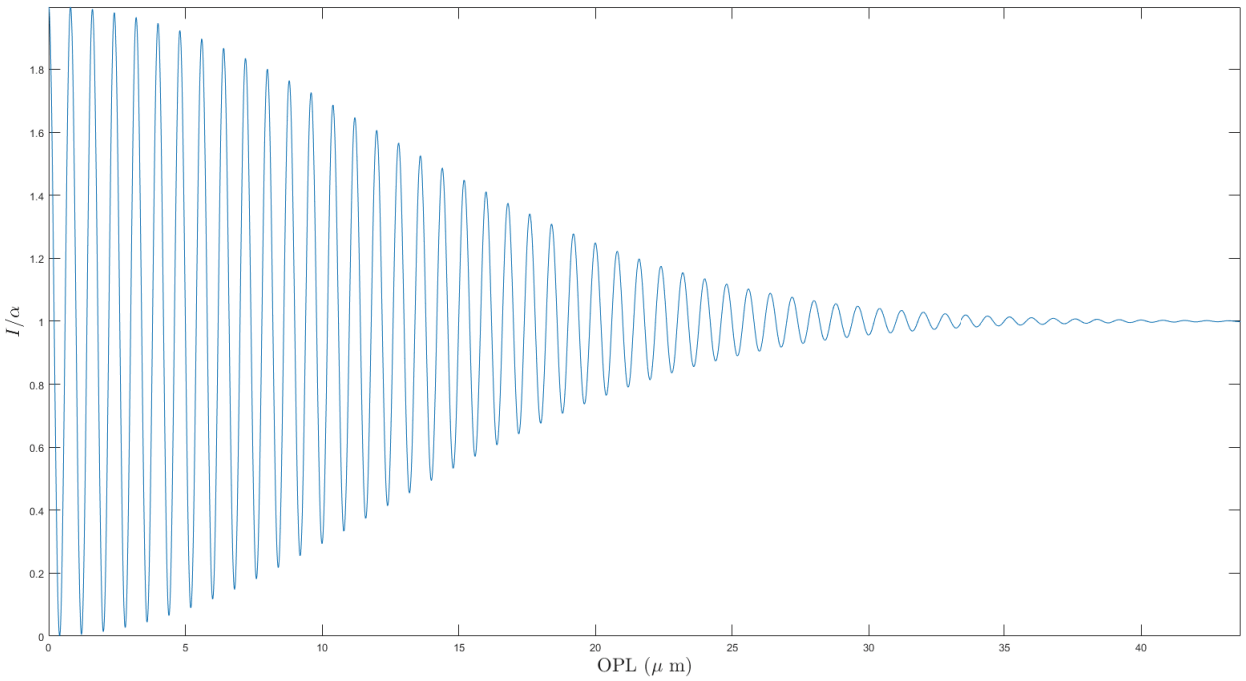

Figure 1.5: Normalized plot of Eq. (1.13) including the effects of coherence with $I_1 = I_2$, $\lambda_0 = 800$ nm, $\Delta\lambda = 2$ nm. The coherence length can be seen to be approximately 32µm.

We can see from Fig. 1.5 that the affect of the complex degree of coherence in the case of a Gaussian line is to slowly reduce the envelope (or visibility) as the delay between the two arms increases (plotted as OPL). Eventually, the OPL (delay) is so long that no interference is observed, i.e. the sinusoidal term in Eq. (1.13) has almost disappeared entirely. This effect thus limits the amount of OPL difference which can be introduced in an interferometer before no interference is apparent. This gives rise to the notion of the ***coherence length*** L_c or ***coherence time*** τ_c. The coherence time has a direct relationship to the complex degree of coherence [13]:

$$\tau_c \equiv \int_{-\infty}^{\infty} |\gamma(\tau)|^2 d\tau$$

The coherence length - a physical distance - can then be calculated as:

$$L_c = c\tau_c$$

The coherence time is typically on the order of $\lambda_0^2/(2\pi c\Delta\lambda)$, and so the coherence length is typically on the order of $\lambda_0^2/(2\pi n\Delta\lambda)$. See [13] for details. We continue by discussing a particular QPI technique, wavelength shifting interferometry (WSI), which takes advantage of these interference concepts to produce highly sensitive phase images.

Chapter 2

Neural Networks for Wavelength Shifting Interferometry

2.1 Wavelength Shifting Interferometry

Wavelength shifting interferometry is a QPI technique that has been recently gaining more attention as an effective QPI technique [16, 17, 18, 19, 20]. It involves taking Q interferograms at different wavelengths (wavenumbers) that adhere to an equation similar to Eq. 1.13, and then combining the information from each to solve for the optical pathlength in the following expression directly [20]:

$$I(L) = \alpha \left[1 + V \cos k_n L\right], \quad k_n = k_1, k_2, \ldots, k_Q \tag{2.1}$$

where k_n are the Q *evenly spaced* wavenumbers of the interferograms. The requirement that the wavenumbers be evenly spaced is a typical assumption that ensures the interferogram is being sampled optimally for a given Q. For instance, for a fixed L, one might chose 4 evenly spaced k_n that provide sample phases $\phi_n = k_n L$ which are evenly spaced by $\pi/2$ between samples. In general, however, L is not a constant, especially for non-uniform phase objects such as unstained biological samples, e.g. red blood cells, cheek cells, etc. Therefore, in wavelength shifting interferometry we cannot rely on a constant phase shift between

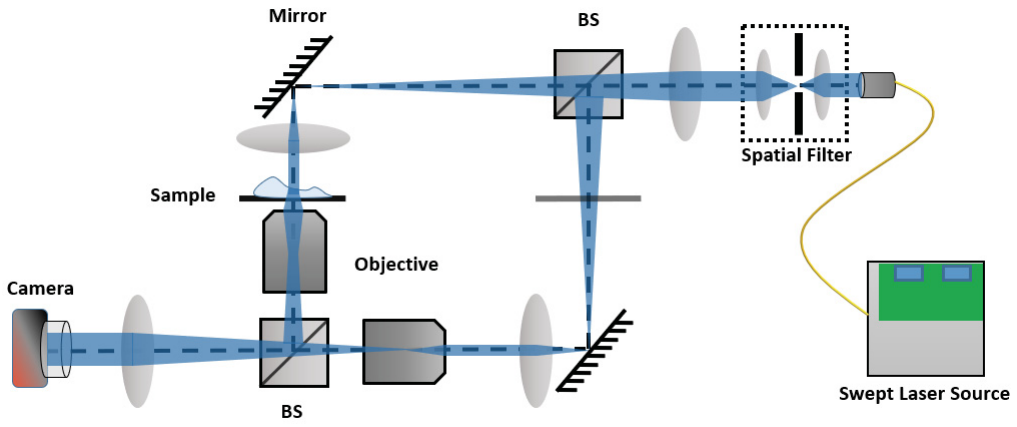

Figure 2.1: A typical WSI system consisting of a Mach-Zehnder interferometer containing two identical beamsplitters (BS), objectives, and mirrors. Adapted with permission from [1].

samples. Rather, we can only ensure that the phase shift between samples is fixed for a given L. In the context of a QPI system which contains some type of interferometer in which the interferogram is recorded on a digital camera, we can take the OPL distribution to be dependent on the position on the camera sensor. That is, for a camera sensor with $M \times N$ pixels, we can write:

$$L \equiv L(x, y), \quad 0 \leq x \leq N, \quad 0 \leq y \leq M$$

Therefore, the goal of the WSI system is to solve for $L(x, y)$ for the entire field of view, and plot this distribution to generate a quantitative phase image. Because we have calculated the OPL (phase) directly, we can also use the phase information to calculate certain parameters of the sample (such as dry mass), or investigate the surface features of a cell with high sensitivity. We must now develop a way to calculate $L(x, y)$ accurately and reliably to achieve high OPL sensitivity.

The answer, in part, depends on Q. If many interferograms are taken, algorithms exist to demodulate $L(x, y)$ in Eq. 2.1 such as a Fourier transform based approach [21], where

L can be obtained using the information contained in an FFT [15, 22]. This requires a sophisticated setup, such as an expensive laser. A simpler case is so called 4-band WSI, in which the interferogram in Eq. 2.1 is sampled at four wavelength bands centered at the different $k_n = [k_1, k_2, k_3, k_4]$ [20]. We will focus on this configuration for the remainder of this work, as it has several advantages as well as being simpler to implement requiring only four evenly spaced k_n.

Figure 2.1 shows a system diagram for WSI as implemented in [20]. Light from a continuous wave (CW) swept laser source (Superlum, BS-840-1) with tunable wavelength range 805-880 nm is collimated and spatially filtered before entering a Mach-Zehnder interferometer. The Mach-Zehnder interferometer can be analyzed in precisely the same way that the Michelson interferometer in Fig. 1.2 was, but the Mach-Zehnder configuration provides some practical advantages. For example, with identical beam splitters and mirrors, the light in the two arms is guaranteed to travel the same distance in air and glass, therefore closely matching their OPLs. The only OPL difference should be due purely to the sample (and potentially an inherent difference in the lengths of the two arms, d_1 and d_2). Light from the two arms (reference and sample - analogous to I_1 and I_2) interfere and the resultant intensity is recorded on a camera. In 4-band WSI, four interferograms are produced which correspond to different wavelengths of the swept source obtained by synchronizing the camera to obtain exposures at a set interval. This has the effect of reducing the effective temporal coherence of the light source to avoid speckle and other coherent artifacts [20]. The mean (noise-free) intensities at any pixel (x, y) recorded by the camera are:

$$\bar{I}_n(L(x,y)) = \alpha[1 + V\cos(k_n L(x,y))], \quad n = 1, 2, 3, 4 \tag{2.2}$$

This equation represents the intensity without the influence of noise. We will discuss how to

account for noise statistics when we introduce the concept of sensitivity in the next section. Noisy (experimental) intensities will be represented simply as I_n. In Eq. 2.2, $\alpha = (I_S + I_R)$, $V(L) = 2\sqrt{I_R I_S}/(I_R+I_S)|\gamma(L)|$ is the visibility. I_R and I_S represent the reference and sample arm intensities, respectively. It is important to note that the OPL in Eq. (2.2) is the total OPL difference between the two arms, including the sample. That is, $L(x,y) = L_S(x,y) - L_R$ where L_R is the OPL in the reference arm and L_S is the OPL in the sample arm, presumed to vary much more strongly with position (x,y) than the reference arm OPL. The phase shift between the 4 samples is $\Delta\phi = \Delta k L(x,y)$. This constant (for a given position (x,y)) although unknown due to the unknown $L(x,y)$, allows for the application of the well-known Carré equation [23, 24] which gives the wrapped OPL in terms of the noisy intensity samples, I_n [20]:

$$L' = \frac{1}{k_0}\tan^{-1}\left\{\frac{\sqrt{[3(I_2-I_3)-(I_1-I_4)](I_2-I_3+I_1-I_4)}}{(I_2+I_3-I_1-I_4)\times \mathrm{sgn}(I_2-I_3)}\right\} \tag{2.3}$$

where the $'$ refers only to the wrapped OPL obtained via Eq. (2.3). This equation can be derived by combining the four equations implied Eq. 2.2 and by making the assumption that the different wavenumbers are evenly spaced [23]. An example of a QPI image generated with this equation is given in Fig. 2.2, in which Zernike polynomial background subtraction [25] and 2-D phase unwrapping has been applied to $L'(x,y)$ to minimize the influence of the background illumination and eliminate wrapped phase in the final image, respectively. Also shown are the corresponding intensities I_1, I_2, I_3, I_4. The timing between the swept source and the camera was configured to obtain the following evenly spaced k_n: $[k_1, \ldots, k_4] = [7.22, 7.36, 7.50, 7.63]$ rad/µm, which is consistent with the original paper [20].

We have defined the basic operation of a 4-band WSI system, we move on to discuss the concept of sensitivity and how it can be used to evaluate a QPI system like WSI.

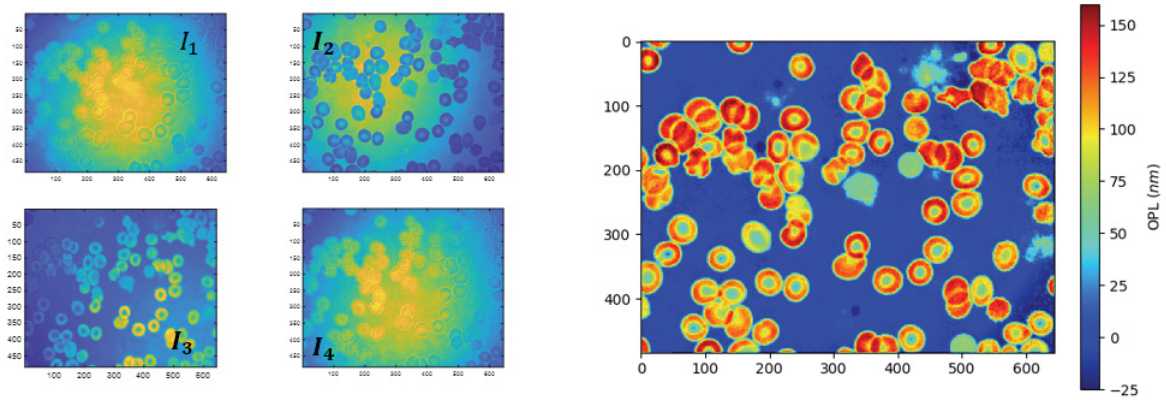

Figure 2.2: **Left**: Interferograms obtained from a WSI system with $M = 484$ pixels, $N = 644$ pixels. The subscripts refer to the 4 different wavenumbers, respectively. **Right**: OPL distribution obtained from the Carré equation (2.3). The cells OPL varies on top of a reference OPL of approximately 13.5 µm. The horizontal and vertical axes are in units of pixels, and the colorbar has units of nm.

2.2 Sensitivity

The concept of sensitivity is of utmost importance in the context of QPI. We are mainly concerned with temporal sensitivity. Temporal sensitivity is defined as the standard deviations of a time series of OPL measurements. For example, if we generate 400 interferograms in a 4-band WSI system, this gives us 100 OPL images, such as in Fig. 2.2, corresponding to 100 $L_i(x, y)$. If the I_n were completely noise-free, then we would expect all of the $L_i(x, y) = L_1 = L_2 = \cdots = L_{100}$ given by Eq. (2.3) to be precisely identical and the temporal standard deviation would be zero. In reality, this is not the case and there is always some noise present in an optical system which leads to a non-zero sensitivity. In particular, many QPI systems are ***shot noise-limited***, which is noise related to the detection of photons themselves and can never be completely eradicated. This means there will always be some inherent variance in the measurement of L in any QPI system.

The standard deviation of the L_i in our example is the temporal sensitivity:

$$\sigma_L = \sqrt{\mathrm{Var}(L_i)} \tag{2.4}$$

Sensitivity as defined above captures a given QPI system's ability to reliably discern fine features of microscopic objects or samples. For example, if $\sigma_L = 100$ nm, then we cannot rely on our QPI system to accurately characterize any OPL changes less than roughly 100 nm as they will be hidden in the variance of the system. On the other hand, if $\sigma_L = 1$ nm, then OPL fluctuations can reliably be detected down to single nanometers, allowing for significantly smaller features to be reliably imaged and quantified as phase.

2.2.1 Cramér-Rao Bound

Since sensitivity is defined in terms of a variance of a parameter being estimated, there is a lower bound to this variance. This lower bound is known as the Cramér-Rao bound (CRB). The CRB is formally the reciprocal of the Fisher information when estimating a single parameter, or inverse of the Fisher information matrix (FIM) when multiple parameters are to be estimated. Given a vector of Q observations, which are random variables $\vec{X} = (X_1, X_2, \ldots, X_Q)$, and a vector of underlying parameters to be estimated $\vec{\theta}$ with length K, the FIM is the expected value of the log likelihood function [26]:

$$J_{ij} = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j}[\log[p(\vec{X}|\vec{\theta})]\right], \quad i, j = 1, 2, \ldots,\ Q \tag{2.5}$$

where $p(\cdot)$ is the conditional joint probability mass function for $\vec{X}$ conditioned on $\vec{\theta}$. The CRB is then given by:

$$(\sigma_{CRB})_{ij} = \sqrt{(J^{-1})_{ij}}, \quad i, j = 1, 2, \ldots, \ Q \tag{2.6}$$

It is clear from Eqs. (2.5), (2.6) that the CRB only depends on the observational data, and the parameters we are interested in estimating. For WSI, our observations are the raw data obtained from the system and the unknown parameters in the system to be estimated are $\vec{\theta} = (\alpha, V, L)$. To completely specify the CRB, we must also account for the noise our signal is likely to be corrupted by. As mentioned previously, in our WSI system and indeed many QPI systems, the limiting noise in the system is typically shot noise. We briefly discuss this phenomenon below before deriving the CRB for our signal model (Eq. (2.2)).

2.2.2 Shot Noise

Shot noise is an inherent limitation present in all optical systems due to the quantum nature of photons [27]. It represents the fundamental limit for the signal-to-noise ratio when measuring intensity by the collection of photons. The quantum nature of photon detection means that at a given position on a photo-detector, there is only a probability that a photon will be detected at a given time (for example, during the time the photo-detector is exposed to light). In a digital camera, photons are collected during an exposure by the sensor and converted to photo-electrons before being 'dumped' into an analog-to-digital converter. In the case of optical hetorodyne detection (which Eq. (1.13) can be considered a form of), a general SNR can be derived in terms of the reference arm power, P_R, and the signal power P_s as (see [27] for details):

$$SNR = 8\pi^3 \frac{2G^2(P_s P_R)(e\eta/h\omega)^2}{\Delta\omega(2G^2 e(i_d + 2\pi P_R e\eta/h\omega) + \mathcal{N}_T)} \tag{2.7}$$

where G is the ***conversion gain*** which converts photo-electrons generated by the detector into 'analog digital units' (ADU) within the A/D converter. G thus has units e^-/ADU. h is Planck's constant, e is the electron charge, η is the quantum efficiency, ω is the frequency of the source light, $\Delta\omega$ is the bandwidth over which our signal exists. $\mathcal{N}_T$ is a thermal noise term known as Johnson noise. The term $2G^2e(i_d + 2\pi P_R e\eta/h\omega)$ represents the total shot noise. Note that the total shot noise depends on P_R. As $P_R \to \infty$, the denominator becomes dominated by the shot noise term and the SNR becomes:

$$SNR_{P_R\to\infty} = 4\pi^2 \frac{P_s}{h\omega\Delta\omega}\eta \tag{2.8}$$

This means that, with a very efficient camera ($\eta \to 1$), and sufficient laser power (P_R), we become limited by the photon arrival process i.e shot noise. This arrival process has been shown to follow Poisson statistics [27], which allows us to model the noise in a QPI system as Poisson distributed (assuming all other forms of noise are accounted for). In the absence of other significant noises associated with the sensor, as long as we can saturate the photo-detector, our SNR becomes limited by Eq. (2.8). Since shot noise is the fundamental limiting factor in a QPI system, it must be the case that the minimum variance associated with the CRB can only be realized when our system is in fact shot noise-limited.

2.2.3 WSI Sensitivity Analysis

We are now prepared to analyze the sensitivity of a 4-band WSI system. First, we must specify the Poisson statistics which corrupt our noise-free signal in Eq. (2.2). Following the approach detailed in [2], let $\vec{X} = [X_1, \ldots, X_4]$ represent the electrons collected by a particular pixel on the camera for each of the four wavelength bands. In a shot noise-limited system, each of the X_n is an independent Poisson random variable with $\mathrm{E}(X_n) = \mathrm{Var}(X_n) = s_n$,

where the s_n are the average photo-generated electrons. For digital cameras, the s_n are then converted to intensities in ADU by the conversion gain G. Therefore, $\bar{I}_n = s_n/G$ represents the noise-free intensity for each band, while $I_n = X_n/G$ represents the shot noise-corrupted intensities. To derive the CRB, we follow the approach in Sec. 2.2.1 by calculating the fisher information matrix as follows with $\vec{\theta} = (\alpha, V, L)$:

$$p(\vec{X}|\vec{\theta}) = \prod_{n=1}^{4} \mathrm{Po}(X_n|s_n) = \prod_{n=1}^{4} \frac{s_n^{X_n}}{X_n!} e^{-s_n}$$

$$J_{ij} = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j}[\log[p(\vec{X}|\vec{\theta})]\right], \quad i,j = 1,2,\ldots,\ Q$$

$$J_{ij} = -\,\mathrm{E}\left[-\sum_{n=1}^{4} \frac{X_n}{s_n^2} \frac{\partial s_n}{\partial\theta_i} \frac{\partial s_n}{\partial\theta_j}\right]$$

Noting that we can interchange the sum and the expectation and that $\mathrm{E}(X_n) = s_n$, we obtain:

$$J_{ij} = \sum_{n=1}^{4} \frac{1}{s_n} \frac{\partial s_n}{\partial\theta_i} \frac{\partial s_n}{\partial\theta_j}$$

Carrying out the partial derivatives and using $s_n = G\bar{I}_n$ where $\bar{I}_n$ is simply Eq. (2.2), we find the following 3×3 FIM, J:

$$J = \begin{pmatrix} \sum_{n=1}^{4} \frac{G}{\bar{I}_n} & \sum_{n=1}^{4} \frac{G\cos(k_n L)}{\bar{I}_n} & \sum_{n=1}^{4} \frac{-Gk_n\alpha V\sin(k_n L)}{\bar{I}_n} \\ \sum_{n=1}^{4} \frac{G\cos(k_n L)}{\bar{I}_n} & \sum_{n=1}^{4} \frac{G\cos^2(k_n L)}{\bar{I}_n} & \sum_{n=1}^{4} \frac{-Gk_n\alpha V\sin(2k_n L)}{2\bar{I}_n} \\ \sum_{n=1}^{4} \frac{-Gk_n\alpha V\sin(k_n L)}{\bar{I}_n} & \sum_{n=1}^{4} \frac{-Gk_n\alpha V\sin(2k_n L)}{2\bar{I}_n} & \sum_{n=1}^{4} \frac{Gk_n^2\alpha^2 V^2\sin^2(k_n L)}{\bar{I}_n} \end{pmatrix} \tag{2.9}$$

Diagonal terms of the inverse of the FIM, J^{-1}, gives the lower bound for a different parameter given one or two of the others. For example, if we took the upper left entry, $(J^{-1})_{11}$, the associated lower bound would be:

$$(\sigma_{11})_{CRB} = \sqrt{(J^{-1})_{11}} = (\sigma_\alpha)_{CRB}$$

That is, this entry gives us the lower bound on the standard deviation for an unbiased estimator of α with V, L unknown. For WSI, however, we are interested in the lower bound on the variance of estimating L, which is given by the lower right entry in J^{-1}:

$$(\sigma_{33})_{CRB} = \sqrt{(J^{-1})_{33}} \equiv \sigma_{CRB} \tag{2.10}$$

For the remainder of this work, whenever we refer to the CRB sensitivity, we are referring to the above equation. This is the CRB for estimating L without knowledge of α, V. Furthermore, all sensitivities from now on will refer specifically to the estimation of OPL. It is also important to discuss the effect of particular demodulation schemes (e.g. Eq.(2.3)) on the sensitivity. This sensitivity is referred to as the algorithmic sensitivity and is defined as $\sigma_{ALG} = \sqrt{\mathrm{Var}(L_{ALG})}$. In general, a particular demodulation algorithm degrades the sensitivity when compared to the CRB, as suggested in Fig. 2.3. This is due, in part, to the fact that Eq. (2.3) represents an unbiased estimator of L. This implies that the estimate, L_{ALG},

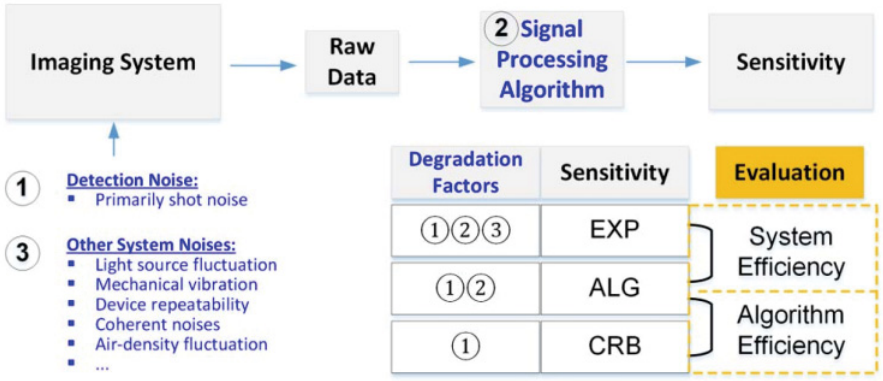

Figure 2.3: Graphical representation of the general sensitivity evaluation framework for QPI systems. Reproduced with permission from the author [2].

must come with some variance greater than or equal to that given by the CRB [14]. We can approximate σ_{ALG} by expanding (2.3) in a first order Taylor series (ignoring all other terms):

$$\mathrm{Var}(L_{ALG}) = \sum_{n=1}^{4} \left(\frac{\partial L}{\partial I_n}|_{\bar{I}_n} \right)^2 \mathrm{Var}(I_n)$$

where $\mathrm{Var}(I_n)$ is the variance associated with the intensity distributions themselves (ideally, shot noise limited). The full expression for σ_{ALG} can be found in Appendix A.1.

Finally, we discuss the experimentally obtained sensitivity, σ_{EXP}. This is the sensitivity observed in practice when using a particular algorithm to obtain L from raw experimental data. σ_{EXP} can be obtained by recording many interferograms of the same sample (e.g. a blank sample), and taking the standard deviation of the OPL distributions so obtained i.e $\sigma_{EXP} = \sqrt{\mathrm{Var}(L_{EXP})}$. This sensitivity is degraded further from σ_{ALG}, due to the possible presence of many different types of noises and imperfections present in a real QPI system. Therefore, in general, we have:

$$\sigma_{EXP} \geq \sigma_{ALG} \geq \sigma_{CRB}$$

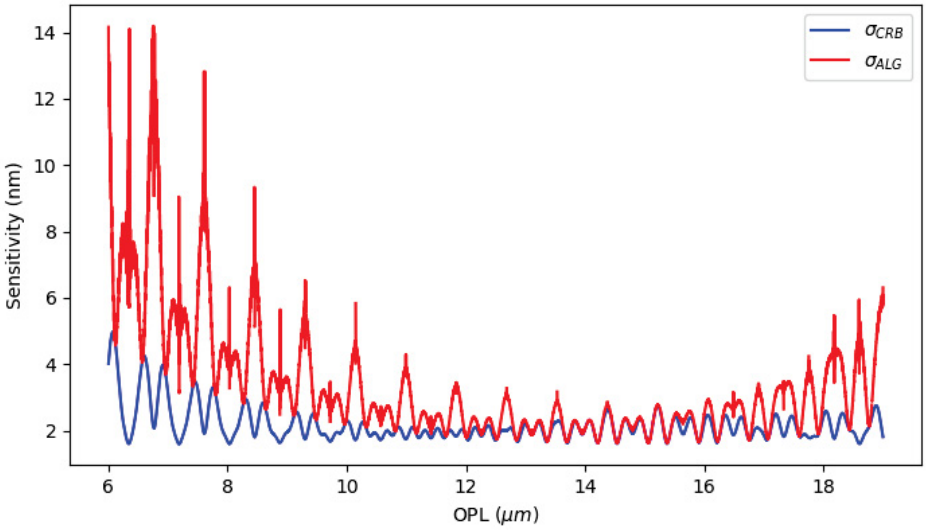

Figure 2.4: Simulated sensitivities σ_{ALG} and σ_{CRB}. Clearly, the CRB is a lower bound for the Carré equation sensitivity (σ_{ALG}), and there are significant gaps between the two suggesting room for improvement through the range of OPL plotted. Note that the OPL includes the sample and the background OPL difference between the two arms. These curves were plotted for typical values of $\alpha = 128$ ADU and $V = 0.7$.

σ_{ALG} and σ_{CRB} are plotted together in 2.4 holding α and V constant at 128 ADU and 0.7, respectively. L is then swept from $[2, 18]$ µm. It is clear that the Carré equation only approaches the limits predicted by the Cramér-Rao bound(s) for a relatively small range of OPLs. This suggests that there is room for improvement, especially towards shorter (longer as well) OPLs, where the difference between σ_{ALG} and the theoretical minimum(s) becomes increasingly large. This is a major motivation for the application of NNs for calculating $L(x, y)$, as there are large gains in sensitivity to be made across a wide range of OPLs. We will show how a DNN can be trained to surpass algorithmic sensitivity and achieve CRB.

2.3 Neural Networks

Neural Networks are a class of algorithms capable of effectively modeling highly non-linear processes [28, 29] through a procedure in which training data is provided as input to the network, and the output of the network is then compared against expected outcomes. This is generally referred to as supervised learning. The training data consists of input-output pairs, denoted by (T_X, T_Y), where T_Y represents the ground truth, or the value which we expect the network to output in response to the given input, T_X. In our case, the input to the DNN will be intensities ($I_n(x, y)$) from each pixel on the camera's sensor, and the output will be the OPL $L(x, y)$. After training, the output of our DNN, L_{DNN}, will be shown to closely approximate L in Eq. (2.2). The network is repeatedly fed this training dataset a certain number of times known as the number of epochs, defined at training time.

For this application, we chose to use a traditional, fully connected DNN consisting of 9 layers. This type of network has been shown to approximate the optimal minimum mean square (MMS) estimator when used for parameter estimation [30]. Figure 2.5 shows a representative plot of our network's structure with the weights, $\mathbf{w}$, denoted by lines of differing thicknesses. Note that the number of nodes pictured in Fig. 2.5 is not meant to represent the exact number of nodes in the networks used. See Appendix A.2.2. for the exact neural network model definition (given as Python code) which contain thousands of nodes. There are two dropout layers included immediately after the first hidden layer and prior to the output layer with rate 0.01 to help prevent over fitting. The activation function for each layer is the standard sigmoid function $S(x) = 1/(1 + e^{-x})$, while the loss function is the usual mean-squared error defined by:

$$E(\mathbf{w}) = \frac{1}{N} \sum_{i=1}^{N} ||L^{(i)}(\mathbf{w})_{DNN} - T_Y^{(i)}||^2 \tag{2.11}$$

where $L^{(i)}(\mathbf{w})_{DNN}$ is the neural network output for the i^{th} training input, and $T_Y^{(i)}$ is the corresponding ground truth value. $E(\mathbf{w})$ represents the objective function being minimized during the training process, and it is emphasized that it is only a function of the weights $\mathbf{w}$ (a three dimensional array) in the network [31]. Clearly, the training set presented to the network will define how the weights are optimized. To understand this somewhat better, a brief description of the concept of *gradient descent* is helpful. When the DNN receives a training input T_X, this input is passed through the network and transformed by the weights matrices within the layers until the output layer is reached. The loss in Eq. (2.11) is then calculated. The weights for the $(n+1)$th ($\mathbf{w}_{n+1}$) epoch over the training set are then adjusted according to the following algorithm for a DNN with K layers [32]:

1 First, the gradient of the loss function E is calculated:

$$\nabla E_n = \left(\frac{\partial E}{\partial (\vec{w}_1)_n}, \frac{\partial E}{\partial (\vec{w}_2)_n}, \ldots, \frac{\partial E}{\partial (\vec{w}_K)_n} \right)$$

2 Since the gradient contains information about the 'direction' (which way, increase or decrease, we alter the weights $\mathbf{w}$) of greatest increase of E, the weights are updated according to the *negative* of ∇E:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \gamma \nabla E_n$$

where γ is a scalar known as the learning rate, which affects the size of the weight update by directly scaling the gradient [33].

3 These steps are repeated until the final weights are determined.

The above algorithm is the simplest implementation of gradient descent. Many more sophisticated techniques are available, including stochastic gradient descent (in which train-

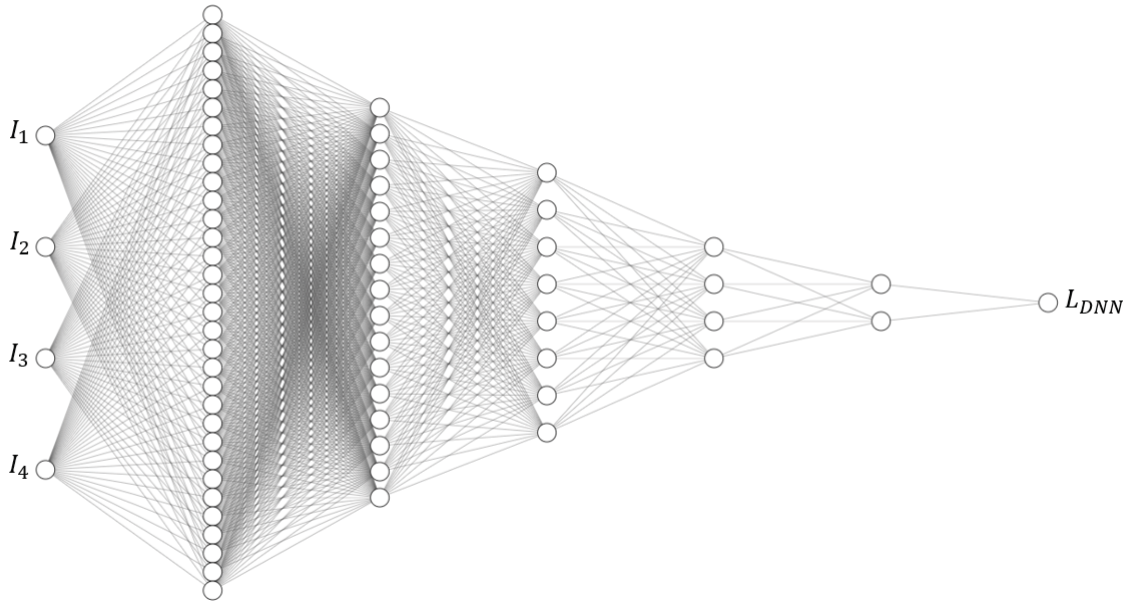

Figure 2.5: Representative fully connected DNN diagram, excluding dropout layers. Each layer's nodes have a direct connection to each node in the next layer. The output is the total OPL ($L_o + L_S$), labeled L_{DNN}.

ing samples are repeatedly shuffled together [34]), and the Adam optimizer used in our algorithms([35]).

As with all neural network based approaches, we must define T_X and T_Y to train a network to demodulate OPL in the context of WSI. This is done in the following section.

2.3.1 Training Data Generation

Consider a particular pixel of an $M \times N$ pixel sensor located at (x_o, y_o), where $0 \leq x_o \leq M$ and $0 \leq y_0 \leq N$. For the purpose of calculating $L(x_o, y_o)$ from the four noise-corrupted intensities $\mathbf{I} = [I_1, I_2, I_3, I_4]$, we require our training data to accurately model the data from our WSI system. To this end, we assume each I_n takes on integer values in the range $[0, 255]$ ADU. This range depends on the particular camera being used, in our case a high-speed

8-bit camera (Allied Vision Mako G030) synchronized with the swept laser source to capture interferograms corresponding to the evenly spaced $[k_1, \ldots, k_4] = [7.22, 7.36, 7.50, 7.63]$ rad/µm. These particular values were chosen to be consistent with [20], but any evenly spaced k_n can be used. Therefore, to generate our training and test sets, we begin by choosing a range of values for the parameters $[\alpha, V, L]$ that reflect typical values encountered in experiment. In our case, the full ranges are $\alpha \in [0, 255]$ ADU (integers) and $V \in [0, 1]$, and $L \in [0, \infty)$. Noise-free (mean) intensity values are then generated according to Eq. (2.2). These mean values, $\bar{I}_n$ (in ADUs), are then converted to photo-electrons, $s_n = \text{Floor}(\bar{I}_n g)$. The s_n are then the mean values for the Poisson distribution which we sample to obtain x_n e.g. $\text{Po}(\text{x}_\text{n}; \text{s}_\text{n})$. Finally, the noisy I_n are obtained from the x_n through $I_n = \text{Floor}(x_n/g)$. Additionally, since it is possible for the same OPL to have different α, V, for each L in the training set, K (α, V) pairs are randomly generated which correspond to a single OPL. Therefore, if the vector of L has length P, T_X is a $K \times P$ matrix where the rows correspond to different noise realizations (due to the Poisson sampling) and (α, V) pairs. The columns then correspond to the different OPLs being trained over. The final input training data matrix, T_X, is:

$$T_X = \begin{pmatrix} \vec{I}_{11} & \ldots & \vec{I}_{1P} \\ \vdots & \ddots & \vdots \\ \vec{I}_{K1} & \ldots & \vec{I}_{KP} \end{pmatrix}$$

where each $\vec{I}_{i,j}$ is a 4-vector corresponding to the different k_n. The truth data for this training

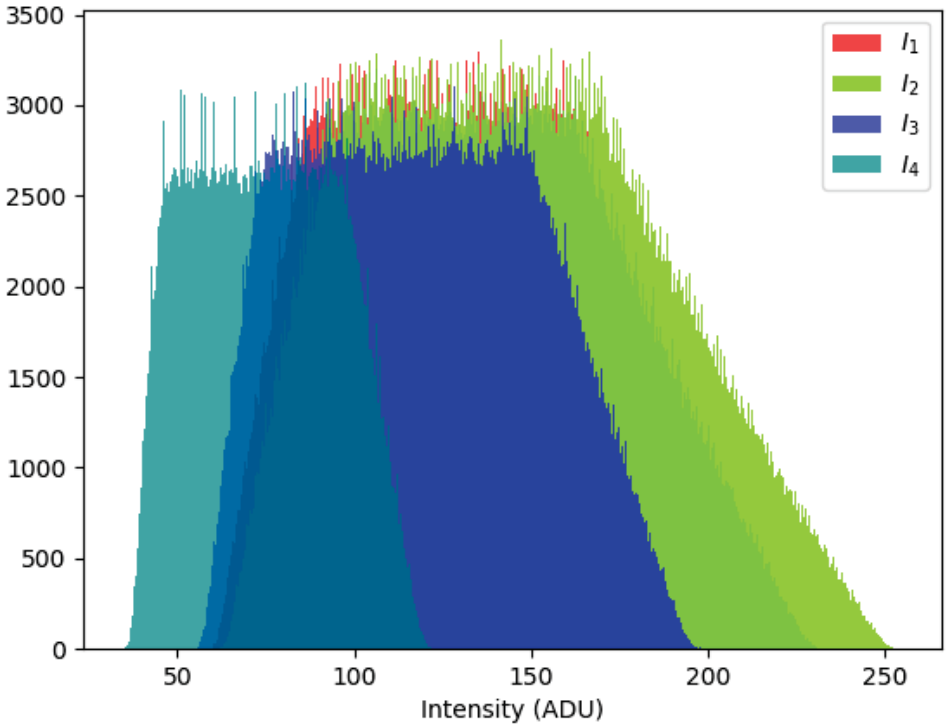

Figure 2.6: Histograms of the individual intensity distributions used for training the DNN to be used for WSI.

set, T_Y, is simply the P OPLs repeated K times:

$$T_Y = \begin{pmatrix} L_{11} & \dots & L_{1P} \\ \vdots & \ddots & \vdots \\ L_{K1} = L_{11} & \dots & L_{KP} = L_{1P} \end{pmatrix}$$

The ground truth data T_Y is normalized into the range $[0, 1]$ before being passed into the DNN for training. The input data T_X is clipped into the range $[0, 255]$, reflecting the range of possible values in the WSI system. A typical set of noisy intensities (T_X) for $L = 6$ µm, $\alpha \in [50, 130]$ ADU, $V \in [0.3, 0.95]$ is shown in Fig. 2.6. The code needed to generate the training data is provided for convenience in Appendix A.2.1.

Chapter 3

Performance Validation

3.1 Simulation Results

We will consider particular training sets and determine the sensitivity of our DNN. In each case the network(s) were trained, and then tested on a new data set to judge the performance. The output of the network(s), L_{DNN}, is then recorded. To calculate a network's sensitivity, a Monte Carlo simulation was performed with α, V fixed and L allowed to vary. The Poisson shot noise statistics are also included in this test set. The resulting $\vec{I}_{ij}$ in T_X were passed into the network (with $K = 25,000$) with shot noise simulated as described previously. Thus, for each OPL, the DNN tries to determine the true OPL while the intensities change due to the simulated shot noise. The standard deviation, $\sigma_{DNN} = \sqrt{\mathrm{Var}(L_{DNN})}$, of the resulting $25,000$ outputs by the DNN was then taken. For a highly sensitive DNN, each of the K values output by the DNN should be similar. This process is repeated for each OPL in the test set (in this case, 500 different OPLs) to produce the results shown in Fig. 3.1. Each point along the σ_{DNN} curve in Fig. 3.1 represents the standard deviation of the $25,000$ outputs of the DNN at each OPL plotted. Note that σ_{DNN} closely follows the CRB predicted by Eq. (2.6), and outperforms the analytical approach across much of the range. Figure 3.2 shows the outputs from the NNs trained including noise for two particular OPLs ('L_{true}') as a histogram. Note that the mean DNN output, L_{avg}, is very close to true value in both cases. The sensitivity is simply the standard deviation of the Gaussian curves in Fig. 3.2

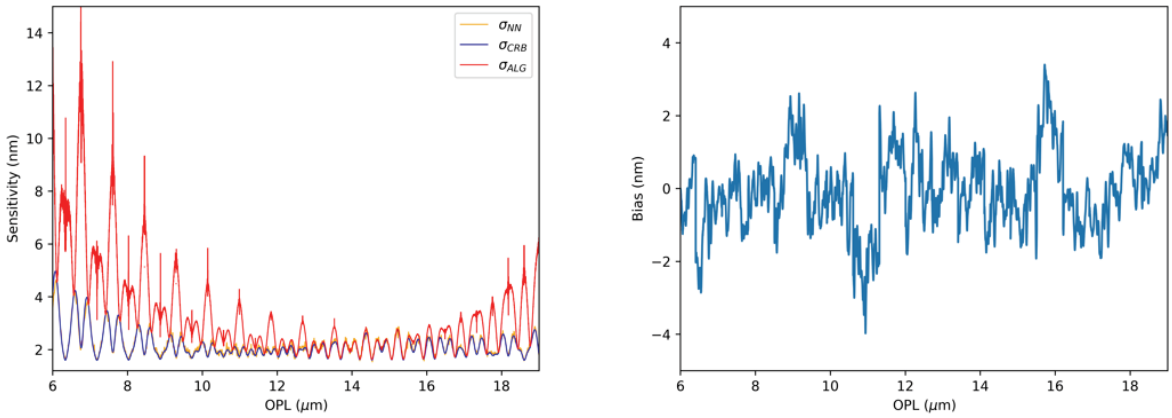

Figure 3.1: **Left:** Sensitivity vs. OPL plot including the DNN sensitivity. The blue (σ_{CRB}) and orange (σ_{DNN}) traces are nearly identical, suggesting the DNN achieves the lower bound set by Eq. (2.6). **Right:** Bias (average error) of the neural network vs. OPL. Across much of the range, the neural network is effectively unbiased (< 5 nm absolute bias).

and precisely corresponds to the sensitivities along the σ_{DNN} curve in Fig. 3.1.

To generate these results, multiple neural networks were trained on overlapping ranges for L (with the same ranges for α, V), each with width $L_{max} - L_{min} = 0.8$ µm (e.g. one network corresponds to $L \in [7, 7.8]$ µm, the next to $L \in [7.7, 8.5]$ µm, and so on). This is done to avoid periodic ambiguities within the intensity data. In general, for optimal sensitivity, we have found that the range of OPLs trained on for each network must satisfy $L_{max} - L_{min} \leq 2\pi / \max(\mathrm{k_n})$. As such, careful selection of neural network is required when applied to experimental data. In practice, this is not a strict limitation as many biological samples can be made to have an optical thickness of less than 0.8 µm OPL. Furthermore, the improvement in sensitivity compensates for this restriction, with a sensitivity increase of nearly an order of magnitude at lower L (see Fig. 3.1). Each range for parameters $[\alpha, V, L]$ has length $K = P = 1000$ for a total of 10^6 (T_X, T_Y) pairs for each network. In Fig. 3.1, the ranges of α and V for both sets of networks were $[70, 140]$ ADU and $[0.59, 0.95]$, respectively. These ranges cover many values typically observed in our system. Other systems may require training over other ranges.

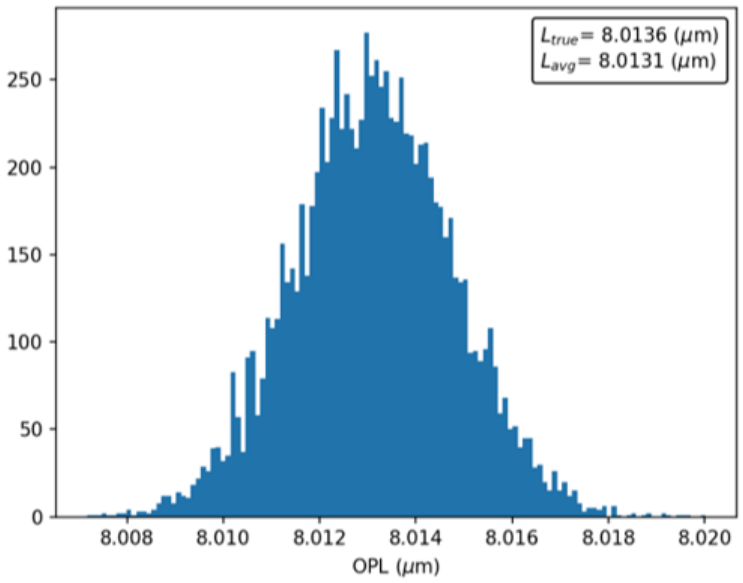

Figure 3.2: Histogram of DNN output for a true OPL in the test set of 8.0136 µm. The mean of this curve minus L_{true} is the *bias*, and the standard deviation is σ_{NN} in Fig. 3.1 for this OPL.

Our networks were implemented in Python using the Keras [36] deep learning library with Tensorflow GPU, and the ADAM optimizer was used [35] with learning rate $\beta = 0.001$. The resulting training time for a single DNN on an Intel (Santa Clara, California) Core i7-7700HQ CPU @ 3.5 GHz and an NVIDIA (Santa Clara, California) GTX 1070 GPU is around 7 minutes.

3.1.1 Randomly Generated OPLs

We performed an additional simulation where the testing data for the DNN is generated differently from the training data. In the training data, the OPL vector is a linear, evenly spaced vector of different values over a 0.8 µm range. For the previously presented simulated results, the test set was also a linear, evenly spaced vector of L values. To demonstrate our network's robustness to different distributions of L [37, 38], we randomly generate (by

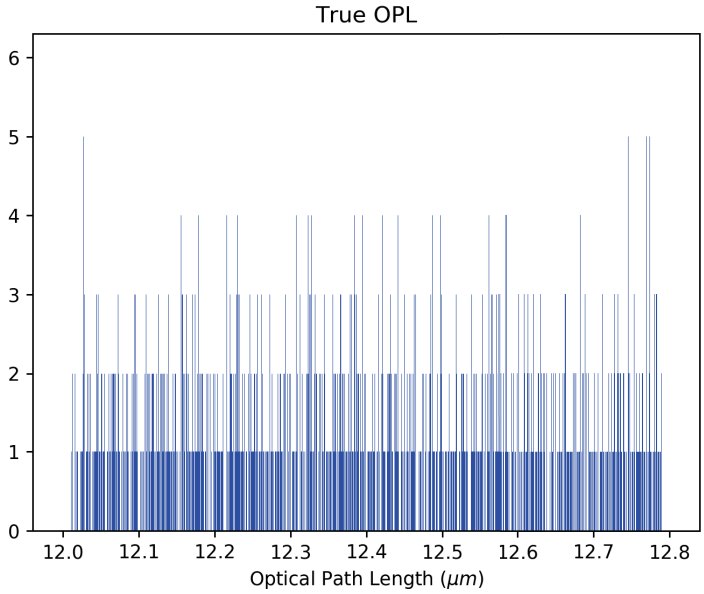

Figure 3.3: Randomly generated OPLs to create a new test set. These OPLs were generated by sampling a uniform distribution.

sampling a uniform distribution) a vector of L values to test the NNs on. Figure 3.3 shows the new, randomly generated OPLs. 2000 OPLs were sampled within the range $L \in [12, 12.8]$µm, and 2000 noise realizations were generated for each OPL giving a total test set size of $4 \cdot 10^6$. These OPLs are then used in Eq. (2.2) with $\alpha = 128$ ADU, $V = 0.7$ (average values for these parameters) to produce the noise-free intensities $\bar{I}_n$. These are then converted to the noisy intensities according to the procedure detailed in Sec. 2.3.1. The output of the DNN is then compared to the 'true' OPL in fig. 3.3 to compute the bias, which is shown in 3.4.

This result shows that the DNN remains accurate even with a significantly different set of data from which is was trained on. This point is reinforced when we apply the NNs to experimental data, where $L(x, y)$ is an unknown random distribution depending on the sample being imaged. In this case, the NNs are still able to accurately demodulate $L(x, y)$ despite not being trained on experimental data. See Secs. 3.2.2, 3.2.3. Overall, these results suggest that the NNs are able to reach the fundamental sensitivity limit for WSI, the first

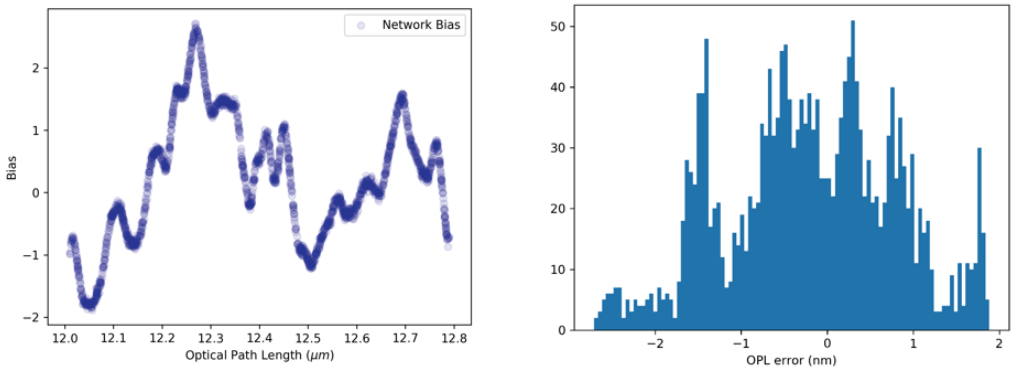

Figure 3.4: Left: Bias of the neural network vs. OPL for the randomly generated test set plotted as a scatter plot. The darker regions correspond to the points in the data set with the same OPL. Right: The bias for the network plotted as a histogram to show that, on average, the error is close to 0 nm.

algorithm known to be able to accomplish this. It has been known since at least the early 90s that neural networks are capable of achieving the CRB when trained on simulated data [39], but this is the first application and evidence to suggest that NNs are capable of achieving the CRB in the context of optical interferometry. The simulations are encouraging, but we must attempt to validate these results by taking raw intensity data from a WSI system and using it as the input for our NNs. These results are shown in the next section.

3.2 Experimental Results

3.2.1 Experimental Sensitivity

To further validate our NNs using experimental data, we make use of the WSI system illustrated in Fig. 2.1 with a blank sample (i.e. a glass cover slip) with $L(x, y) = L_S - L_R$ tuned to particular, fixed values. As the first example, we tuned the OPL difference between the two arms of the interferometer to $\bar{L} \approx 9.8$ µm and made 2000 consecutive acquisitions to produce 500 phase images via our NNs and Eq. (2.3). The resolution of our camera

is 0.3 megapixels (484 × 644). To clarify which sensitivities are which, we refer to the experimentally obtained sensitivities as $(\cdot)_{EXP}$. For example, the experimental sensitivity associated with Eq. (2.3) is $(\sigma_{Carre})_{EXP}$. At this OPL, we expect that based on Fig. 2.4 at 9.8 µm OPL that Eq. (2.3) will be comparable to the DNN, but that the DNN will have a slight advantage. The results are shown in Fig. 3.5, where based on the histogram, $(\sigma_{DNN})_{EXP}$ indeed has the best sensitivity distribution, being narrower and centered at a lower sensitivity than $(\sigma_{Carre})_{EXP}$. These results partially confirm the inequality presented in Eq. (2.2.3). Additionally, they suggest that the NNs are capable of providing an efficient OPL demodulation. This is significant, especially whenever the total OPL in the system is such that the sensitivity of Eq. (2.3) is extremely poor. The processing time for each of the 500 OPL images was about 2.2s on our PC, which is slower than the Carré equation (as it is an analytical solution), but still quite fast. We note that all of the experimental results in this work have had the background phase variations eliminated (subtracted) via Zernike polynomial based background subtraction [40]. This is done to eliminate some of the constant background fluctuations, which can improve temporal sensitivity.

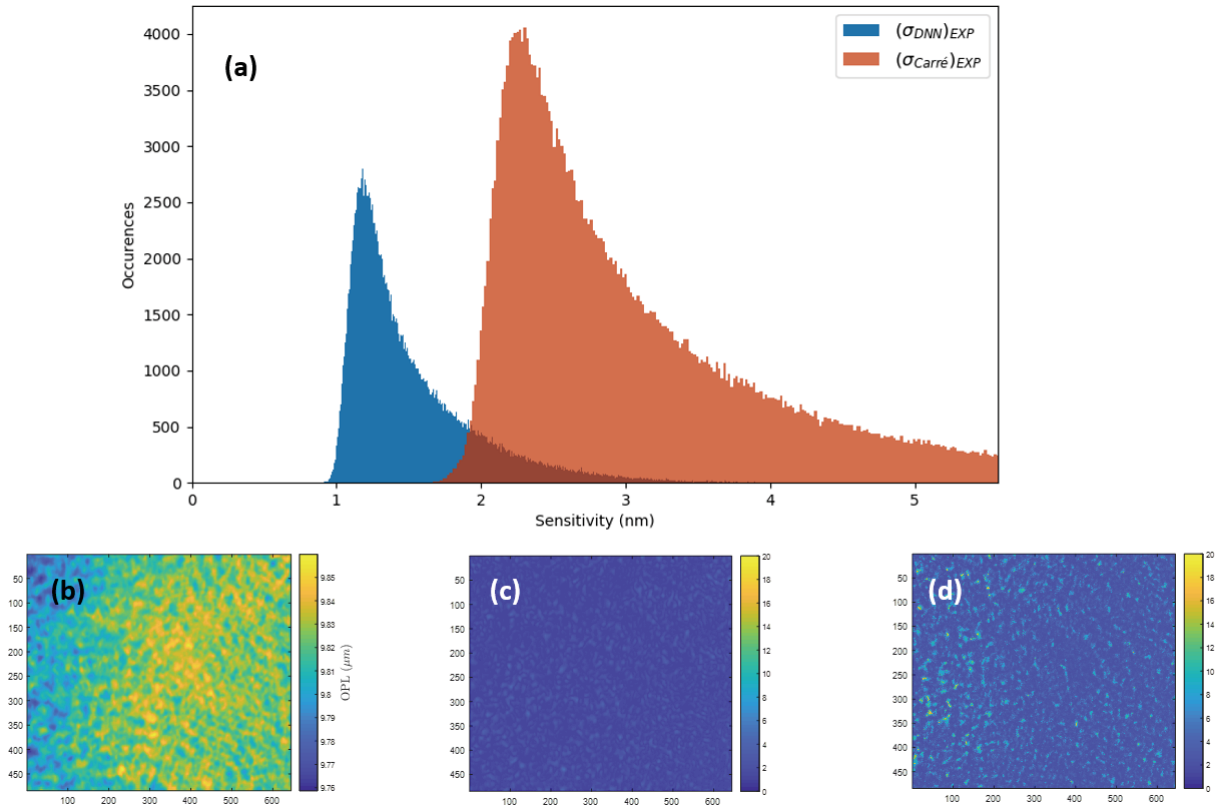

Figure 3.5: (a) Histogram of the sensitivity distribution of the entire field of view for both Carré and the DNN. The DNN has a much narrower distribution as well as a lower mean sensitivity. (b) Average OPL output of the DNN over 500 images, showing a distribution of OPL across the blank sample with mean value ≈ 9.8 µm, as expected. (c) Sensitivity distribution for the DNN. (d) Sensitivity distribution given by Eq. (2.3).

A more granular view of the differences between the two methods can be obtained by looking at the OPL output of a single pixel across all 500 OPL images and plotting the output of the two methods together on a histogram. This is shown in Fig. 3.6 for the center pixel in our field of view, where the difference between the DNN and the Carré equation are quite apparent. The spread (or variance, sensitivity) of the distribution associated with the Carré equation is much broader, leading to the lower sensitivity when compared with the DNN. The DNN, on the other hand, has a much tighter distribution and is thus more consistent in demodulating the OPL at the center pixel than Carré.

It is worth noting that the sensitivity map shown in Fig. 3.5c together with the histogram in Fig. 3.5a serve to partially confirm the results predicted in Fig. 3.1, where for an OPL of $\approx$ 9.8 µm we should in fact expect that the DNN will have a sensitivity around 2 nm. Likewise, the Carré equation has approximately 3.5 nm sensitivity for the same OPL, showing good agreement between simulation and experiment. This is particularly remarkable for the DNN, as we have trained the DNN on purely simulated data based on a relatively simple signal model (Eq. 2.2) paired with a simple shot noise model (see 2.2.2).

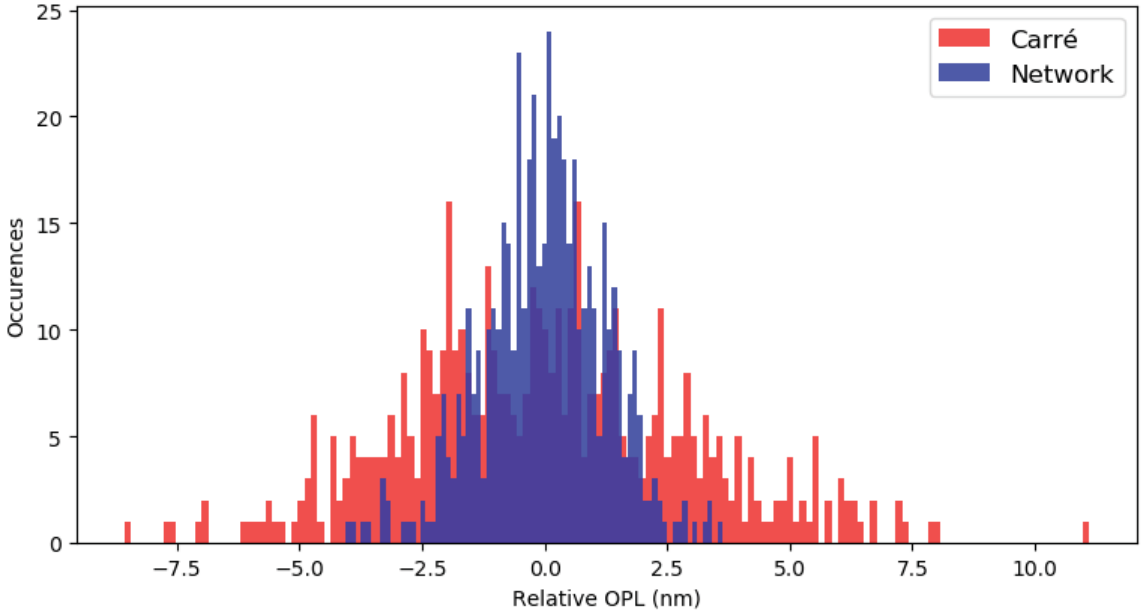

Figure 3.6: Histograms of the output of the Carré equation and the DNN for the center pixel in the field of view across 500 phase images. Shaded regions corresponds to the points at which the two methods agree. Note the much narrower distribution for the DNN.

3.2.2 Red Blood Cells

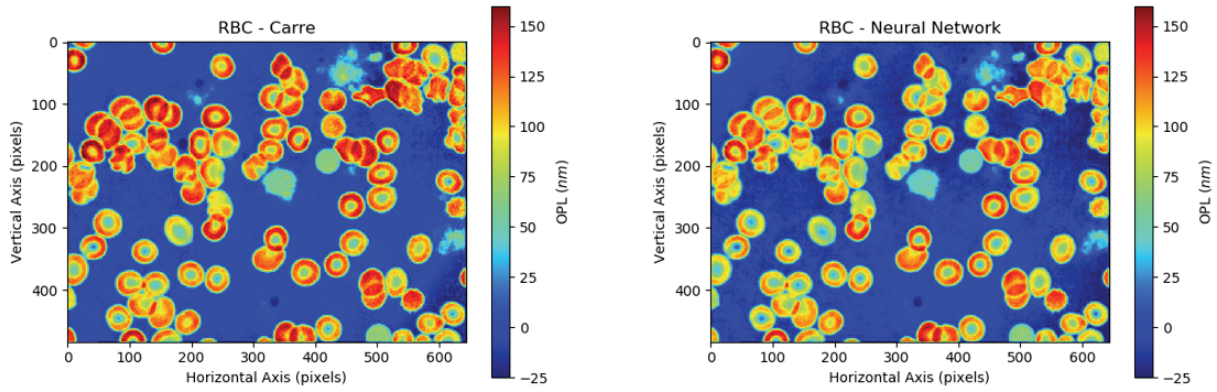

Figure 3.7: Left: Phase image of RBCs produced using Eq. 2.3. Right: Phase image of RBCs produced by a DNN trained to encompass the range $L \in [13, 13.8]$ µm. The two images are nearly identical, as expected.

In this section we demonstrate the use of NNs in obtaining cell images, in particular red blood cells (RBCs). The reference or background OPL in the system for the phase images shown in Fig. 3.7 was again tuned to 13.5 µm, meaning that based on Figs. 2.4, 3.1 we expect that the output of the DNN and Eq. (2.3) to produce a very similar phase image, and this is indeed the case. To obtain these images, we simply acquire the experimental data from the WSI system and pass it into a DNN trained on data correspondind to an OPL range which contains 13.5µm. We can further confirm that the DNN and the Carré equation are outputting similar phase distributions by plotting a line across the images in Fig. 3.7 together to evaluate the differences between the two approaches. This plot is shown in Fig. 3.8, in which we the two approaches output a nearly perfectly matched OPL distribution across row 409 of the field of view. This quantitatively confirms the visual (qualitative) similarity present in the two RBC images.

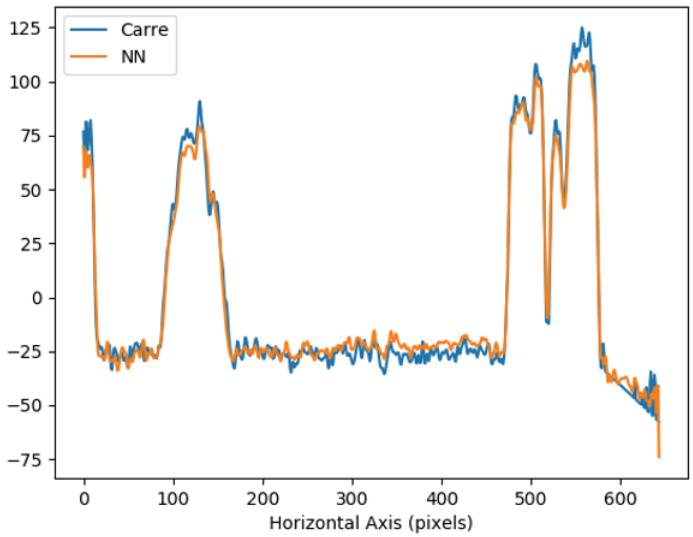

Figure 3.8: Row 409 of both RBC images in Fig. 3.7 are plotted together for all columns of the image (644). THe two approaches are in very good agreement, with only minor differences between Carré and the NN across the image, confirming our qualitative observations.

We also performed a RBC experiment where the reference OPL was set to $L = 7\mu m$. This time, we have used a DNN trained on data corresponding to an OPL range which contains 7µm (for example, $L \in [6.7, 7.5]\mu m$). In this region, we expect that the DNN will output a better image due to the enchanced sensitivity of the DNN. The result is shown in fig. 3.9. This time, we see a clear advantage for the DNN when compared to the traditional, analytical approach. The Carré equation is beginning to fail due to the increasingly poor sensitivity as the OPL approaches zero.

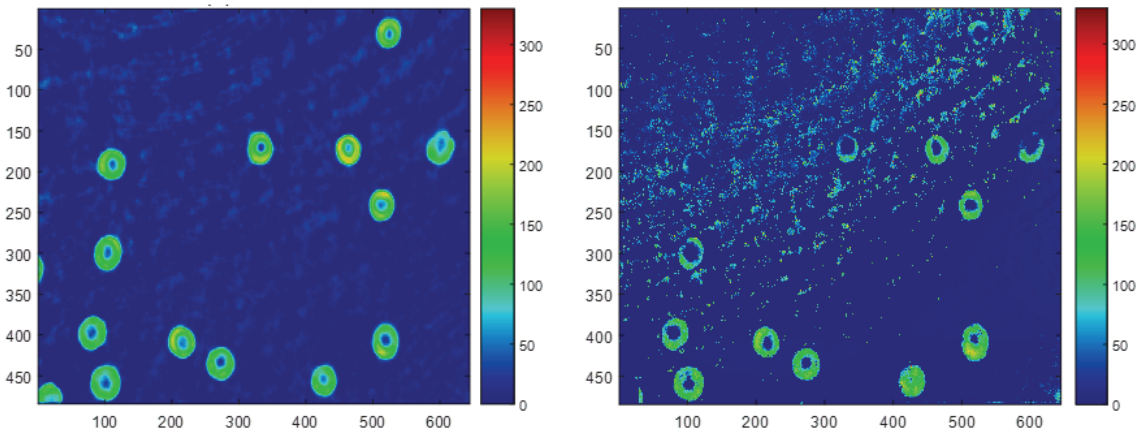

Figure 3.9: **Left**: RBC image produced by the DNN, showing a relatively smooth background with well defined RBC. **Right**: RBC image generated by Eq. (2.3), where we can see some artifacts due to the use of the Carré equation starting to appear. Additionally, some of the RBCs visible in the DNN image are difficult to see in this image.

3.2.3 Cheek Cells

Finally, we consider a comparison between the two techniques by imaging a human cheek cell. A human cheek cell has an OPL distribution that is quite different from the RBCs just presented, so this experiment represents a good test of our DNN's robustness to different biological samples. This experiment was performed with the background OPL set to $\approx 7\mu m$ as in the previous RBC image. The results are presented in fig. 3.10, where we can clearly see the DNN outperforming the Carré equation by resolving the details of the cheek cell more clearly. The Carré equation has recovered most of the important features, but they are marred by an overall grainy, discontinous representation due to the reduced sensitivity of Eq. (2.3).

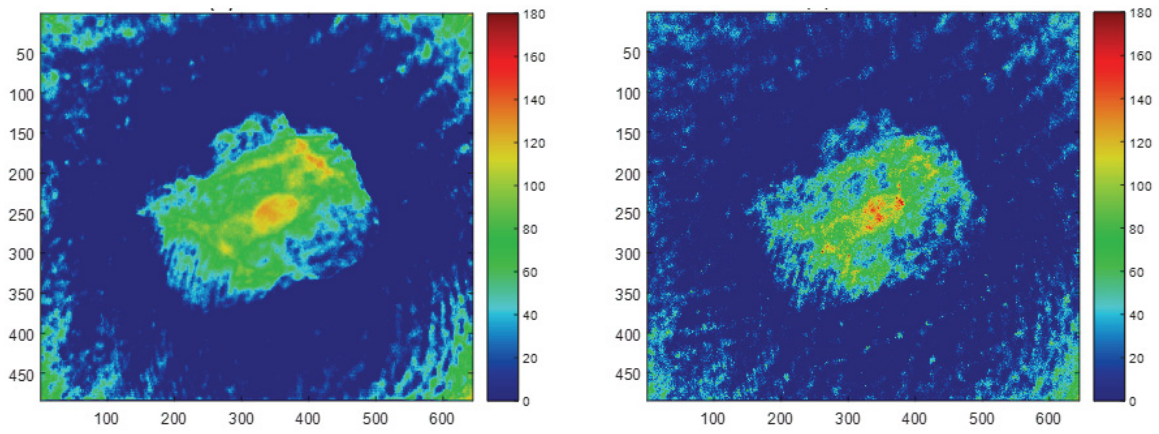

Figure 3.10: **Left**: Cheek cell image from the DNN. **Right**: RBC image generated by Eq. (2.3), where we can see some artifacts due to the use of the Carré equation starting to appear. Additionally, some of the RBCs visible in the DNN image are difficult to see in this image.

Chapter 4

Summary

We have introduced a deep neural network based technique for the demodulation of optical pathlength (phase) in a quantitative phase imaging system (QPI). The concept of quantitative phase imaging was discussed and contrast was provided between QPI and traditional qualitative techniques such as Zernike phase contrast microscopy. Then, we specified the objective of QPI which is to calculate phase distributions. The particular type of phase distributions we are interested were introduced through the example of plane wave solutions to Maxwell's equations in free space. The phase of interest is the phase of electromagnetic waves. Typically, it is the difference in phase between two optical waves (for example, a 'reference' wave and a 'sample' wave) which is of interest in QPI. This leads to a natural discussion of wave optics and in particular the phenomenon of wave interference. Interference can be used to determine the phase difference between waves, and this practice of extracting information from the interference of waves is known generally as optical interferometry. We analyzed the simple case of a two-beam interferometer known as the Michelson interferometer and studied some general properties of interference which apply to most two-beam interferometric configurations. This simple introduction was presented in the context of monochromatic light sources, which simplifies the analysis somewhat. In reality, light sources are never perfectly monochromatic, making the simple wave optics analysis somewhat limited. A different perspective was presented in the context of Statistical Optics, which leads to a natural definition of temporal coherence which provides insight as to when we can expect to observe

interferograms of the kind produced by the interference of two monochromatic optical waves. This concept is introduced and some of the effects are illustrated in a basic example.

We then discuss a QPI technique known as Wavelength Shifting Interferometry, where multiple interferograms are taken successively to obtain the phase via the Carré equation. This technique has been proven to be highly sensitive within a certain range of OPLs, being comparable to Cramér-Rao bound. However, outside this range, the sensitivity degrades and no longer achieves the CRB. This suggests the possibility for other algorithms to fill this gap and achieve the CRB for a much wide range of OPL. Our approach was to use a fully connected deep neural network to achieve the CRB no matter the total OPL in the system. To this end, we provided a brief review of DNN concepts and then described our training data process and how we will use the DNN in the context of WSI. By incorporating the shot noise-limited nature of our WSI system into our training data, we were able to show in simulation that our NNs approach the CRB for all simulated OPLs, providing a significant advantage over the Carré equation in many cases. We then validated some of our simulations by performing experiments with an implementation of a 4-band WSI system. The experiments confirmed our simulations to a large degree, although the entirety of the simulated sensitivity vs. OPL curve was not able to be validated due to some constraints of our WSI implementation. We demonstrated that the DNN has a distinct advantage over the traditional analytical technique at $L = 9.8$ µm. We also established parity between the two techniques in the case of imaging Red Blood Cells with total system OPL ≈ 13.5 µm.

Based on the experimental and simulation results, we can conclude that NNs offer an efficient solution for obtaining the optical pathlength in the context of Wavelength Shifting Interferometry. Not only do the NNs achieve the fundamental sensitivity limit predicted by the Cramér Rao-bound, a first in the field, but they do so accurately and without a significant loss in processing speed. These results suggest that DNN could be an extremely

valuable tool for demodulating optical pathlength in a variety of optical interferometry systems. We have only analyzed a single example of WSI in this work, but due to the generality of our signal model, we believe these techniques can be easily extended to some other QPI techniques and may provide similar gains in sensitivity.

Two conference publications were generated while developing this research:

1. J.A. Black, J.G. Thomas, S. Chen, Y. Zhu, "Performance assessment of interferometric sensing," Proc. SPIE 11000, Fiber Optic Sensors and Applications XVI, 110000C (14 May 2019).

2. J. A. Black, S. Chen, and Y. Zhu, "Neural Network for Optical Pathlength Demodulation in Wavelength Shifting Interferometry ," Frontiers in Optics, Optical Society of America, (Sep 2019).

Bibliography

[1] Shichao Chen, Junghyun Ryu, Kiho Lee, and Yizheng Zhu. Swept source digital holographic phase microscopy. *Opt. Lett.*, 41(4):665–668, Feb 2016.

[2] Shichao Chen, Chengshuai Li, and Yizheng Zhu. Sensitivity evaluation of quantitative phase imaging: a study of wavelength shifting interferometry. *Opt. Lett.*, 42(6):1088–1091, Mar 2017.

[3] Zhuo Wang, Larry Millet, Mustafa Mir, Huafeng Ding, Sakulsuk Unarunotai, John Rogers, Martha U. Gillette, and Gabriel Popescu. Spatial light interference microscopy (slim). *Opt. Express*, 19(2):1016–1026, Jan 2011.

[4] D. Gillies, W. Gamal, A. Downes, Y. Reinwald, Y. Yang, A.J. El Haj, and P.O. Bagnaninchi. Real-time and non-invasive measurements of cell mechanical behaviour with optical coherence phase microscopy. *Methods*, 136:126 – 133, 2018. Methods in Quantitative Phase Imaging in Life Science.

[5] Doyeon Kim, Nuri Oh, Kyoohyun Kim, SangYun Lee, Chan-Gi Pack, Ji-Ho Park, and YongKeun Park. Label-free high-resolution 3-d imaging of gold nanoparticles inside live cells using optical diffraction tomography. *Methods*, 136:160 – 167, 2018. Methods in Quantitative Phase Imaging in Life Science.

[6] Gabriel Popescu. *Quantitative Phase Imaging of Cells and Tissues*. 01 2011.

[7] Jay Nadeau, YongKeun Park, and Gabriel Popescu. Methods in quantitative phase imaging in life science. *Methods*, 136:1 – 3, 2018. Methods in Quantitative Phase Imaging in Life Science.

[8] Christopher A. Metzler, Philip Schniter, Ashok Veeraraghavan, and Richard G. Baraniuk. prdeep: Robust phase retrieval with a flexible deep network. In *ICML*, 2018.

[9] David Jeffrey Griffiths. *Introduction to electrodynamics*. Cambridge University Press, 2018.

[10] Jian-Ming Jin. *Electromagnetic field theory and computation*. Wiley-Blackwell, 2010.

[11] Bahaa E. A. Saleh. *Fundamentals of photonics*. Wiley, 2019.

[12] Ruibo Shang, Shichao Chen, Chengshuai Li, and Yizheng Zhu. Spectral modulation interferometry for quantitative phase imaging. *Biomedical optics express*, 6(2):473–479, Jan 2015. 25780737[pmid].

[13] Joseph W. Goodman. *Statistical Optics*. Wiley, 2015.

[14] Alberto Leon-Garcia. *Probability, Statistics, and Random Processes for Electrical Engineering*. Pearson/Prentice Hall, Upper Saddle River, NJ, third edition, 2008.

[15] Joseph W. Goodman. *Introduction to Fourier optics*. Freeman, 2017.

[16] Takeshi Takahashi, Yukihiro Ishii, and Ribun Onodera. Phase-shifting interferometric profilometry with a wavelength-tunable diode source. *Optical Review*, 21(3):410–414, May 2014.

[17] Kimio Tatsuno and Yoshito Tsunoda. Diode laser direct modulation heterodyne interferometer. *Appl. Opt.*, 26(1):37–40, Jan 1987.

[18] Yukihiro Ishii. Laser-diode interferometry. volume 46 of *Progress in Optics*, pages 243 – 309. Elsevier, 2004.

[19] Xiaobo Tian, Xingzhou Tu, Kimiko Della Croce, Guang Yao, Haijiang Cai, Neal Brock, Stanley Pau, and Rongguang Liang. Multi-wavelength quantitative polarization and phase microscope. *Biomed. Opt. Express*, 10(4):1638–1648, Apr 2019.

[20] Shichao Chen, Chengshuai Li, and Yizheng Zhu. Low-coherence wavelength shifting interferometry for high-speed quantitative phase imaging. *Opt. Lett.*, 41(15):3431–3434, Aug 2016.

[21] C. Li and Y. Zhu. Cramer–rao bound for frequency estimation of spectral interference and its shot noise-limited behavior. *IEEE Journal of Selected Topics in Quantum Electronics*, 23(2):410–416, March 2017.

[22] Mitsuo Takeda, Hideki Ina, and Seiji Kobayashi. Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *J. Opt. Soc. Am.*, 72(1):156–160, Jan 1982.

[23] Pedro Americo Almeida Magalhaes Junior, Perrin Smith Neto, and Clovis Sperb de Barcellos. Generalization of carré equation. *Optik*, 122(6):475 – 489, 2011.

[24] Pedro Americo Almeida Magalhaes, Perrin Smith Neto, and Cristina Almeida Magalhães. A modified carre algorithm for phase shifting interferometry. *Journal of Optics*, 39(1):5–21, Mar 2010.

[25] Chuen-Lin Tien, Siao-Shan Jyu, and Huei-Min Yang. A method for fringe normalization by zernike polynomial. *Optical Review*, 16:173–175, 03 2009.

[26] D. Rife, M. Goldstein, and R. Boorstyn. A unification of cramér-rao type bounds (corresp.). *IEEE Transactions on Information Theory*, 21(3):330–332, May 1975.

[27] Amnon Yariv and Pochi Yeh. *Photonics: optical electronics in modern communications*. Oxford Univ. Press, 2009.

[28] Athanasios Margaris and Miltiadis Adamopoulos. Solving nonlinear algebraic systems using artificial neural networks. *CEUR Workshop Proceedings*, 284, 01 2007.

[29] Mo Deng, Shuai Li, and George Barbastathis. Learning to synthesize: splitting and recombining low and high spatial frequencies for image recovery. *CoRR*, abs/1811.07945, 2018.

[30] Michael T. Manry, Steven J. Apollo, and Qiang Yu. Minimum mean square estimation and neural networks. *Neurocomputing*, 13(1):59 – 74, 1996.

[31] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *CoRR*, abs/1611.02862, 2016.

[32] JONATHAN BARZILAI and JONATHAN M. BORWEIN. Two-Point Step Size Gradient Methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 01 1988.

[33] Pavlo M. Radiuk. Impact of training set batch size on the performance of convolutional neural networks for diverse datasets. *Information Technology and Management Science*, 20, 12 2017.

[34] Herbert Robbins and Sutton Monro. A stochastic approximation method. *Ann. Math. Statist.*, 22(3):400–407, 09 1951.

[35] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.

[36] François Chollet et al. Keras. https://keras.io, 2015.

[37] X. Chai, Q. Ba, and G. Yang. Characterizing robustness and sensitivity of convolutional neural networks in segmentation of fluorescence microscopy images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3838–3842, Oct 2018.

[38] P. Kostyla, T. Lobos, and Z. Waclawek. Neural networks for real-time estimation of signal parameters. In *Proceedings of IEEE International Symposium on Industrial Electronics*, volume 1, pages 380–385 vol.1, June 1996.

[39] J. Ting-Ho Lo. Synthetic approach to optimal filtering. *IEEE Transactions on Neural Networks*, 5(5):803–811, Sep. 1994.

[40] Vasudevan Lakshminarayanan and Andre Fleck. Zernike polynomials: A guide. *Journal of Modern Optics - J MOD OPTIC*, 58:1678–1678, 04 2011.

Appendices

Appendix A

A.1 Algorithmic Sensitivity Of The Carré Equation

In this section we provide the full expression for σ_{ALG} for the solution of $L(x, y)$ given by Eq. (2.3). Recall that we had:

$$\mathrm{Var}(L_{ALG}) = \sum_{n=1}^{4} \left(\frac{\partial L}{\partial I_n} |_{\bar{I}_n} \right)^2 \mathrm{Var}(I_n)$$

If we take the sqrt root of this equation and expand the sum, we obtain σ_{ALG} [2]:

$$\sigma_{ALG} = \sqrt{\mathrm{Var}(L)} = \frac{\sqrt{BC}A}{k_0(A^2 + BC)} \left\{ \begin{array}{c} (\frac{1}{A} - \frac{1}{2B} + \frac{1}{2C})\sqrt{\bar{I}_1/G} \\ +(-\frac{1}{A} + \frac{3}{2B} + \frac{1}{2C})\sqrt{\bar{I}_2/G} \\ +(-\frac{1}{A} - \frac{3}{2B} - \frac{1}{2C})\sqrt{\bar{I}_3/G} \\ +(\frac{1}{A} + \frac{1}{2B} - \frac{1}{2C})\sqrt{\bar{I}_4/G} \end{array} \right\} \tag{A.1}$$

where $k_0 = (k_1 + k_4)/2$, and $A = \bar{I}_2 + \bar{I}_3 - \bar{I}_1 - \bar{I}_4$, $B = 3\bar{I}_2 - 3\bar{I}_3 - \bar{I}_1 + \bar{I}_4$, and $C = \bar{I}_2 - \bar{I}_3 + \bar{I}_1 - \bar{I}_4$. This gives the algorithmic sensitivity as function of mean (noise-free) intensities $\bar{I}_n$, which makes this expression particular convenient for performing simulations, as was done in much of this work.

A.2 DNN Model Specification & Data Generation

In this appendix we give the model specification as Python code. The framework is Keras with Tensorflow.

A.2.1 Data Generation

```
1  from numpy import asarray, sin, cos, sqrt, linspace, pi, arctan2,
       unwrap, hstack, zeros, ones, floor, matmul, repeat, clip, tile
2  from numpy.random import randn, randint, uniform
3  from scipy.stats import poisson
4
5  def trainingDataGeneration(min_a, max_a, min_vis, max_vis, min_opl,
       max_opl, noise=True):
6          I_train = []
7          P_carre = []
8          I_crb   = []
9          L_train = []
10         num_samples = int(1e3) # Number of noise-samples to go with
               each OPL.
11         train_size = int(1e3) # Number of OPLs in the training set
12         G = 34.4 # Camera gain in e-/ADU.
13
14         # Wavelengths to "sample" at
15         wl = asarray([869.65, 853.45, 837.85, 822.8])*1e-9
16         k = 2*pi/wl
17
```

```
18          L = linspace(min_opl, max_opl, train_size)
19
20          L = asarray([repeat(L, num_samples)]) # Repeat the OPL Vector '
                num_samples' times, corresponding to the number of noise
                realizations.
21
22          phase = matmul(k.T, L) # Create a matrix with shape (
                num_samples*train_size, 4) corresponding to the phase for
                each data point
23
24          a = floor(uniform(min_a, max_a, num_samples*train_size)) #
                Generate random alpha values for each point
25
26          V = uniform(min_vis, max_vis, num_samples*train_size) #
                Generate a random visibility for each point
27
28          I_true = a*(1 + V*cos(phase)) # Noise free intensities. Shape
                is same as a, V, and Phase.
29
30          s_true = I_true*G # Noise free pre-A/D signal
31
32          if noise:
33              #I_train = floor(I_true + multiply(sqrt(I_true*G), randn(
                    phase.shape[0], phase.shape[1])/G)).T # Gaussian approx
34              x_n = poisson.rvs(s_true).T # Simulated noisy signal on the
                     sensor
35              I_train = x_n/G # Noisy intensity
```

```
36          else:
37              I_train = I_true.T
38
39          I_train = clip(I_train, 0, 255) # Clip signal in case the noise
                resulted in values outside [0, 255]
40
41          L_train = L
42          return [L_train, I_train]
```

A.2.2 DNN Model

```
43              opt = Adam(lr=0.001) # Adam optimizer initializtion
44              model = Sequential()
45              model.add(Dense(1024, input_dim=4, activation='sigmoid'))
46              model.add(Dropout(0.01))
47              model.add(Dense(512, activation='sigmoid'))
48              model.add(Dense(256, activation='sigmoid'))
49              model.add(Dense(128, activation='sigmoid'))
50              model.add(Dropout(0.01))
51              model.add(Dense(32, activation='sigmoid'))
52              model.add(Dense(16, activation='sigmoid'))
53              model.add(Dense(1))
```