# Computational Imaging for Phase Retrieval and Biomedical Applications

Thesis by

Cheng Shen

In Partial Fulfillment of the Requirements for

the Degree of

Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2023

Defended May 22, 2023

© 2023

Cheng Shen
ORCID: 0000-0001-7136-4715

# ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Changhuei Yang, for his invaluable guidance and unwavering support throughout my PhD study. He consistently set a high standard for research quality and encouraged me to aim for excellence in all aspects of my work. Whenever I faced technical challenges, he patiently brainstormed with me to find effective solutions. His creativity and innovative thinking have inspired me to become a better researcher. I am particularly grateful to him for his guidance in developing my communication and writing skills, which were areas where I felt least confident prior to starting my PhD. Without his help, I would not have been able to achieve what I have accomplished today.

I am also grateful to my committee members, Professors Lihong Wang, Magdalena Zernicka-Goetz, and Katherine L. Bouman for their precious support throughout my research journey. Professor Lihong Wang's biomedical imaging course provided me with a comprehensive overview of different classical imaging modalities, and I gained numerous insights from his teachings. Professor Katherine L. Bouman's computational imaging course perfectly aligned with my research interests and inspired me greatly. Additionally, I had the honor of working on a research project co-supervised by Professor Magdalena Zernicka-Goetz, and I am deeply appreciative of her domain knowledge and guidance. Her great passion and dedication to her work have always been a source of encouragement for me.

Members of the Biophotonics Lab have been instrumental in supporting me throughout my PhD study. I would like to express my heartfelt gratitude to Dr. Chi Shing (Antony) Chan for imparting his expertise in optical experimental skills and for teaching me how to formulate questions accurately and divide and conquer technical problems. I am also grateful to Dr. Michelle Cua for her selfless guidance and training, which helped me develop my critical thinking abilities. Her generous help during the initial stages of my PhD study, especially in setting up connections during the sample hunting for my first project, is something that I cannot thank her enough for. I would also like to extend my special thanks to Ruizhi Cao for his constant support during my hard time in the pandemic and our fruitful

# ABSTRACT

In conventional imaging, optimizing hardware is prioritized to enhance image quality directly. Digital signal processing is viewed as supplementary. Computational imaging intentionally distorts images through modulation schemes in illumination or sensing. Then its reconstruction algorithms extract desired object information from raw data afterwards. Co-designing hardware and algorithms reduces demands on hardware and achieves the same or even better image quality. Algorithm design is at the heart of computational imaging, with model-based inverse problem or data-driven deep learning methods as approaches. This thesis presents research work from both perspectives, with a primary focus on the phase retrieval issue in computational microscopy and the application of deep learning techniques to address biomedical imaging challenges.

The first half of the thesis begins with Fourier ptychography, which was employed to overcome chromatic aberration problems in multispectral imaging. Then, we proposed a novel computational coherent imaging modality based on Kramers-Kronig relations, aiming to replace Fourier ptychography as a non-iterative method. While this approach showed promise, it lacks certain essential characteristics of the original Fourier ptychography. To address this limitation, we introduced two additional algorithms to form a whole package scheme. Through comprehensive evaluation, we demonstrated that the combined scheme outperforms Fourier ptychography in achieving high-resolution, large field-of-view, aberration-free coherent imaging.

The second half of the thesis shifts focus to deep-learning-based methods. In one project, we optimized the scanning strategy and image processing pipeline of an epifluorescence microscope to address focus issues. Additionally, we leveraged deep-learning-based object detection models to automate cell analysis tasks. In another project, we predicted the polarity status of mouse embryos from bright field images using adapted deep learning models. These findings highlight the capability of computational imaging to automate labor-intensive processes, and even outperform humans in challenging tasks.

# PUBLISHED CONTENT AND CONTRIBUTIONS

- R. Cao, **C. Shen**, and C. Yang, "High-resolution, large field-of-view label-free imaging via aberration-corrected, closed-form complex field reconstruction," *In Preparation*

  *C.S. participated in designing and constructing the optical system, prepared samples, conducted experiments, collected and analyzed the data, and participated in writing the manuscript.*

- **C. Shen**, S. Rawal, R. Brown, H. Zhou, A. Agarwal, M. A. Watson, R. J. Cote, and C. Yang, "Automatic detection of circulating tumor cells and cancer associated fibroblasts using deep learning," *Scientific Reports* 13: 5708 (2023). DOI: 10.1038/s41598-023-32955-0

  *C.S. participated in conceiving the project and planning the project direction, built up the imaging system, developed data processing algorithms as well as deep learning cell detection models, scanned samples, processed and analyzed the data, wrote the draft, and participated in editing the manuscript.*

- H. Zhou, **C. Shen**, M. Liang, and C. Yang, "Analysis of postreconstruction digital refocusing in Fourier ptychographic microscopy," *Optical Engineering* 61(7): 073102 (2022). DOI: 10.1117/1.OE.61.7.073102

  *C.S. participated in conceiving the project and discussing the project direction, built up the experimental setup, and participated in editing the manuscript.*

- **C. Shen**, A. Lamba, M. Zhu, R. Zhang, M. Zernicka-Goetz, and C. Yang, "Stain-free detection of embryo polarization using deep learning," *Scientific Reports* 12: 2404 (2022). DOI: 10.1038/s41598-022-05990-6

  *C.S. participated in conceiving the project and planning the project direction, designed and optimized the deep learning model, conducted human volunteer trials, and participated in interpreting results, writing and editing the manuscript.*

- **C. Shen**, M. Liang, A. Pan, and C. Yang, "Non-iterative complex wave-field reconstruction based on Kramers-Kronig relations," *Photonics Research* 9(6): 1003-1012 (2021). DOI: 10.1364/PRJ.419886

*C.S. conceived the project idea, conducted simulations, built up the imaging system, developed the reconstruction algorithm, conducted experiments, collected and analyzed the data, wrote the draft, and participated in editing the manuscript.*

- G. Zheng, **C. Shen**, S. Jiang, P. Song, and C. Yang, "Concept, implementations and applications of Fourier ptychography," *Nature Reviews Physics* 3(3): 207-223 (2021). DOI: 10.1038/s42254-021-00280-y

  *C.S. contributed to all aspects of manuscript preparation, revision and editing.*

- **C. Shen**, A. C. S. Chan, J. Chung, D. E. Williams, A. Hajimiri, and C. Yang, "Computational aberration correction of VIS-NIR multispectral imaging microscopy based on Fourier ptychography," *Optics Express* 27(18): 24923-24937 (2019). DOI: 10.1364/OE.27.024923

  *C.S. participated in conceiving the project and planning the project direction, built up the imaging system, developed aberration correction algorithms, prepared samples, conducted experiments, collected and analyzed the data, wrote the draft, and participated in editing the manuscript.*

# TABLE OF CONTENTS

# LIST OF ILLUSTRATIONS

# LIST OF TABLES

*C h a p t e r   1*

## INTRODUCTION

Imaging refers to the visual representation or reproduction of an object's profile. This process involves the formation of an image, which can be achieved using various materials and methods. The concept of imaging has been around for a long time and has always been important, even before it was formally introduced. In fact, the ability to see is a fundamental process of imaging. Throughout the course of human evolution, our species has developed a highly sophisticated imaging system in the form of our eyes. The human eye lens, for example, is a flexible and tunable lens that enables us to see with remarkable clarity, while the human retina is perhaps the most advanced camera sensor known to exist, capable of capturing images with exceptional dynamic range [1]. Scientists and engineers have long been fascinated by these mechanisms and have sought to replicate their functions through the creation of artificial imaging systems [2]. Thanks to advances in lens design and manufacturing, as well as the invention of digital camera sensors, it is now possible to achieve high-quality imaging even with a small camera module integrated into our cell phones. Furthermore, we have been able to extend our imaging capabilities beyond human vision, allowing us to observe objects on scales ranging from atoms [3] to stars [4].

Figure 1.1: Conventional imaging paradigm.

While imaging systems may have diverse structures and forms, they can be broken down into four basic elements: illumination, sample, optical system, and sensor, as illustrated in Figure 1.1. In the case of microscopic imaging, the illumination source may be a halogen light, while the sample could be a pathology slide. A simple bright field microscope's optical system comprises an objective lens and a tube lens, and the sensor could be a charge-coupled

device (CCD) or complementary metal-oxide-semiconductor (CMOS) camera. The final image is directly captured by the sensor device. To obtain high-quality images, all imaging conditions must be optimized. For instance, the illumination intensity should be high enough to ensure that the photon flux is sufficient to carry information about the sample. The optical system's resolution must be high enough to resolve the fine details in the microscope slide samples. Finally, the sensor should have a low noise level and high dynamic range to capture high signal-to-noise ratio (SNR) and contrast images. Achieving these requirements primarily depends on hardware system improvements. Optical engineers continuously optimize objective lens designs to pursue a large numerical aperture (NA), resulting in high optical resolution, without sacrificing field-of-view (FOV) [5]. Meanwhile, camera engineers have been improving sensor designs by enlarging the frame, reducing pixel size, and enhancing noise control [6].

As our demand for high-quality imaging continues to grow, conventional imaging paradigms become no longer sufficient. Current hardware manufacturing technology cannot keep up with our design specifications, and physical limits present additional obstacles. For instance, the resolution of optical imaging systems, such as microscopes or telescopes, may be limited by factors such as lens aberrations. Even if we eliminate aberrations through sophisticated lens design, there is still a fundamental limit to the resolution of any optical system due to the physics of wave diffraction. The diffraction-limited angular resolution of an instrument is proportional to the observed wavelength of light and inversely proportional to the NA of the imaging lens, as quantitatively described by the Abbe diffraction limit [7]. To surpass this limit and achieve imaging beyond the Abbe diffraction limit, we can consider challenging the current conventional imaging paradigm.

There is another prominent case in the coherent imaging domain. When the illumination source is coherent, like a laser, the sample's absorption will modulate the light amplitude, while its thickness and refractive index distribution will modulate the light field phase. Although absorption contrast is often the main focus, phase profile can be highly informative, especially in life science research [8]. Phase images of unstained living cells can reveal

structures that are otherwise invisible. However, it is impossible to directly measure the phase of the light wave field striking the sensor since the frequency of visible light far exceeds the response rate of even state-of-the-art cameras. Therefore, recovering a phase map from intensity measurements by camera sensors has become a topic of interest for optical imaging scientists in recent decades. In addition, imaging non-line-of-sight objects or in the wavelength range where commercial sensor arrays are still unattainable also requires a rethinking of the conventional imaging framework.
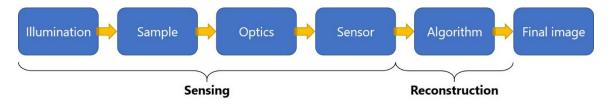


Figure 1.2: Computational imaging paradigm.

All these scenarios drive the emergence of the notion of computational imaging illustrated in Fig. 1.2. Notably, computational imaging involves an additional step between the sensor and the final image: the algorithm. Unlike simple image de-noising or color image de-mosaicking algorithms, computational imaging aims to co-design the sensing process and the reconstruction algorithm to enable the optical coding of information from the sample of interest, which is later decoded digitally. This approach reduces the requirements for hardware design and leverages the power of computation to solve challenging imaging problems. Mathematically, the hardware design can be viewed as a forward process, while the computational algorithms address the inverse problem, which is a popular topic in applied mathematics [9]. There are numerous mathematic tools available that can be utilized to solve the inverse problem.

Based on the methods to solve the inverse problem, current computational imaging techniques can be broadly categorized into two groups: model-based and data-driven methods.

Model-based methods require the mathematical depiction of the forward process [10]. If the desired sample information is denoted as $x$ and measurements are $y$, the forward process can be expressed as

$$A(x) = y, \tag{1.1}$$

where $A$ represents the system function. It relies on the accurate physical modeling for each imaging component. Ideally, we wish to find the inverse operator of $A$ so that

$$x = A^{-1}(y). \tag{1.2}$$

However, it is often challenging due to the nonlinearity of $A$. Moreover, the inverse problem may not only be nonlinear but also ill-posed. For instance, a phase retrieval problem can have multiple solutions that satisfy the same intensity measurement constraint, making it difficult to determine the correct solution. Thus, it is more effective to reframe the problem as an optimization one and incorporate the prior knowledge into the optimization framework as follows,

$$\operatorname*{argmin}_{x} loss(A(x), y) + prior(x). \tag{1.3}$$

Here, the loss function used in model-based computational imaging methods depends on the specific algorithm design, while the prior knowledge is typically based on the characteristics of the sample being imaged. To solve this, various optimization algorithms can be employed, such as gradient descent. It iteratively updates the estimated image by descending along the negative gradient of the total cost function. Other optimization algorithms, such as conjugate gradient or stochastic gradient descent, can also be used depending on the specific problem and available resources. Overall, the choice of loss function, prior knowledge, and optimization algorithm plays a crucial role in the success of model-based computational imaging, and researchers continue to explore and develop new techniques to improve the accuracy and efficiency of this approach.

Model-based computational imaging methods offer several significant advantages. Firstly, they harness the underlying physical models of light propagation and image formation to reconstruct images with remarkable precision and fidelity. The entire reconstruction process is transparent and comprehensible. Secondly, these methods can employ prior knowledge of the scene or imaging system to enhance the reconstruction quality or reduce the required data

for accurate reconstruction. For instance, by integrating prior information about the scene's sparsity or structure, one can reconstruct high-quality images from a limited number of measurements, thereby lowering the acquisition time and data storage demands.

The clarity offered by model-based methods can be a double-edged sword. While it can be a boon, there are situations where constructing the physical model from $x$ to $y$ may be arduous or even impossible. Moreover, determining the nonlinearity characteristics of each imaging component or noise statistics during the imaging process may be unattainable. As a result, the reconstruction using model-based optimization method could fail or be impractical to implement [11]. To tackle this obstacle, data-driven computational imaging methods have been proposed.

Data-driven computational imaging methods leverage machine learning, computer vision, and other advanced algorithms to enhance image quality and extract valuable information from images [12]. By exploiting the statistical properties of image data, these techniques can enhance the entire imaging pipeline, from image acquisition and processing to analysis, reducing both hardware demands and the need for human labor. For example, data-driven computational imaging can enhance resolution, contrast, and SNR using algorithms that take advantage of the underlying image data's statistical patterns. Furthermore, deep learning models trained on labeled image datasets can be used for high-level image analysis tasks, such as object recognition and image segmentation.

Unlike model-based computational imaging techniques that rely on analytical forward models, data-driven computational imaging methods rely on large datasets of input-output data pairs. Based on deep neural network models, these techniques learn the direct mapping from the output $y$ to the input $x$ by leveraging the power of big data. The ability of deep neural networks to approximate any function ensures that, with sufficient data and careful training, the learned mapping from output to input data is both accurate and reliable.

Both model-based and data-driven computational imaging methods are highly versatile and can be adapted to a wide range of imaging scenarios. Furthermore, hybrid models combining

the benefits of both have been proposed recently [13]. These approaches have enormous potential to revolutionize various fields, including medical imaging, remote sensing, robotics, and autonomous vehicles.

The focus of this thesis is on computational imaging, and I will introduce my research in this area. Based on the dichotomy discussed earlier, my projects can be divided into two groups. Firstly, I will discuss two model-based computational imaging projects that focus on solving the phase retrieval problem in coherent imaging. These two projects are consecutive work and highly coherent to each other. Secondly, I will discuss two data-driven computational imaging projects that utilize deep learning techniques to solve biomedical image analysis problems. Despite both being based on deep learning, these two projects address distinct challenges in biomedical imaging. Before formally introducing my research, I would like to provide a brief overview of the basic background knowledge in each domain.

## 1.1 Coherent imaging theory

In this section, some important theoretical tools of coherent imaging are introduced for convenience of later discussion.

### Scalar diffraction theory

Scalar diffraction theory is the fundamental research tool in coherent imaging studies [7]. It is used when the size of obstacles and optical elements in light propagation and transformation is not close to the light wavelength scale, and the diffraction problem is not studied near the diffraction plane. In this scenario, scalar diffraction theory treats the diffracting wave as a scalar field, neglecting the vector nature of the electromagnetic wave. Therefore, it does not explicitly consider the coupling between the electric and magnetic fields described by Maxwell's equations. Only the scalar amplitude of one transverse component of the light vector is considered, assuming that any other relevant component can be treated independently. Under this theoretical framework, the physical process of wave propagation can be exactly described by the angular spectrum model.

When far from the radiation sources or in source-free media, Maxwell's equations in a three-dimensional infinite medium space can be formulated as

$$\nabla \cdot E = 0, \qquad \nabla \cdot B = 0,$$
$$\nabla \times E = -\frac{\partial B}{\partial t}, \qquad \nabla \times B = \varepsilon\mu\frac{\partial E}{\partial t}, \tag{1.4}$$

where $E$ is the electric field, $B$ is the magnetic field, $\varepsilon$ represents the permittivity of medium space and $\mu$ represents the permeability of medium space. If we take a curl of both sides of equations on the second line in Eq. (1.4), we will have

$$\nabla^2 E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = 0, \nabla^2 B - \frac{1}{c^2}\frac{\partial^2 B}{\partial t^2} = 0, \tag{1.5}$$

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the Laplace operator and $c = \frac{1}{\sqrt{\varepsilon\mu}}$. They are wave equations.

Without loss of generality, we can focus on the electric field $E$ only. Assuming the $E$ field is time-harmonic, it can be expressed as $E = U(x, y, z)\exp(-i2\pi vt)$, where $U(x, y, z)$ is the complex amplitude at the space point $(x, y, z)$. Substituting it to Eq. (1.5), we have

$$(\nabla^2 + k^2)U(x, y, z) = 0, \tag{1.6}$$

where $k = \frac{2\pi v}{c}$ is the wave vector. The wavelength in this medium space will be $\lambda = \frac{2\pi}{k}$.

Equation (1.6) is the Helmholtz equation.
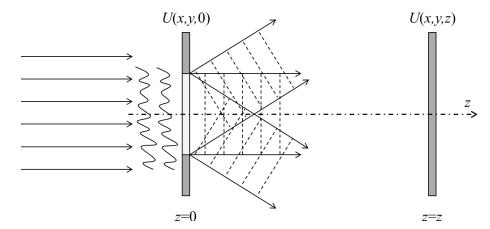


Figure 1.3: Wave propagation under scalar diffraction theory.

Now, we can model the wave propagation. Starting from the simplest case, we assume the wave propagates along $z$ axis and from the plane $z = 0$ to another plane at $z$, as shown in Fig. 1.3. If the wave field on any plane at $z$ is denoted as $U(x, y, z)$ and its Fourier spectrum is $G_z(f_x, f_y)$, we will have

$$U(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_z(f_x, f_y) \exp[i2\pi(f_x x + f_y y)] df_x df_y. \tag{1.7}$$

Substituting this to Eq. (1.6) and because any $(x, y)$ should make Eq. (1.6) hold, we conclude that

$$(\nabla^2 + k^2)\{G_z(f_x, f_y) \exp[i2\pi(f_x x + f_y y)]\} = 0. \tag{1.8}$$

Considering $G_z(f_x, f_y)$ is only dependent on $z$,

$$\frac{\partial}{\partial x} G_z(f_x, f_y) = \frac{\partial}{\partial y} G_z(f_x, f_y) = 0, \quad \frac{\partial}{\partial z} G_z(f_x, f_y) = \frac{d}{dz} G_z(f_x, f_y). \tag{1.9}$$

And we also have

$$\frac{\partial}{\partial x} \exp[i2\pi(f_x x + f_y y)] = (i2\pi f_x) \exp[i2\pi(f_x x + f_y y)],$$

$$\frac{\partial}{\partial y} \exp[i2\pi(f_x x + f_y y)] = (i2\pi f_y) \exp[i2\pi(f_x x + f_y y)], \tag{1.10}$$

$$\frac{\partial}{\partial z} \exp[i2\pi(f_x x + f_y y)] = 0.$$

Using Eq. (1.9) and Eq. (1.10), we can simplify Eq. (1.8) to

$$\frac{d^2}{dz^2} G_z(f_x, f_y) + \left(\frac{2\pi}{\lambda} \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right)^2 G_z(f_x, f_y) = 0. \tag{1.11}$$

It can be seen that this is still a Helmholtz equation, with respect to $G_z(f_x, f_y)$. If the wave field $G_0(f_x, f_y)$ at $z = 0$ is known, it is a particular solution to Eq. (1.11). Based on differential equation theory, the general solution of Eq. (1.11) can be expressed as

$$G_z(f_x, f_y) = G_0(f_x, f_y) \exp\left[i\frac{2\pi}{\lambda} z\sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right]. \tag{1.12}$$

This equation shows that the propagation of a wave along the $z$-direction is equivalent to multiplying the Fourier spectrum of wave field on the starting plane by a $z$-dependent phase

delay factor. In the linear system theory, it says that the transfer function of free space diffraction after the distance of $z$ is $H_z(f_x, f_y) = \exp\left[i\frac{2\pi}{\lambda}z\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right]$.

To further understand the physical picture, we can re-write Eq. (1.7) to

$$U(x,y,z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_z(f_x,f_y)\exp[i\frac{2\pi}{\lambda}(\lambda f_x x + \lambda f_y y)]df_x df_y. \tag{1.13}$$

It says that any wave field can be decomposed into a series of planar waves with amplitude of $\left|G_z(f_x,f_y)df_x df_y\right|$ and direction cosines of $(\lambda f_x, \lambda f_y, \sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2})$. The infinite integral limit indicates that the decomposed planar wave direction covers all possible directions in space, as shown in Fig. 1.3. The analysis of a complex wave field has now been replaced by the analysis of a series of plane waves with the same frequency but varying propagation angles, which is commonly referred to as the angular spectrum method. From Eq. (1.12), we can also see that the angular spectrum components when $1-(\lambda f_x)^2-(\lambda f_y)^2 < 0$ will decrease exponentially with the increase of $z$. It only exists in a very close proximity after the diffraction starting plane and is known as the evanescent wave. Only the angular spectral component that satisfies $1-(\lambda f_x)^2-(\lambda f_y)^2 \geq 0$ or equivalently $f_x^2 + f_y^2 \leq \frac{1}{\lambda^2}$ can reach the observation plane at $z$. Therefore, the diffraction of coherent wave field in free space can be viewed as passing through an ideal low-pass filter of radius $1/\lambda$.

**Fourier transform property of a lens**

Using Fourier transform, Equation (1.12) can be written as

$$U(x,y,z) = \Im^{-1}\left\{\Im\{U(x,y,0)\}\exp\left[ikz\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right]\right\}, \tag{1.14}$$

where $\Im$ is the Fourier transform operator. Expanding the root part of the phase factor of $\exp\left[ikz\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right]$, we obtain

$$\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2} = 1-\frac{1}{2}\lambda^2(f_x^2+f_y^2)+\frac{1}{8}\lambda^4(f_x^2+f_y^2)^2+\text{L}. \tag{1.15}$$

When $f_x$ and $f_y$ are small enough, we can keep only the first two terms. Equation (1.14) will be approximated as

$$U(x,y,z) = \mathfrak{I}^{-1}\left\{\mathfrak{I}\{U(x,y,0)\}\exp\left[ikz\left(1-\frac{1}{2}\lambda^2(f_x^2+f_y^2)\right)\right]\right\}. \tag{1.16}$$

Since the Fourier pair of $\exp\left[ikz\left(1-\frac{1}{2}\lambda^2(f_x^2+f_y^2)\right)\right]$ is $\dfrac{\exp(ikz)}{i\lambda z}\exp\left[i\dfrac{k}{2z}(x^2+y^2)\right]$, we

can rewrite Eq. (1.16) to the form of convolution in spatial domain as

$$U(x,y,z) = U(x,y,0)*\frac{\exp(ikz)}{i\lambda z}\exp\left[i\frac{k}{2z}(x^2+y^2)\right], \tag{1.17}$$

where * is the convolution operator. For conciseness, we denote $U(x,y,z)$ as $U_z(x,y)$ and the coordinates on the starting plane as $(x_0, y_0)$ to distinguish it from the ones on the receiving plane. Thus Eq. (1.17) can be written as

$$U_z(x,y) = \frac{\exp(ikz)}{i\lambda z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}U_0(x_0,y_0)\exp\left\{i\frac{k}{2z}\left[(x-x_0)^2+(y-y_0)^2\right]\right\}dx_0dy_0. \tag{1.18}$$

This is the well-known Fresnel diffraction integral formula, which was derived by the French scholar Fresnel prior to the establishment of Maxwell's equations. It can be interpreted as the Fresnel approximation of the angular spectrum diffraction formula. We can further expand Eq. (1.18) to

$$U_z(x,y) = \frac{\exp(ikz)}{i\lambda z}\exp\left[i\frac{k}{2z}(x^2+y^2)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}U_0(x_0,y_0)\exp\left[ik\frac{(x_0^2+y_0^2)}{2z}\right]$$
$$\times\exp\left[-ik\left(\frac{x}{z}x_0+\frac{y}{z}y_0\right)\right]dx_0dy_0. \tag{1.19}$$

If the propagation distance $z$ is large enough compared to the aperture size on the $(x_0, y_0)$ plane, the phase term $\exp\left[ik\dfrac{(x_0^2+y_0^2)}{2z}\right]$ will be approximately one. We will have

$$U_z(x,y) \propto \exp\left[i\frac{k}{2z}(x^2+y^2)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}U_0(x_0,y_0)\exp\left[-ik\left(\frac{x}{z}x_0+\frac{y}{z}y_0\right)\right]dx_0dy_0. \tag{1.20}$$

This is the well-known Fraunhofer diffraction equation. For a thin lens, its transmission function under the paraxial approximation is $t(x_0, y_0) = \exp\left[-ik\dfrac{\left(x_0^2 + y_0^2\right)}{2f}\right]$, where $f$ is its focal length. If we put it just behind the starting plane and the receiving plane is set as the lens focal plane, the phase term $\exp\left[ik\dfrac{\left(x_0^2 + y_0^2\right)}{2z}\right]$ in Eq. (1.19) can be canceled by phase modulation from lens. We will get

$$U_f(x, y) = \frac{\exp(ikf)}{i\lambda f}\exp\left[i\frac{k}{2f}(x^2 + y^2)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}U_0(x_0, y_0)$$
$$\times \exp\left[-i2\pi\left(\frac{x}{\lambda f}x_0 + \frac{y}{\lambda f}y_0\right)\right]dx_0 dy_0. \tag{1.21}$$

It tells that the focal plane complex field distribution of a lens is a Fourier transform of the input plane complex field distribution multiplied with a quadratic phase term and some constant.
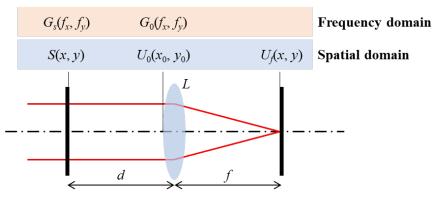


Figure 1.4: Fourier transform property of a lens. L, lens.

Now, we consider a more particular case, as shown in Fig. 1.4. The sample $S(x, y)$ is placed before the lens with the distance of $d$. We can depict the propagation from the exit plane after the sample to the front plane of lens with the Fresnel diffraction formula. According to Eq. (1.16), we can have the following expression in the spatial frequency domain,

$$G_0(f_x, f_y) = G_s(f_x, f_y) \exp\left[ikd\left(1 - \frac{1}{2}\lambda^2(f_x^2 + f_y^2)\right)\right]. \tag{1.22}$$

Then, from the front plane of lens to the receiving plane, we directly apply Eq. (1.21). Based on the definition of Fourier transform, it can also be written as

$$U_f(x, y) = \frac{\exp(ikf)}{i\lambda f} \exp\left[i\frac{k}{2f}(x^2 + y^2)\right] G_0(\frac{x}{\lambda f}, \frac{y}{\lambda f}). \tag{1.23}$$

Substituting Eq. (1.22) to Eq. (1.23), we have

$$
\begin{aligned}
U_f(x, y) &= \frac{\exp(ikf)}{i\lambda f} \exp\left[i\frac{k}{2f}(x^2 + y^2)\right] G_s(\frac{x}{\lambda f}, \frac{y}{\lambda f}) \\
&\times \exp\left[ikd\left(1 - \frac{1}{2}\lambda^2\left(\left(\frac{x}{\lambda f}\right)^2 + \left(\frac{y}{\lambda f}\right)^2\right)\right)\right].
\end{aligned}
\tag{1.24}
$$

Simplifying it, we can arrive at

$$U_f(x, y) = \frac{\exp\left[ik(f + d)\right]}{i\lambda f} \exp\left[i\frac{k}{2f}(x^2 + y^2)(1 - \frac{d}{f})\right] G_s(\frac{x}{\lambda f}, \frac{y}{\lambda f}). \tag{1.25}$$

If $d = f$, it becomes

$$U_f(x, y) = \frac{\exp(i2kf)}{i\lambda f} G_s(\frac{x}{\lambda f}, \frac{y}{\lambda f}). \tag{1.26}$$

If we ignore the constant factor, $U_f(x, y)$ is exactly the Fourier transform of $S(x, y)$. To conclude, a lens can connect its front and back focal plane with an optical Fourier transform.
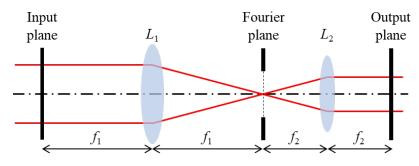


Figure 1.5: 4$f$ system. L, lens.

**4$f$ system and pupil function**

In this thesis, we focus more on the Fourier transform property of a lens rather than its imaging capability. A 4*f* optical system is an architecture based on Fourier optics, depicted in Fig. 1.5. It consists of two lenses. The input plane is positioned one focal length in front of Lens 1, while the output plane is placed one focal length after Lens 2. The distance between Lens 1 and Lens 2 is equal to the sum of their focal lengths. The Fourier plane, which corresponds to Fourier transform of the input plane wave front, is situated one focal length behind Lens 1 or equivalently one focal length in front of Lens 2. The magnification of the entire system is determined to be $-f_2/f_1$. If the two lenses have the same focal length, an identical relay with a magnification of -1 can be achieved. For this kind of system, it is easy to apply frequency domain modulation.

If we further put an aperture step on the Fourier plane as shown in Fig. 1.5, the 4*f* imaging system could become double-telecentric [14]. This type of system is highly accurate, as the magnification and FOV remains completely unchanged by any shifts in the object position or the sensor position. Double-telecentric lenses offer the advantages of both object and image space telecentricity, making them a popular choice in modern microscope design. In reality, the microscope objectives often comprise multiple lens elements. The 4*f* system is just a simplified physical model for it. The aperture stop on the Fourier plane, also called pupil, makes the whole system a low-pass filter. Its size is indicated by the nominal NA of objective. In the following discussion, the 4*f* system model will be intensively used.

In the coherent imaging framework, the optical imaging system can be viewed as a linear system. The coherent transfer function (CTF), also known as the pupil function, describes the impact of an optical imaging system on a light wave in the frequency domain. It is a complex function of the position within the pupil that indicates the relative changes in amplitude and phase of the light wave. Its phase captures all optical aberrations that occur between the sample plane and the image plane. Its amplitude could be different at different positions within the pupil, sometimes deliberately for the purpose of apodization. The support of the pupil function is typically circular with the radius determined by the objective

NA. Any imperfections in the optics can directly affect the pupil function, making it a critical tool for analyzing the performance of the optical imaging system.

If the input is a point source, the resulting wave front at the output plane is the Fourier transform of the pupil function. This wave front is referred to as the point spread function (PSF) of the optical imaging system. Therefore, any aberrations present in the system will be manifested in the PSF.

**Aberration and Zernike decomposition**

Aberration refers to the deviation of an optical system's performance from the idealized performance of its intended design. This deviation can cause the resulting image to be distorted or blurred. Aberration analysis can be conducted using either geometrical optics or wave optics techniques [15].

There are two types of aberrations: monochromatic and chromatic. Monochromatic aberrations arise due to the geometry of the lens or mirror and can occur when light is either refracted or reflected. They manifest even when monochromatic light is utilized, hence their name. On the other hand, chromatic aberrations arise due to dispersion, which is the variation of a lens's refractive index with wavelength. Because of this dispersion, light of varying wavelengths converges at different points, leading to chromatic aberration. This type of aberration is not observed when monochromatic light is used.

We will focus on monochromatic aberrations, as chromatic aberrations can be analyzed by wavelength using the same approach. The most common types of monochromatic aberrations include defocus, spherical aberration, coma, astigmatism, field curvature, and image distortion. Optical aberrations can be reflected in the phase of the pupil function, as introduced in the last section. Zernike polynomials can mathematically model circular wave front profiles of the pupil function associated with aberrations. These polynomials, developed by Frits Zernike in the 1930s, are orthogonal over a circle of unit radius. Using Zernike polynomials, a complex, aberrated wave front profile can be fitted to generate a set of fitting coefficients that represent different types of aberrations. These Zernike coefficients

are linearly independent, allowing for individual aberration contributions to be isolated and quantified separately.
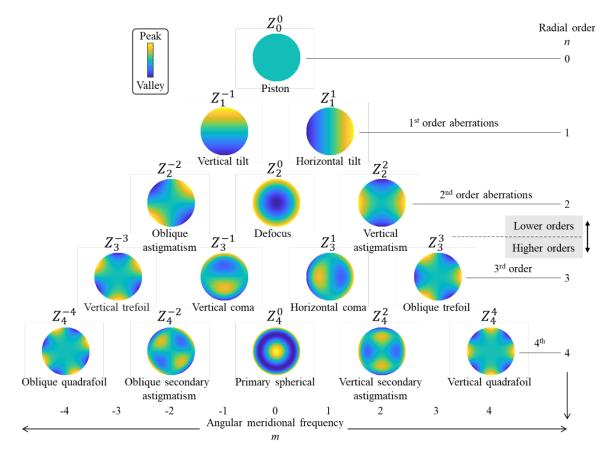


Figure 1.6: Zernike mode pyramid.

There are even and odd Zernike polynomials. The even Zernike polynomials are defined as

$$Z_n^m(\rho,\varphi) = R_n^m(\rho)\cos(m\varphi). \tag{1.27}$$

The odd Zernike polynomials are defined as

$$Z_n^{-m}(\rho,\varphi) = R_n^m(\rho)\sin(m\varphi), \tag{1.28}$$

where $m$ and $n$ are nonnegative integers with $n \geq m$, $\rho$ is the normalized radial distance and $\varphi$ is the azimuthal angle in radians. The radial polynomials $R_n^m$ have no azimuthal dependence and are defined as

$$R_n^m\left(\rho\right) = \sum_{k=0}^{\frac{n-m}{2}} \frac{\left(-1\right)^k\left(n-k\right)!}{k!\left(\frac{n+m}{2}-k\right)!\left(\frac{n-m}{2}-k\right)!}\rho^{n-2k}, \qquad (1.29)$$

for an even number of $n$ - $m$, while it is 0 for an odd number of $n$ - $m$. A special value is $R_n^m(1)=1$. We visualize the phase profile of each Zernike mode in Fig. 1.6. It is noteworthy that aberrations within the FOV of a microscope objective varies spatially.

## 1.2 Deep-learning-based computer vision tasks

Computer vision is a specialized field of computer science that aims to teach computers to identify and comprehend objects and people in images and videos. Similar to other forms of artificial intelligence (AI), computer vision seeks to automate tasks that replicate human capabilities.

Conventional computer vision techniques require hand-crafted features like edges or corners to detect objects or patterns in an image. This is followed by classification or recognition using traditional machine learning techniques such as decision tree [16] or support vector machine (SVM) [17]. Sometimes it also requires knowledge of imaging principles. In contrast, deep learning-based computer vision [18] relies on artificial neural networks to learn features directly from the data, without explicit feature extraction. This approach involves using deep neural networks, which are trained on large datasets to recognize complex patterns in images. The network architecture usually comprises multiple layers of interconnected neurons, with each layer learning increasingly complex features.

As large benchmark datasets with annotation become open-source and computer resources based on graphics processing unit (GPU) become cheaper, deep learning-based approaches have become the dominant method for achieving state-of-the-art performance on a wide range of image recognition tasks. Deep learning-based computer vision has been shown to outperform conventional computer vision in many applications, particularly in tasks that require a high degree of accuracy and robustness, such as image classification, object

detection, and segmentation. However, in cases with limited data or resources, conventional computer vision techniques still have their advantages.

**Different computer vision tasks**

Current mainstream computer vision tasks can be divided into two categories: discriminative and generative tasks. In a discriminative task, the goal is to predict a class label from input data by learning the decision boundary that separates different classes, which can be highly nonlinear and exist in high-dimensional space. On the other hand, a generative task aims to model the joint distribution of input data and labels and utilize it to generate new samples. Figure 1.7 provides an overview of the most common computer vision tasks where deep learning has achieved remarkable performance. However, in this thesis, we exclusively focus on discriminative tasks as, for biomedical applications, the possible hallucination in generative tasks is generally not acceptable.
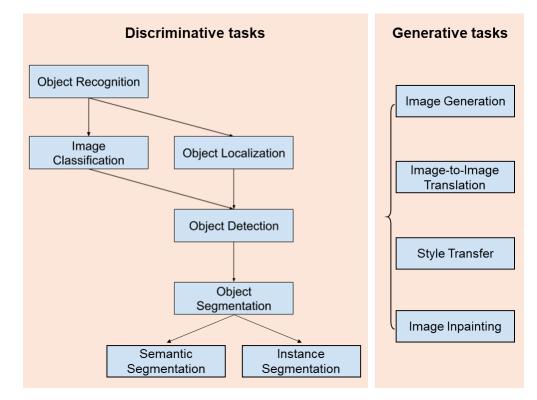


Figure 1.7: Different computer vision tasks.

**Image classification**

Image classification is a fundamental task in computer vision that involves assigning a label to an input image based on its visual content [19]. Common techniques involve recognizing patterns and features in images, such as shape, color, and texture, using feature extraction methods or handcrafting. Then we can train a machine learning model, such as boosted decision trees or SVM, to classify these features.

In the 2010s, the development of large visual databases like ImageNet [20] and the availability of GPU computational resources led to the emergence of groundbreaking models, starting with AlexNet [21]. This model revolutionized the image classification field and inspired subsequent successful models, including the widely used ResNet [22]. Convolutional neural networks (CNNs) have become a new paradigm in computer vision research, serving as the backbone not only for image classification but also for other tasks such as object detection and segmentation.

In recent years, researchers have continued to refine and optimize deep learning techniques for image classification, resulting in even higher accuracy rates and improved performance on a wide range of image recognition tasks [19].

**Object detection**

Object detection is a vital computer vision task that aims to identify and locate objects within images or videos and assign them to specific categories [23]. This task comprises two fundamental processes: object localization and object classification, which are illustrated in Fig. 1.7. Object localization entails identifying the object's position within the image, typically using bounding boxes annotations. Meanwhile, object classification involves assigning a label or category to the identified object. Deep learning models can perform both tasks in either a cascade or simultaneous manner. The former is referred to as a two-stage detector, such as the well-known Faster R-CNN [24], while the latter is known as a one-stage detector, such as SSD [25].

Object detection is a challenging problem due to the wide range of variations in object appearance, scale, pose, and occlusion, as well as the intricacies of real-world scenes. Despite

these challenges, object detection is an essential task with numerous applications, such as autonomous driving, robotics, and biomedical imaging [23]. Consequently, researchers have been continuously improving deep learning detection models to tackle these challenges and enhance the accuracy and efficiency of object detection.

## 1.3 Thesis outline

The remaining chapters of this thesis are structured as follows.

In Chapter 2, we introduce a computational imaging project based on Fourier ptychographic microscopy (FPM). FPM is a powerful computational imaging technique that was developed in our lab about a decade ago [26]. It can achieve high resolution, large FOV, phase retrieval, and aberration correction simultaneously. In this project, we use FPM's phase retrieval and pixel super-resolution abilities to address the issue of chromatic aberration in multispectral imaging.

Moving on to Chapter 3, we propose a new complex field imaging method based on Kramers-Kronig relations. By designing the pupil modulation method carefully, this method utilizes the analyticity of measurement signals to provide closed-form solutions. Its non-iterative reconstruction algorithm using as few as two measurements greatly reduces the measurement time and optimization uncertainty of FPM.

Building upon the developments discussed in Chapter 3, we further enhance the proposed method based on Kramers-Kronig relations by introducing two additional algorithms in Chapter 4. Leveraging the valuable insight gained from the relative locations between different spectrum components and the pupil function, we analytically correct aberrations and extend the spectrum to higher frequency bands utilizing dark field images. The resulting whole package scheme outperforms FPM with less raw data, higher reconstruction quality and robustness to severe aberrations. It provides an advantageous high-resolution, large field-of-view aberration-free coherent imaging modality.

Starting from Chapter 5, we shift our focus to data-driven computational imaging tasks. In this chapter, we showcase how co-designing a scanning scheme and an image processing pipeline can solve the defocus issue caused by uneven surfaces of microfilters. Additionally, we train a deep learning model on high-focus-quality images to automate the detection of circulating tumor cells and cancer-associated fibroblasts captured by microfilters. Our model demonstrates superior performance compared to the conventional computer vision method.

Finally, in Chapter 6, we tackle the challenging task of predicting the polarity phenomenon in embryo development from bright field images. Previously, this phenomenon could only be determined by fluorescence imaging techniques. We adapt deep learning models to address this problem and prove that they outperform human volunteers. Our findings demonstrate the potential of leveraging deep learning models to detect high-level patterns in image data, which are uninterpretable by human eyes, to solve seemingly impossible tasks.

*C h a p t e r   2*

# COMPUTATIONAL ABERRATION CORRECTION OF VIS-NIR MULTISPECTRAL MICROSCOPY BASED ON FOURIER PTYCHOGRAPHY

Due to the chromatic dispersion properties inherent in all optical materials, even the best-designed multispectral objective will exhibit residual chromatic aberration. In this chapter, we demonstrate a computational multispectral microscope based on the Fourier ptychographic microscopy (FPM) to correct these effects in order to render undistorted, in-focus images. The microscope consists of four spectral channels ranging from 405 nm to 1552 nm. After the computational aberration correction, it can achieve isotropic resolution enhancement as verified with the Siemens star sample. We imaged a flip-chip to show the promise of our system to conduct fault detection on silicon chips. This computational approach provides a cost-efficient strategy for high quality multispectral imaging over a broad spectral range.

## 2.1 Introduction

Multispectral microscopy, as a powerful method to obtain wide spectral images of a sample, has been applied to various research and industrial fields, from ore mineral inspection [27] and pharmaceutical composition determination [28] to biological tissue analysis [29]. Especially, with the wavelength extended to the near infrared (NIR) regime, the multispectral imager can have even more capabilities, including food quality inspection and wafer and solar cell production monitoring [30, 31]. Thus, multispectral microscope reaching to the NIR regime can be a promising instrument for these scenarios.

A multispectral microscope's performance highly depends on the optical system design, as aberrations are critical barriers for an optical system to achieve the ideal diffraction-limited resolution. Particularly, when the illumination wavelength ranges over a wide spectrum, such as from ultraviolet to NIR light, the chromatic aberration can become quite severe. Conventionally, to achieve an optical system whose aberrations are well-corrected for a wide spectrum, optical engineers need to put in much effort to combine multiple lenses of different surface profiles or materials to compensate for the aberrations with the aid of optical design software. Even if the final design meets the requirement, the manufacturing process can be difficult and costly. It often entails a bulky setup [32]. Furthermore, due to the finite number of optical surfaces, it will still suffer from residual aberrations to some extent [33].

As introduced in Chapter 1, the imperfect phase retardance of lenses that generates aberrations can be represented by the pupil function, a phase-only function. Instead of minimizing aberrations by the optical design, it is also viable to compensate for the phase error afterwards, either via additional hardware components, like a spatial light modulator, or by numerical post-processing. The key is to obtain the accurate wavefront. Its difficulty lies in the fact that the light frequency goes far beyond the response rate of the existing state-of-the-art detectors, unlike other waves such as acoustic waves whose wavefronts can be directly measured in time by sufficiently fast detectors. Therefore, the phase information of an optical wavefront can only be inferred from a detector's intensity measurements. There have been intensive research efforts around the optical wavefront sensing techniques. A simple non-interferometric method is to utilize a Shack-Hartmann wavefront sensor [34, 35], which comprises a set of micro-lenses placed in close proximity to an array detector. From the local focal spot shifts, the phase distortion can be computed. Despite its simple principle, a considerable setup modification is inevitable. Its spatial resolution is limited by the micro-lens density, and thus the phase approximation is relatively rough.

Alternatively, the wavefront can be reconstructed from direct intensity measurements by the image sensor. The reconstruction is performed by computational algorithms which can be categorized into two groups. Non-iterative methods are based on the transport-of-intensity

equation (TIE) [36–40]. It is a second-order elliptic partial differential equation that quantitatively relates the longitudinal intensity variation with the transverse phase distribution of the wavefront. However, this method cannot cope well with significant amplitude modulations as they can cause singularity problems in numerical calculation. The other group of computational wavefront measurement algorithms is categorized as the iterative method, which utilizes the degrees of freedom in the optical system to introduce phase diversity [41–43]. A set of seemingly distorted intensity patterns generated by phase diversity is first recorded by a detector and then combined with an iterative algorithm into the original wavefront's complex function. There are different ways to introduce the phase diversity. One simple way is by defocusing (i.e., axial scanning). Various algorithms for phase retrieval from defocus diversity have been reported [44–47]. It is suitable for lens-free on-chip setups to achieve high-throughput imaging [48, 49]. However, it is prone to pixilation artifacts, as the optical resolution can easily exceed the detector's pixel resolution. Another iterative wavefront reconstruction method operates by laterally scanning the wavefront over a thin sample. This method is termed ptychography [50–52], and it works by shifting an illumination probe over the sample with the coverage at each scanning position overlapping one another so that a highly redundant diffraction pattern dataset is captured. The overlapping region functions much like an interferogram during the iterative reconstruction procedure of the complex functions of both the sample and the probe, which guarantees the convergence of the non-convex phase retrieval computation. Afterwards, its Fourier-domain counterpart, Fourier ptychographic microscopy (FPM), was proposed [26].

FPM is based on the principle that the objective lens's pupil function can act as the probe that scans across the sample's Fourier spectrum under plane-wave illuminations of varied angles. Combined with a simple LED array to realize the spectrum scanning, it can access wider spatial-frequency band of the sample and thus enhance the imaging resolution by aperture synthesis with phase retrieval. Due to its simple setup and robust performance, it has attracted great interest from researchers, spurring theoretical and experimental achievements ranging from better optimization algorithms [53, 54], system error calibration [55], and noise suppression [56, 57], to three-dimensional imaging [58, 59]. Furthermore, a

variant of FPM, termed aperture-scanning FPM, was demonstrated to directly scan a physical aperture over the Fourier plane [60, 61], similar to the coded aperture method [62]. It can be easily realized for both transmissive and reflective imaging, and it is fit for applications where the sample-objective distance is not adjustable, such as wafer inspection [61] and retinal imaging [33]. In addition, without the varied-angle illumination requirement, the thin-sample assumption of the original FPM is circumvented [60].

Applying the aperture-scanning FPM to optical aberration removal has been previously explored in [63] which demonstrates simultaneous optimization of the pupil function estimate and the sample spectrum estimate via the simulated annealing algorithm. Unlike the simultaneous optimization algorithm proposed for angular-diversity FPM [53], this method was shown to be quite sensitive to input parameters and sample as it does not have the advantage of redundant pupil function information scanning across the sample spectrum. Moreover, its optimization computation workload is significantly great due to the high non-convex property of its formulation, which is unacceptable when calibrating the spatially varying aberration over a wide field-of-view (FOV), let alone for multiple wavelengths.

In this chapter, we report a multispectral microscope based on aperture-scanning FPM that is applicable to general imaging scenarios. With its ability to reconstruct complex wavefront functions, it removes aberrations present in the optical system in post-processing to recover sharp, in-focus images. To overcome the limited resolution of the array detectors for both visible and NIR wavelengths, we first use aperture-scanning FPM to synthesize pixel-super-resolution complex fields of a sample [61] over 4 illumination wavelengths (405 nm, 532 nm, 638 nm, and 1552 nm). Pixel super-resolution is especially critical for NIR wavelengths as the existing short-wave infrared (SWIR) cameras are all made of Indium Gallium Arsenide (InGaAs), whose noise control is much more difficult than the mainstream visible light (VIS) cameras based on silicon. Its pixel size is thus bigger (around 10-15 μm) to achieve sufficient signal-to-noise ratio (SNR). As such, it can easily violate the Nyquist sampling requirement of the objective lens in the microscope. Then with the pixel-super-resolved complex fields at hand, multiple defocused complex field slices can be generated and fed into the defocus-

diversity-based aberration reconstruction method proposed by [64] to account for the spatially varying aberrations.

## 2.2 Experimental setup

Our multispectral microscope system consists of the illumination path and the detection path, as shown in Fig. 2.1 (a). To obtain the multispectral information, 4 laser diodes with different wavelengths ranging from 405 nm to 1552 nm are coupled into a multi-mode fiber (FT200EMT-CUSTOM Thorlabs, 0.39 NA, Ø200 μm, FC/PC, 5m), which is jittered by a vibrating motor to wash out the speckle in the output. The diodes are sequentially powered by a programmable microcontroller (Arduino UNO) during the image capturing process. Depending on the sample to be imaged, transmission or reflection illumination mode is selected. In the detection part, the VIS and NIR imaging paths share the same objective (10X Mitutoyo Plan Apo NIR Infinity Corrected Objective, 0.26 NA), 4$f$ relay system and a mechanical scanning aperture. The paths are separated by a 50:50 beam splitter. Each path has a tube lens and a camera designed for the respective wavelengths. This dual-detection-path design is necessary because it is hard to find a camera with such a wide spectral response. The VIS camera (PROSILICA GX 6600) used for wavelengths ranging from 405 nm to 638 nm has 6576×4394 pixels with the pixel size of 5.5 μm. The magnification for VIS imaging path is 10 so that the effective FOV in the object space is around 3.62 mm×2.42 mm. The NIR camera (WIDY SWIR 640U-S) used for 1552 nm has 640×512 pixels with the pixel size of 15 μm. The magnification for NIR imaging path is intentionally set as 5× so that the effective FOV in the object space is approximately 1.92 mm×1.54 mm. These two cameras are first manually aligned to match their centers in the experimental setup, as shown in Fig. 2.1 (c), and then digitally registered in the signal processing step to rectify the slight tilting and rotation between two camera sensors. Here, the different magnification factor selected for NIR imaging path is the result of a tradeoff between FOV and the effective pixel sampling rate, which will be discussed together with the scanning aperture size in the following.
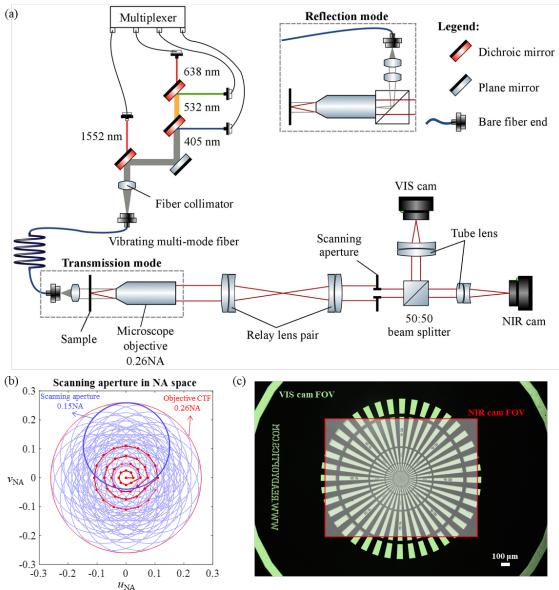
Figure 2.1: Multispectral, aperture-scanning Fourier ptychographic microscope system design. (a) The experimental setup schematic, switchable between transmission and reflection illumination mode. (b) FPM scanning strategy, where the red dots and thick red lines represent the scanning trajectory. The blue dotted circles represent the NA coverage of each scanning aperture. (c) Spatial position of VIS and NIR camera FOVs on the sample plane. For visualization, the VIS camera FOV is displayed in green and the NIR camera FOV in transparent gray.

The key to the aperture-scanning FPM in our system is a physical iris with adjustable aperture mounted on a 2D motorized translation stage. It is located at the relayed Fourier plane and scanned in a spiral pattern, as shown in Fig. 2.1 (b), for each laser diode. Then, a set of sub-

aperture intensity images are acquired for each wavelength. The overlapping rate between two adjacent aperture coverages is about 85%, satisfying the redundancy demand by the FPM algorithm. The total number of images required to fully exploit the objective NA is 47. The image sets will be fed into the reconstruction algorithm described in the following section. In our experiment, the aperture radius is chosen to be 3 mm, equal to 0.15 NA in the system. Thus, the effective pixel size in the object space should be less than 1.35 μm (calculated at 405 nm) for the VIS imaging path and less than 5.17 μm (calculated at 1552 nm) for the NIR imaging path in order to meet the Nyquist sampling requirement. According to the setup, the effective pixel size of the VIS camera is 0.55 μm under 10× magnification, meeting the sampling requirement. For the NIR camera, a large FOV is desired without going under the sampling limit. Thus 5× magnification is a balanced choice, where the effective pixel size is 3 μm.

## 2.3 Aberration calibration and correction scheme

Aberration correction for the multispectral optics is a difficult task. Here, we first utilize FPM to reconstruct the complex amplitude field and overcome the pixel sampling issue. Then, we assume the FOV's center region to be free from aberration and use it as the reference for determining the spatially variant pupil functions across the FOV, as has been done in [64]. Finally, the pupil functions are decomposed into the eight Zernike modes to be used for subsequent aberration-compensation procedures.

Dataflow pipeline of the aberration calibration and correction scheme is shown in Fig. 2.2. Here, polystyrene microspheres (Polybead, Polysciences) are used as the standard sample for calibration due to their following advantages. First, they have an isotropic shape and a consistent diameter with a variety of choices in size. We choose microbeads of different sizes for different wavelengths so that they can be treated as the quasi-point object for our optical system and used for the following calibration. Second, they can be homogeneously distributed across the full FOV and thereby avoid the mechanical movement of sample to calibrate the spatially varying aberration. As we can observe from Fig. 2.2 (a), the microsphere located at the peripheral region of FOV (denoted as off-axis bead) is distorted

into an elliptical shape by the residual aberration, while the one in the central FOV (denoted as on-axis bead) is circular.
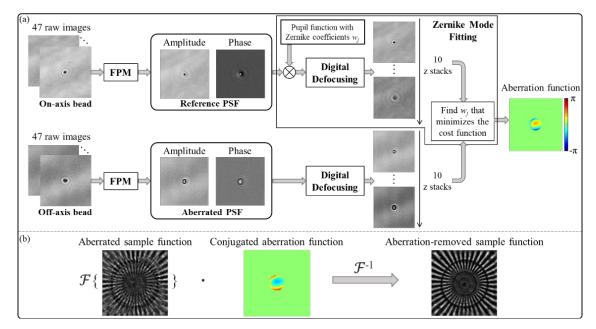


Figure 2.2: Dataflow pipeline of the aberration calibration and correction scheme. (a) Aberration calibration algorithm pipeline based on the microbead sample. It consists of three steps, FPM reconstruction, digital defocusing and Zernike mode fitting. (b) Aberration correction algorithm pipeline. The pupil function at some specific spatial location is compensated as its conjugate in the Fourier domain.

Shown in Fig. 2.2 (a), our aberration calibration algorithm pipeline can be divided into 3 steps. We assume that the following discussions are for a single wavelength.

**(I) FPM reconstruction**

The imaging process can be modeled as follows. The spatial coordinates on the sample plane, the Fourier plane and the image plane are respectively denoted as $(x, y)$, $(u, v)$ and $(x_1, y_1)$. The exit wave out of the sample plane is described by a complex-valued function $o(x, y)$. Then, for a single wavelength $\lambda_0$, the set of low-resolution raw images can be written as

$$I_k(x_1, y_1) = |F^{-1}\{F\{o(x,y)\} \cdot P_k(u,v)\}|^2, k = 1,2,\dots,47, \qquad (2.1)$$

where each $k$ corresponds to a scanning position and $P_k(u,v)$ is the shifted aperture function at that location. The goal of FPM algorithm is to reconstruct the phase from the

highly redundant dataset. First, a high-resolution spatial frequency guess $O'(u, v)$ is generated. Then the computational task is formulated as an optimization problem to minimize the following cost function:

$$\text{argmin} \sum_{k=1}^{47} ||F^{-1}\{O'(u, v) \cdot P_k(u, v)\}|^2 - I_k(x_1, y_1)|^2. \qquad (2.2)$$

Here, we adopt the sequential Gauss-Newton method to optimize it, taking the convergence speed, noise robustness, and computational cost into account [54]. Since the FPM algorithm has been discussed in detail in [54], it will not be elaborated on in this part.

**(II) Digital defocusing**

After the complex amplitude functions of on-axis and off-axis beads are reconstructed, they are digitally defocused to 10 different $z$-axis positions by the angular spectrum method [7]. If the reconstructed complex amplitude field is denoted as $B(x, y)$, the set of 10 $z$-stack images can be written as

$$I_t(x, y) = |F^{-1}\{F\{B(x, y)\} \cdot H_t(u, v)\}|^2, t = 1,2,\ldots,10. \qquad (2.3)$$

Here, the $H_t(u, v)$ is the free-space propagation kernel, which is

$$H_t(u, v) = exp\left[i\frac{2\pi}{\lambda_0}z_t\sqrt{1 - (\lambda_0 u)^2 - (\lambda_0 v)^2}\right], t = 1,2,\ldots,10. \qquad (2.4)$$

The defocusing distance interval $z_t - z_{t-1}$ is set as a constant linearly proportional to the wavelength in order to produce the enough spatial variations among the intensity images $I_t(x, y)$. The set of $z$-stack intensity images $I_{off,t}(x, y)$ from the off-axis bead function $B_{off}(x, y)$ will be used as the constraints in the following fitting step.

**(III) Zernike mode fitting**

Based on the assumption that the FOV's center is minimally distorted and well corrected by the objective lens's manufacturer, it can be taken as the aberration-free region with the diffraction-limited resolution performance. Therefore, the reconstructed on-axis bead $B_{on}(x, y)$ can be taken as the aberration-free image of the bead. The differences between this reference image and the aberrated bead image at other FOV locations in the Fourier domain are attributed to the residual aberrations. Considering that the minimization in the Fourier domain is sensitive to noise, we formulate the optimization problem in the spatial domain. This is why the step (2) is needed to obtain the $z$ stacks in the spatial domain.

The aberration function can be decomposed into orthogonal Zernike modes $Z_n^m(u,v)$, each of which corresponds to a principle aberration basis [15]. In this chapter, we focus on the top eight common and dominant Zernike modes, which are $Z_1^{-1}$ (tilt$_x$), $Z_1^1$ (tilt$_y$), $Z_2^{-2}$ (astigmatism$_x$), $Z_2^2$ (astigmatism$_y$), $Z_2^0$ (defocus), $Z_3^{-1}$ (coma$_x$), $Z_3^1$ (coma$_y$), and $Z_4^0$ (spherical aberration). For simplicity, we re-notate them as $Z_j(u,v), j = 1,2\ldots,8$. Consequently, a pupil function with a set of guessed Zernike coefficients is

$$P(u,v;w_1,\ldots,w_8) = exp\left[i2\pi \sum_{j=1}^8 w_j Z_j(u,v)\right], \tag{2.5}$$

where $w_j$ is the coefficient for each Zernike mode. Then we find the pupil function with the Zernike mode coefficients that can minimize the following optimization argument:

$$\text{argmin} \sum_{t=1}^{10} \left| |F^{-1}\{F\{B_{on}(x,y)\}P(u,v;w_1,\ldots,w_8) \cdot H_t(u,v)\}|^2 - I_{off,t}(x,y)\right|^2. \tag{2.6}$$

Once the optimal solution is found, the pupil function at the FOV location $(x,y)$ can be subsequently used to computationally compensate for the aberrations, as shown in Fig. 2.2 (b). Here, since the aberration is spatially varying and wavelength-dependent, the process above will be repeated for different locations in the whole FOV and for all four wavelengths.

## 2.4 Calibration of multispectral spatially varying aberration

One fact implied in the account of last section is that the aberration is spatially varying. This is more obvious for the low NA objective with a large FOV, such as the one used in our multispectral microscope system. Thus, the Zernike mode coefficient $w_j$ is a function of the spatial location $(x,y)$ on the sample plane and the wavelength $\lambda$, which can be written as $w_j(x,y;\lambda)$.

For a single fixed wavelength $\lambda_0$, a series of discrete data points can be acquired for $w_j(x,y;\lambda_0)$ by repeating the calibration procedures in last section, as displayed by the blue dots in Fig. 2.3 ($\lambda_0$ = 532 nm). It has been reported that the spatial dependency of different Zernike mode coefficients can be fitted with the polynomials of different orders [15]. Thus, these discrete data points are fitted to 2D surface in Fig. 2.3 so that the local Zernike coefficients for any spatial location in the FOV can be queried. From the magnitude of eight

Zernike coefficient distributions, it can be seen that astigmatism is the dominant aberration in the objective lens of our study.
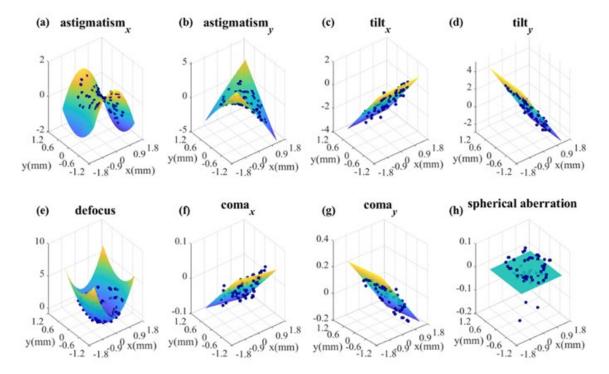


Figure 2.3: Spatial varying aberration of our multispectral microscope system when the wavelength is 532 nm. (a)-(h) Zernike mode coefficients as function of the spatial coordinate on the sample plane. Each blue-dot data point represents the calculated Zernike coefficient weight from one off-axis bead. ~70 beads are identified over the entire FOV. These data points are fitted to a 2D surface for each type of aberration.

The same aberration characterization procedures are repeated for each wavelength and the spatial variance of their Zernike coefficients is summarized in Fig. 2.4 and Table 2.1. Considering the tilt and defocus aberration modes are sample-dependent and can be compensated by the digital refocusing on local tiles, they are not shown here. In the remainder of this chapter, we only apply the reported three Seidel aberrations (astigmatism, coma, and spherical aberration) to correct residual aberrations and it can be seen in the next section that these Zernike terms are sufficient because of their dominance in the aberration function.

Table 2.1: Fitted spherical aberration.

| Wavelength (nm) | 405 | 532 | 632 | 1552 |
|---|---|---|---|---|
| Spherical aberration coefficient | 0.0089 | 0.0104 | 0.0116 | 0.1148 |
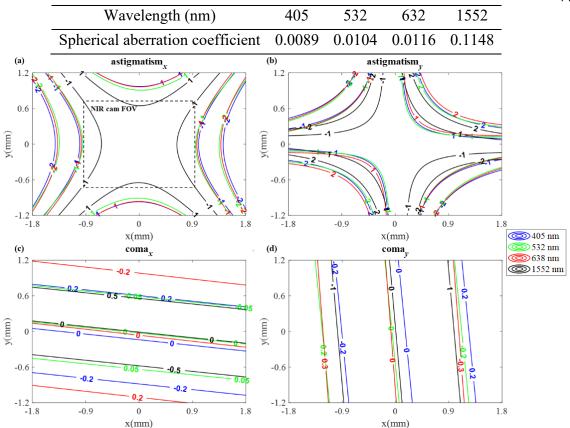


Figure 2.4: Spatial varying aberration of our multispectral microscope system for multiple wavelengths. The NIR fitted surface result is based on the beads located in the NIR camera's FOV, shown by the dash-line box in (a). For comparison, it is extrapolated to the same size as the VIS results. (a)-(d) Astigmatism and coma coefficients as function of the spatial coordinate on the sample plane. Each color represents a single wavelength.

For explicit visualization and comparison among four wavelengths, the aberration characterization results computed from the NIR camera's FOV and VIS camera's FOV are registered and plotted together. The result under the NIR wavelength is further extrapolated to the same size as the VIS ones to observe their varying trends. Since the spherical aberration coefficient is constant for individual wavelength, they are displayed in Table 2.1 instead of the contours in Fig. 2.4. From the contour plots, we can tell that the spatial dependence of different Zernike mode coefficients has similar distribution for each wavelength. However, the aberration is more severe for the NIR regime than for the VIS regime, as indicated by the absolute values of all aberration coefficients. They increase the fastest under the wavelength of 1552 nm when deviated from the FOV center.

| Wavelength | 405 nm | 532 nm | 638 nm | 1552 nm |
|---|---|---|---|---|
| Offset from optical center | 1567 μm | 1592 μm | 1910 μm | 792 μm |



(a)

**Transmission mode**

Before correction

After correction

(b)

**Reflection mode**

Before correction

After correction

20 μm

(c1) (c2) (c3) (c4)
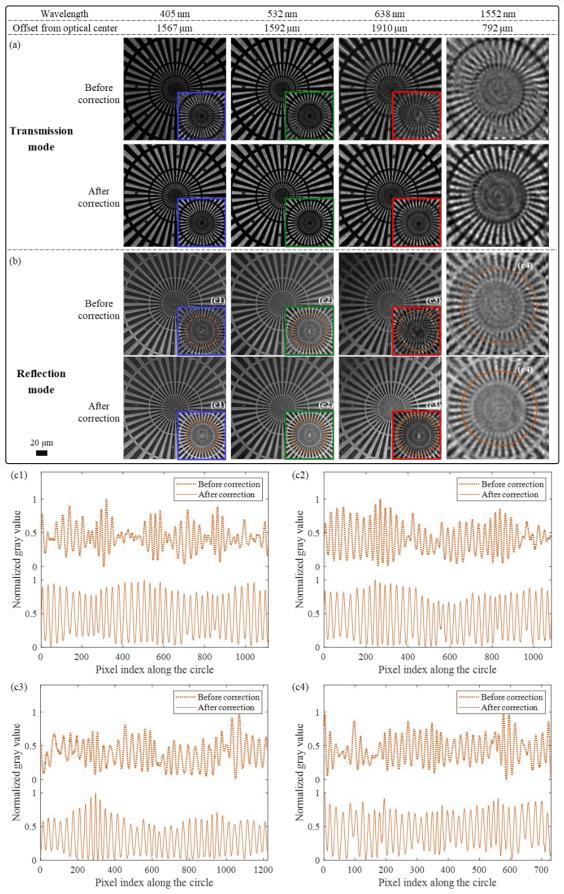
Figure 2.5: Demonstration of aberration calibration and correction. The Siemens star target was randomly offset from the optical center for different wavelengths. Transmissive (a) and reflective (b) reconstructions before and after correction are compared and the inset boxes zoom in on the sample center to show the resolution enhancement. (c1)-(c4) draws the line profile of a circle segment in the reflective reconstructions before and after correction.

## 2.5 Aberration correction for Siemens star imaging

After the spatially varying aberration for each wavelength is calibrated, a simple and direct way to verify its correctness is to apply it for imaging of a standard sample, such as the Siemens star suggested by [65]. In the experiment, the Siemens star target is randomly offset from the optical axis for each wavelength and the transmissive and reflective image datasets are acquired at the same location.

Following the correction scheme described in Section 2.3, reconstructions before and after correction are presented in Figs. 2.5 (a)-2.5 (b). To observe the resolution enhancement, the Siemens star's center is enlarged and shown in the inset boxes. It can be clearly seen that the aberrations at the arbitrary locations are well corrected under both transmission and reflection illumination modes across all wavelengths.

To further quantify the resolution enhancement, the line profile of a circle section from the reflective reconstructions under four wavelengths is drawn [66] and compared before and after aberration correction in Figs. 2.5 (c1)-2.5 (c4). The interval between peaks and troughs of Siemens star before correction is uneven. The intensity of some troughs is even higher than that of the peaks and several peaks and troughs are totally invisible. After compensation, the even space between peaks and troughs tells that the aberration-corrected contrast is uniform at all orientations, which indicates an isotropic resolution improvement. This verifies the efficacy of our aberration compensation method. Notably, the overall reconstruction quality under 1552 nm is not comparable to the VIS results, as the NIR camera has higher thermal noise.

## 2.6 Aberration correction for silicon chip imaging

A useful application of our multispectral microscope is for the silicon chip inspection. Here, a flip-chip is utilized as our imaging sample. Flip-chip is a prevalent method to interconnect

semiconductor devices to external circuitry with solder bumps that have been deposited onto the chip pads, as shown in the schematic of Fig. 2.6 (a). The flip-chip we used has nine metallic layers stacking above the silicon die with the thickness of around 305 μm. On top of the metallic layers bonded are eleven micro-ball bumps.
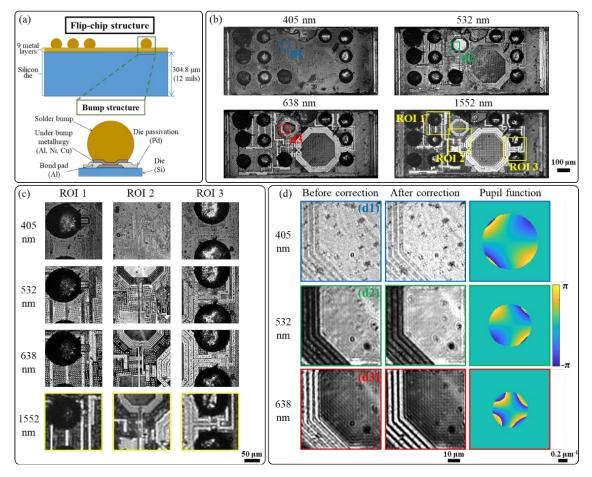


Figure 2.6: Schematic of the flip-chip and its frontside imaging. (a) Schematic of the flip-chip and its micro-ball bump. (b) Multispectral frontside images of the flip-chip. (c) Comparison between the visible and NIR images in three regions of interest (ROI) marked by the yellow box in (b). (d) Enlargements of the same area from three visible-light channel images, where the results before and after aberration correction are compared, together with the local pupil function for each wavelength.

In failure analysis of silicon chips, nondestructive imaging is important to avoid disturbing the functionality of integrated circuits (IC). High-resolution imaging techniques such as transmission electron microscope (TEM) or scanning electron microscope (SEM) require the transistors to be exposed destructively. Optical microscopy techniques may be used in the

frontside imaging on the top metallic layers. However, for the bottom layers or in situations where the flip-chip is bonded, optical imaging is not practical as silicon is opaque in the visible spectrum [32]. Considering silicon's bandgap at 300 K is around 1.12 eV, the radiation at wavelengths longer than 1.12 μm does not have enough energy to excite valence band electrons into the conduction band. Thus, silicon is nearly transparent in the NIR regime. The use of NIR backside imaging is common in the semiconductor failure analysis. Much effort has been put in the resolution improvement, such as using solid immersion lenses [67] and using a deformable mirror to compensate for aberrations [68]. We expect that the NIR imaging results on silicon chip can be further enhanced with our cost-efficient digital aberration correction method.

First, the frontside multispectral images are displayed in Fig. 2.6 (b). It can be seen that different information can be obtained under different wavelengths. Remaining adhesive and scratches are the clearest under 405 nm while the solder mask and the copper wires respectively provide contrast under 532 nm and 638 nm. The image under 1552 nm is similar to the one under 638 nm. The difference is more explicit in Fig. 2.6 (c), the zoom-in of three regions marked in yellow boxes in Fig. 2.6 (b). Considering the resolution limit imposed by the wavelength, it is preferable to image the frontside with the visible light. Furthermore, the aberration for three visible channels is corrected. As shown in Fig. 2.6 (d), the width of metal wires is not uniform due to aberrations before correction. After the aberration correction is applied with the calibrated local pupil function, the wire width becomes consistent, indicating the resolution is more isotropic. Another interesting observation from the local pupil functions is that their patterns are similar to each other except from being rescaled by the wavelength, which tells us that the spatial dependence of aberration has a similar pattern for each wavelength, just as shown in Fig. 2.4.

As shown in Fig. 2.7 (a), the 1552 nm wavelength can penetrate the silicon substrate of 305 μm thick while the 532 nm wavelength is totally blocked at the surface. Figure 2.7(c1) displays the bond pad array and (c2) is the backside image of the inductor corresponding to the ROI 2 in Fig. 2.6 (c).
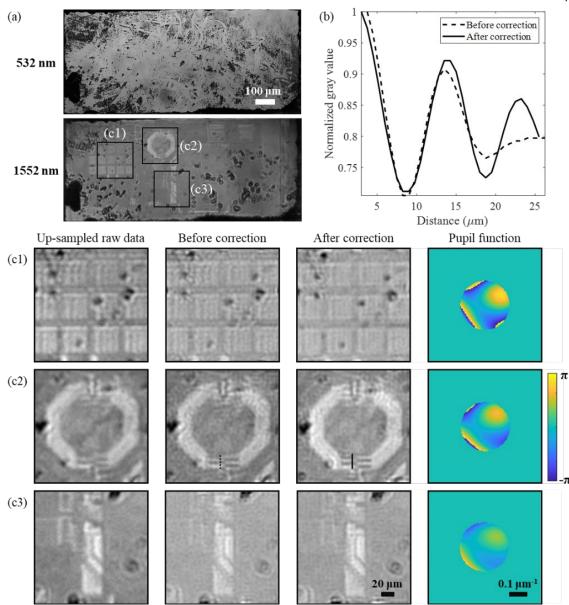
Figure 2.7: Backside imaging of the flip-chip. (a) Comparison between the backside raw image under the wavelength of 532nm and 1552nm. (b) Line profiles through the periodic structure in (c2) to highlight the aberration correction performance. (c1)-(c3) Comparison between up-sampled raw image, FPM-reconstructed image before aberration compensation and the image after compensation along with their local pupil function.

It is clear in Fig. 2.7 (c2) that the side-by-side copper wires of the inductor cannot be resolved by the direct up-sampled raw data due to the pixel aliasing. Although the FPM reconstruction can overcome the pixel sampling issue, the aberration still distorts the wire shape. Only after the aberration correction, the wires can be well resolved, as shown by the line profiles in Fig.

2.7 (b). This could be useful when diagnosing the faults, for example the short circuit, of the IC. Without the aberration correction, it is difficult to resolve the integrity of the wires.

To sum up, our multispectral microscope can provide an informative inspection to the flip chip. From the frontside imaging, the visible light can provide high-resolution multi-channel images for top layers. From the backside imaging, the NIR light can penetrate the optically opaque silicon wafer and see the structures beneath. For all the regimes, our aberration correction scheme can overcome the aberration barrier, which is important as the microscale details of chip can be easily distorted by the residual aberrations.

## 2.7 Summary

In this chapter, a scheme to calibrate and correct the aberrations of a multispectral microscope system is reported. It is based on the aperture-scanning FPM, which can be easily integrated into the traditional microscope system. With the pixel super-resolution ability of FPM [61], it relaxes the sampling requirement for the detector. Thus, it is especially suitable for the NIR imaging considering the current limitation in manufacturing detectors in the NIR regime. Moreover, our multispectral microscope is based on computational aberration correction which is much simpler in design and more cost-effective than other optical correction efforts. Once calibrated, it can be applied for subsequent imaging without further calibrations.

To demonstrate the effectiveness of our scheme, we have built and tested a multispectral microscope system with the operating wavelength ranging from 405 nm to 1552 nm. Due to the large FOV of the objective, the aberration in the periphery region is severe. We first calibrate the system with the standard polystyrene microspheres, and then the isotropic resolution enhancement is demonstrated with the Siemens star. Finally, we have shown a promising application to image the silicon-wafer flip-chip. Some IC details are only resolved after the aberration correction. Our scheme could find applications in nondestructive fault diagnosis of samples such as microelectromechanical systems (MEMS) devices, heavily doped silicon samples, wafer bonding, and 3D chip stacks.

*Chapter 3*

# NON-INTERFEROMETRIC AND NON-ITERATIVE COMPLEX OPTICAL WAVE-FIELD RECONSTRUCTION BASED ON KRAMERS-KRONIG RELATIONS

*This chapter is adapted from the manuscript, C. Shen, M. Liang, A. Pan, and C. Yang, "Non-iterative complex wave-field reconstruction based on Kramers–Kronig relations," Photonics Research 9(6): 1003-1012 (2021). DOI: 10.1364/PRJ.419886*

In the previous chapter, FPM was shown to be an effective approach for complex wave-field reconstruction. However, it does have some drawbacks. Firstly, it necessitates multiple measurements, resulting in lengthy data acquisition times and large raw data volumes. Secondly, its iterative reconstruction algorithm based on non-convex optimization can be time-consuming. In light of these issues, we present a new non-iterative and non-interferometric computational imaging technique in this chapter, called synthetic aperture imaging based on Kramers-Kronig relations (KKSAI). Our proposed method enables the reconstruction of complex wave-fields with greater efficiency.

By collecting images through a modified microscope system with pupil modulation capability, we show that the phase and amplitude profile of the sample at pupil limited resolution can be extracted from as few as two intensity images using Kramers-Kronig (KK) relations. It is established that as long as each sub-aperture's edge crosses the pupil center, the collected raw images are mathematically analogous to off-axis holograms. This in turn allows us to adapt a recently reported KK-relations-based phase recovery framework in off-axis holography for use in KKSAI. KKSAI is non-iterative, free of parameter tuning, and applicable to a wider range of samples. Simulation and experiment results have proved that it has much lower computational burden and achieves the best reconstruction quality when compared with two existing phase imaging methods.

**3.1 Introduction**

Coherent optical field can be described as a 2D complex function under scalar diffraction theory [7]. In most practical situations, we are not able to fully measure the entire complex function, because optical detectors are only suited for detecting intensity but not the phase of light. Yet phase image profiles are highly informative and useful in various applications, particularly in life science research, where phase images of living unstained cells can reveal cell structures that are otherwise invisible. To address this need, the optical imaging community has actively worked on and developed methods to infer phase information from purely intensity measurements over the past century or so.

To date, existing quantitative phase measurement methods can be generally categorized as interferometric and non-interferometric methods. Digital holography [69–73], phase shifting interferometry [74, 75], and optical coherence tomography [76, 77] belong in the former group. Non-interferometric methods include iterative phase (diversity) retrieval [42, 47, 78], (Fourier) ptychography [26, 50, 51, 79], transport of intensity equation [37, 39], and quantitative differential phase contrast [80–82]. Non-interferometric methods are inherently attractive, as they are generally simpler to implement and more robust to use. Quantitative phase imaging modalities have found numerous applications in a range of fields, including cellular mechanics, biophysics [83, 84], digital pathology [49, 85], X-ray crystallography [86], and optical metrology [87].

In this chapter, we report a novel computational imaging method that is able to compute the complex wave-field from purely intensity measurements. We name the method synthetic aperture imaging based on Kramers-Kronig relations (KKSAI). This method is closely related to pupil modulation Fourier ptychographic microscopy (FPM) [61], pupil modulation quantitative differential phase contrast (DPC) microscopy [82], and digital holographic microscopy [88–90]. However, it possesses certain advantages over all these existing methods. Compared to pupil modulation FPM [61], KKSAI is non-iterative and does not require data redundancy for operation. As such, KKSAI requires less raw data as well as processing time, and hence has a significantly reduced processing burden and an improved
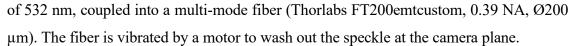
speed advantage. In comparison to pupil modulation DPC [82], KKSAI can work with a wider range of samples as it does not require samples to adhere to the weak sample assumption (both the absorption and phase variations of the sample must be sufficiently small) – a condition that pupil modulation DPC requires for operation. Compared to in-line holography [88], KKSAI does not have a twin-image issue, as guaranteed by the bandlimited signal analyticity. Compared to common-path off-axis holography [89], KKSAI's space-bandwidth product (SBP) can be notably higher by 3-4-fold.

This chapter is organized as follows: First, we present the experimental setup and data acquisition process. Next, we provide a detailed description of the KKSAI reconstruction algorithm and offer a mathematical proof of its validity, followed by two possible scanning schemes. Then, we report on our simulation and experiment results, demonstrating that KKSAI outperforms pupil modulation FPM and pupil modulation DPC methods when given the same measured data. We also clarify two differences between KKSAI and the original off-axis holography paper [90]. Finally, we conclude by addressing the overall efficiency of the KKSAI method compared to other phase imaging methods.

## 3.2 Experimental setup

The schematic of our KKSAI imaging system prototype is shown in Fig. 3.1 (a). It is simply a conventional wide-field microscope that has been modified to incorporate pupil modulation capability. As with other phase imaging methods, the goal here is to recover the phase profile of the sample from intensity image measurements.

In the experiment, the pupil plane is relayed outside the objective (10X Mitutoyo Plan Apo infinity corrected objective, 0.28 NA) onto the spatial light modulator (SLM) so that we can perform amplitude modulation of the pupil. The modulation module consists of a reflective mode liquid crystal on silicon (Holoeye LC-R 1080) SLM and a pair of linear polarizers (P1 and P2) with their polarization directions orthogonal to each other in order to maximize the amplitude modulation contrast. The light is reflected off the SLM and finally projected onto the camera. The pixel size of the camera (Allied Vision Prosilica GX 6600) is 5.5 μm. The illumination is provided by a laser diode (Thorlabs DJ532-40) with the central wavelength

of 532 nm, coupled into a multi-mode fiber (Thorlabs FT200emtcustom, 0.39 NA, Ø200 μm). The fiber is vibrated by a motor to wash out the speckle at the camera plane.
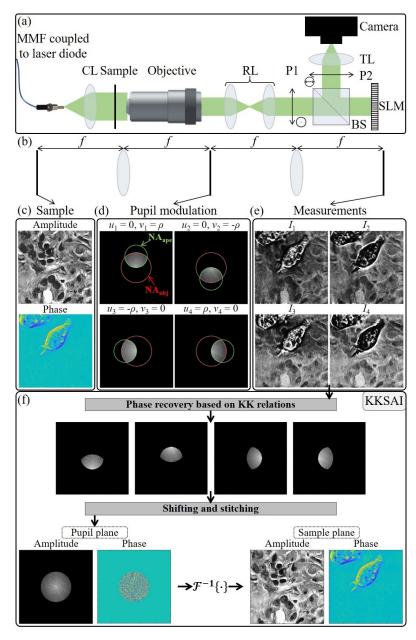


Figure 3.1: Principle of KKSAI. (a) Schematic of experimental setup, where pupil modulation is achieved by a SLM-based module. (b) Simplified $4f$ system corresponding to (a). (c) Simulated complex-valued sample. (d) Amplitude pupil modulation indicated by the green circle, whose center is $(u_i, v_i)$. (e) Measured images corresponding to (d). (f) KKSAI reconstruction algorithm flowchart. It finally recovers the pupil-limited sample spectrum. MMF, multi-mode fiber; CL, collimating lens; RL, relay lens; P, polarizer; BS, beam splitter; SLM, spatial light modulator; TL, tube lens.

The KKSAI system can be simplified and represented as a $4f$ imaging system, as shown in Fig. 3.1 (b). The coordinates at the sample plane, the pupil plane and the camera sensor plane are denoted as $(x', y')$, $(u, v)$ and $(x, y)$ respectively. If the complex-valued sample in Fig. 3.1 (c) is described by $s(x', y') = a(x', y')e^{j\varphi(x', y')}$, its maximum spectrum which can be detected by this system is

$$S(u, v) = F\{s\}(u, v) \cdot C(u, v), \tag{3.1}$$

where $F\{\cdot\}$ is the Fourier transform (FT) operator and $C(u, v)$ is the coherent transfer function (CTF) of imaging system as indicated by the red circle in Fig. 3.1 (d). Its radius is determined by the objective numerical aperture (NA). Our overarching aim is to recover $S(u, v)$, which corresponds to the complex wave field at the imaging system's pupil-limited resolution.

During KKSAI operation, we scan a binary circular aperture $D(u, v)$ at the pupil plane and its edge strictly crosses the pupil center, depicted by green circles in Fig. 3.1 (d). Technically, we display the appropriate aperture on the reflective SLM to implement this pupil modulation. We use four scanning steps to fully cover $S(u, v)$ and its offset distance from the pupil center is $\rho$. At each step, the scanning aperture can be denoted as $D(u - u_i, v - v_i)$, $i = 1,2,3,4$ and its cropped sub-region from sample spectrum is

$$S_i(u, v) = S(u, v) \cdot D(u - u_i, v - v_i). \tag{3.2}$$

Accordingly, the measured intensity image at the camera plane can be expressed as

$$I_i(x, y) = |F^{-1}\{S_i(u, v)\}|^2. \tag{3.3}$$

These measured intensity images are then fed into the KKSAI reconstruction algorithm (explained in the next section). The phase recovery process during reconstruction will be repeated for all 4 $I_i(x, y)$ to recover the complex field $S_i(u, v)$. They are subsequently stitched into a full pupil plane complex-valued spectrum $S(u, v)$. In turn, we can perform an inverse FT to constitute the final amplitude and phase image of sample (see Fig. 3.1 (f) for data flow chart).

**3.3 KKSAI reconstruction algorithm**

The primary aim of the KKSAI reconstruction algorithm is to recover the complex-valued $S_i(u,v)$ from $I_i(x,y)$.

To best explain the process, we can first express $I_i$ as an intensity result from superposition of an unscattered planar wave and the scattered field from sample:

$$I_i(x,y) = |F^{-1}\{S_i(u,v)\}|^2 = |F^{-1}\{S_i'(u,v) + \delta(u,v)\}|^2$$
$$= \left|s_i'(x,y) + e^{-j(0\cdot x + 0\cdot y)}\right|^2, \tag{3.4}$$

where $s_i'(x,y)$ is the scattered field exiting from the sample plane and $e^{-j(0\cdot x + 0\cdot y)}$ represents the ballistic light going through the sample. The ballistic light has a planar wavefront and propagates along the optical axis of our imaging system.

We can see that the Fourier spectrum of $I_i(x,y)$ is quite similar to the one of off-axis hologram, as the FT of $I_1$ in Fig. 3.2 (a) shows. This is because the scattered field $s_i'(x,y)$ is offset to the planar wave $e^{-j(0\cdot x + 0\cdot y)}$ in Fourier domain, and they would interfere with each other. (Here in Fig. 3.2, $\rho$ is decreased so that the scanning aperture will not exceed the support of CTF and $S_1(u,v)$ covers a circular region. This ensures that the cross terms in the spectrum correspond well to the cross terms in the traditional off-axis holography case. But all the following deduction holds for any $0 < \rho \leq \rho_{\mathrm{NA}_{ape}}$.)

To better explain this observation, we can employ the frequency shifting property of FT. If the covered spectrum sub-region $S_1(u,v)$ is shifted to the pupil plane center, only a phase term will be multiplied on the sensor plane. However, our intensity detector is not able to detect the phase. Therefore, the same intensity image would still be measured, as shown in Fig. 3.2 (b).

$$I_1(x,y) = |F^{-1}\{S_1(u,v)\}|^2 = |F^{-1}\{S_1(u+u_1, v+v_1)\}|^2. \tag{3.5}$$

Then, the shifted spectrum sub-region $S_1(u+u_1, v+v_1)$ can be hypothetically separated into a Dirac delta function $\delta(u+u_1, v+v_1)$ corresponding to the original planar wave and the shifted scattered field function

$$\tilde{S}_1(u,v) = S_1'(u+u_1, v+v_1), \tag{3.6}$$

as shown in Fig. 3.2 (c).
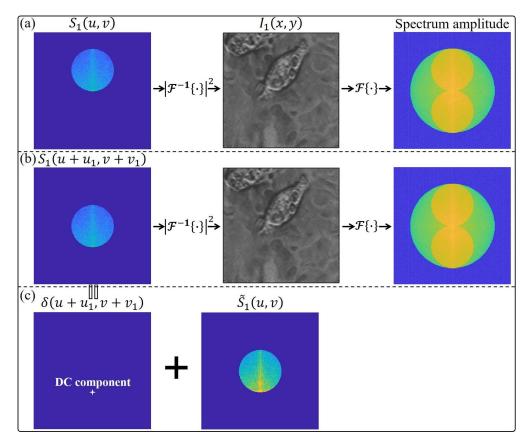


Figure 3.2: Analogy between KKSAI measurement and off-axis hologram. (a) Measurement $I_1$ between its corresponding complex-valued spectrum sub-region and the amplitude of its FT. (b) Shifted spectrum sub-region still brings in the same measurement due to the frequency shifting property of FT and phase loss of square-law detector. (c) Shifted spectrum sub-region can be hypothetically decomposed into a Dirac delta function and the shifted scattered complex-valued function.

Thus, the measurement $I_1$ can be expressed as

$$
\begin{aligned}
I_1(x,y) &= \left|\mathrm{F}^{-1}\{\tilde{S}_1(u,v) + \delta(u+u_1, v+v_1)\}\right|^2 = \left|\tilde{s}_1(x,y) + e^{-j(u_1 \cdot x + v_1 \cdot y)}\right|^2 \\
&= |\tilde{s}_1(x,y)|^2 + 1 + \tilde{s}_1^*(x,y) \cdot e^{-j(u_1 \cdot x + v_1 \cdot y)} + \tilde{s}_1(x,y) \cdot e^{j(u_1 \cdot x + v_1 \cdot y)},
\end{aligned} \tag{3.7}
$$

where $\tilde{s}_1(x,y)$ is the inverse FT of $\tilde{S}_1(u,v)$ and * denotes the complex conjugate operator.

Then its FT will be

$$
\begin{aligned}
\mathrm{F}^{-1}\{I_1\} = \tilde{S}_1(u,v) \star \tilde{S}_1(u,v) + \delta(u,v) + \\
\tilde{S}_1^*\big(-(u+u_1), -(v+v_1)\big) + \tilde{S}_1(u-u_1, v-v_1),
\end{aligned} \tag{3.8}
$$

where $\star$ represents cross-correlation operator. It can be clearly seen that the first two terms correspond to the self-interference terms in the off-axis hologram and the other two terms comes from cross interference. As such, we can see that our KKSAI measurement is analogous to an off-axis hologram.

Next, we can employ KK relations to perform phase recovery. The process is similar to the one reported recently for off-axis holography [90]. As it has been established that our KKSAI measurement shares pertinent similarities to off-axis hologram, we can adapt the mathematical formulation reported in Ref. [90] to recover phase from the KKSAI measurements.

To clearly explain the process, we use $I_1$ as the starting point. We first generate the hypothetical reference plane wave. It is determined by the offset of $S_1(u, v)$ and in turn by the scanning aperture position $(u_1, v_1)$. For $I_1$, this reference plane wave is expressed as

$$r_1(x, y) = \mathrm{F}^{-1}\{\delta(u + u_1, v + v_1)\} = e^{-j(u_1 \cdot x + v_1 \cdot y)}. \tag{3.9}$$

Next, we specify a directional Hilbert kernel. This kernel depends on the direction from the scanning aperture center to the DC term. For $I_1$, we can express it as:

$$H_1(u, v) = -j sgn(v_1) \cdot sgn(v), \tag{3.10}$$

where $sgn(v)$ is the sign function. The last step is based on directional Hilbert transform to retrieve the complex field corresponding to the spectrum sub-region $S_1(u, v)$. We first define an intermediate variable,

$$X = ln(\mathrm{F}^{-1}\{S_1(u + u_1, v + v_1)\}/r_1(x, y)) = ln\big((\tilde{s}_1(x, y) + r_1(x, y))/r_1(x, y)\big). \tag{3.11}$$

Then,

$$\begin{aligned} Re\{X\} &= 0.5 \cdot ln(I_1(x, y)/|r_1(x, y)|^2), \\ Im\{X\} &= \mathrm{F}^{-1}\{\mathrm{F}\{Re\{X\}\} \cdot H_1(u, v)\}, \end{aligned} \tag{3.12}$$

which will be expanded on in the next section. Thus,

$$S_1(u + u_1, v + v_1) = \mathrm{F}\{e^{Re\{X\}+jIm\{X\}} \cdot r_1(x, y)\}. \tag{3.13}$$

After obtaining the shifted version of four sub-regions, we can move them back to the correct position and get

$$S_i(u, v), i = 1,2,3,4. \tag{3.14}$$

Eventually,

$$S(u,v) = \sum_{i=1}^{4} S_i(u,v)/(\sum_{i=1}^{4} D(u-u_i, v-v_i) + \varepsilon), \qquad (3.15)$$

where $\varepsilon = 10^{-5}$ is a small constant for numerical stability in the zero-valued region. As seen in Fig. 3.1 (f), the reconstruction results of KKSAI are in good agreement with the original sample function.

The validity of this KK-relations-based process has been demonstrated in the original paper [90] for off-axis holography. It leverages the analyticity of band-limited signals. To ensure that it also applies for KKSAI, we need to prove our measured image signal is indeed analytical as well.

## 3.4 Signal analyticity proof and scanning scheme

For off-axis holography, the self and cross interference terms are required to be separable in Fourier domain such that one cross term can be cropped out and applied inverse FT to recover the whole complex field. Such separation is usually achieved by adjusting the reference wave incidence angle. However, the recent study [90] relaxed this restriction and proved that as long as the two cross terms do not overlap, the complex field can be fully determined. We can arrive at this conclusion by using Titchmarsh theorem [91].

For conciseness, we use $\vec{l}$ as the vector representative of $(x, y)$ coordinate. Also, without loss of generality, the following variable replacements are used.

$$\begin{aligned} I_1(x,y) &\rightarrow I(\vec{l}) \\ \tilde{s}_1(x,y) &\rightarrow \tilde{s}(\vec{l}) \\ r_1(x,y) &= e^{-j(u_1 \cdot x + v_1 \cdot y)} \rightarrow r(\vec{l}) = e^{-j\vec{\rho}_r \cdot \vec{l}} \end{aligned} \qquad , \qquad (3.16)$$

where $\vec{\rho}_r = (u_1, v_1)$.

Then Eq. (3.7) can be expressed as

$$I(\vec{l}) = \left| \tilde{s}(\vec{l}) + r(\vec{l}) \right|^2. \qquad (3.17)$$

Then

$$I/|r|^2 = |\tilde{s}/r + 1|^2, \qquad (3.18)$$

$$ln|\tilde{s}/r + 1| = \frac{1}{2}\big(ln(I) - ln(|r|^2)\big). \qquad (3.19)$$

Let $X = ln(\tilde{s} + r/r) = ln(\tilde{s}/r + 1) = Re\{X\} + jIm\{X\}$. We will have

$$e^X = \tilde{s}/r + 1 = e^{Re\{X\}} \cdot e^{jIm\{X\}}, \qquad (3.20)$$

$$|\tilde{s}/r + 1| = e^{Re\{X\}}, \qquad (3.21)$$

$$Re\{X\} = ln|\tilde{s}/r + 1| = \frac{1}{2}\left(ln(I) - ln(|r|^2)\right). \qquad (3.22)$$

Equation (3.22) shows that $Re\{X\}$ is a quantity determined by the measurement $I(\vec{l})$ and the reference wave amplitude $|r|$. Since $r(\vec{l}) = e^{-j\vec{\rho}_r \cdot \vec{l}}$ is a planar wave, its amplitude $|r| = 1$ is a constant. Thus, $Re\{X\}$ only depends on $I(\vec{l})$ and is known. Now, if $X$ was analytical, KK relations can retrieve $Im\{X\}$ directly from $Re\{X\}$ and the complex field $\tilde{s}(\vec{l})$ can be reconstructed.

To find the analyticity condition for $X$, we define

$$\alpha(\vec{l}) = \frac{\tilde{s}}{r} = \tilde{s}(\vec{l}) \cdot e^{j\vec{\rho}_r \cdot \vec{l}}. \qquad (3.23)$$

Then

$$X = ln(\alpha + 1) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} \alpha^{n+1}, \qquad (3.24)$$

where the condition $\alpha(\vec{l}) = \tilde{s}/r < 1$ is satisfied as long as the energy of the DC component is several order of magnitude higher than the other frequency components in practice. Equation (3.24) tells that the analyticity of $X$ depends on the analyticity of $\alpha$.

Next, we introduce Titchmarsh theorem to verify the analyticity of $\alpha$. The theorem states that the following conditions are equivalent for a complex-valued function $f(l)$ that is square integrable over the real $l$-axis:

- The real and imaginary parts of $f(l)$ are Hilbert transforms of each other.
- Its Fourier transform $F\{f\}(\rho)$ is 0 or vanishes rapidly for $\rho < 0$.

An important observation is that $\tilde{s}(\vec{l})$ is always a band-limited signal under the pupil modulation, as shown in Fig. 3.3 (a-c). In our experiments, it is characterized by the scanning aperture radius. For the convenience of discussion, $\rho_{\parallel} = \vec{\rho}_r/|\vec{\rho}_r|$ and its perpendicular direction $\rho_{\perp}$ are defined on the $(u, v)$ plane as a new set of coordinates. Their corresponding

axis pair on the $(x, y)$ plane is $l_{\parallel}$ and $l_{\perp}$. If applying 1D FT to both sides of Eq. (3.23) along $l_{\parallel}$, we have

$$A\big(\rho_{\parallel}, l_{\perp}\big) = \tilde{S}\big(\rho_{\parallel} - |\vec{\rho}_r|, l_{\perp}\big), \tag{3.25}$$

where $A\big(\rho_{\parallel}, l_{\perp}\big)$ and $\tilde{S}\big(\rho_{\parallel}, l_{\perp}\big)$ is 1D FT of $\alpha(\vec{l})$ and $\tilde{s}(\vec{l})$ along $l_{\parallel}$ respectively as shown in Fig. 3.3 (d).
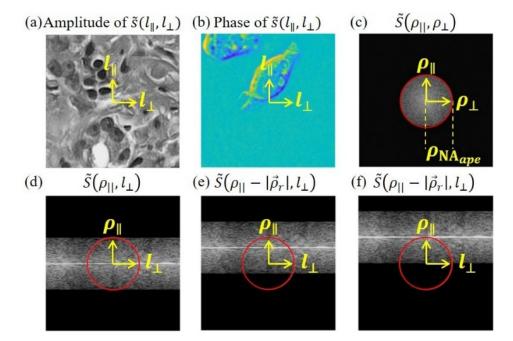


Figure 3.3: Titchmarsh theorem applied to a band-limited signal. (a) Amplitude and (b) Phase of $\tilde{s}(\vec{l})$ with bandwidth of $\rho_{\mathrm{NA}_{ape}}$. (c) Logarithm of its 2D Fourier amplitude spectrum. (d) Logarithm of its 1D Fourier amplitude spectrum along $l_{\parallel}$ axis and its shifted copies by (e) $|\vec{\rho}_r| < \rho_{\mathrm{NA}_{ape}}$ and (f) $|\vec{\rho}_r| = \rho_{\mathrm{NA}_{ape}}$.

Since $\tilde{s}(\vec{l})$ is band-limited, $\alpha(\vec{l})$ must be a complex-valued square integrable function along $l_{\parallel}$. Hence Titchmarsh theorem is valid for $\alpha(\vec{l})$. This implies that as long as $A\big(\rho_{\parallel}, l_{\perp}\big) = 0$ for $\rho_{\parallel} < 0$, its real and imaginary part are Hilbert transforms of each other. Equation (3.25) shows that $A\big(\rho_{\parallel}, l_{\perp}\big)$ is a shifted copy of $\tilde{S}\big(\rho_{\parallel}, l_{\perp}\big)$ along $\rho_{\parallel}$. As such, when the shifting distance $|\vec{\rho}_r|$ is at least equal to $\rho_{\mathrm{NA}_{ape}}$ as illustrated in Fig. 3.3 (f), the condition is met and $\alpha(\vec{l})$ is analytical.

We further note that, although the analyticity condition $|\vec{\rho}_r| \geq \rho_{\mathrm{NA}_{ape}}$ seems 1D along $l_\parallel$, it is actually a sufficient condition of 2D analyticity on the plane $\vec{l}$ in the limit sense, which has been proven in Ref. [92] and discussed in detail in the supplementary of Ref. [90].

To sum up, when $|\vec{\rho}_r| \geq \rho_{\mathrm{NA}_{ape}}$, $X$ satisfies the analyticity condition and as such the KK relations can be applied to recover the complex field $\tilde{s}(\vec{l})$. However, when $|\vec{\rho}_r| > \rho_{\mathrm{NA}_{ape}}$, the DC component cannot enter the measurement and KKSAI's analogy to off-axis hologram fails. Therefore, for KKSAI to function properly, we need to operate under the condition where $|\vec{\rho}_r| = \rho_{\mathrm{NA}_{ape}}$. In other words, the scanning aperture edge must cross the pupil origin exactly.
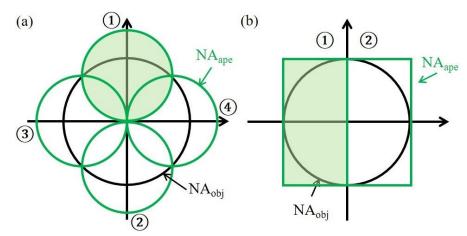


Figure 3.4: Scanning scheme examples to cover the entire pupil. (a) Four circular apertures; (b) Two rectangular apertures. ⓘ is used to label the measurement sequence.

Figure 3.4 (a) shows the scanning scheme discussed above. It is designed to cover the entire pupil and contains some overlaps. However, since our KKSAI method does not require the data redundancy, this scanning is slightly inefficient because of the overlaps. If we stick with scanning a circular aperture to cover the entire pupil, this overlapping issue is inevitable. If we are free to choose any aperture shape, then the scanning scheme with 2 non-overlapping measurements shown in Fig. 3.4 (b) would be the most efficient. The two scans have aperture edges that exactly cross the pupil origin. In principle, as long as the scanning of any convex aperture covers the whole pupil and its edge crosses the origin, the scanning scheme should work for KKSAI.

If not stated, Figure 3.4 (a) is our default scanning scheme used in the rest of this paper. We chose this scheme because the methods that we compare KKSAI to require overlaps in order to accomplish phase recovery. Finally, we note that the camera sampling rate of our method still needs to satisfy the Nyquist sampling limit $2\rho_{\mathrm{NA}_{ape}}$ to avoid sub-sampling aliasing.

## 3.5 Simulation results and discussion

In this section, we report on a series of simulations that were performed to verify the performance of our proposed method. The results are also compared with two existing imaging modalities, pupil modulation DPC (PM-DPC) [82] and pupil modulation FPM (PM-FPM) [60, 61].

To be more precise, we briefly describe these two existing phase imaging methods here. PM-DPC [82] is a counterpart of original quantitative DPC [80] and is different in that it replaces the asymmetric illumination with asymmetric pupil modulation. Specifically, it captured two or four phase gradient images from complementary half-open pupils. Then, when combined with an additional sample's intensity image, the method can compute the complete field of the sample. PM-FPM [60, 61] is the counterpart of the original FPM [26] and is different in that it replaces sample spectrum translation by oblique illumination with aperture translation on the pupil plane. Its reconstruction usually requires at least 20 raw images to converge. Although we note that there are other variants to the original DPC and FPM [93, 94], we chose these two to compare with KKSAI because they all use similar experimental setup.

All the existing complex field imaging methods have tradeoffs between system complexity, sample restriction, measurement data volume, reconstruction time complexity, and reconstruction accuracy. To make our comparison meaningful, we fix the system and measurement data volume so as to examine other aspects. Thus, all three methods in the following discussions utilize the same dataset $I_i$ for fairness.

First, we simulated pure phase samples with small and large phase variations, and the results of three methods are shown in Fig. 3.5. Looking at the reconstructed phases, we can discern

that KKSAI appears to render phase images that are closest to the ground truth regardless of the phase magnitude. To quantitatively compare their performance, we additionally calculated the mean square error (MSE) and feature similarity (FSIM) index [95] between reconstruction and ground truth and report the result in Table 3.1.
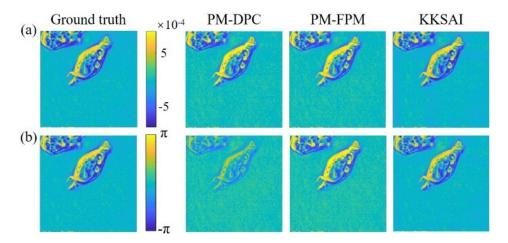


Figure 3.5: Reconstruction of phase-only samples by two existing methods and our proposed method. (a) Weak phase sample; (b) Strong phase sample.

Table 3.1: Quantitative evaluation of reconstructions in Fig. 3.5.

| | Metric | PM-DPC | PM-FPM | KKSAI |
|---|---|---|---|---|
| (a) | MSE | $4.80 \times 10^{-9}$ | $4.53 \times 10^{-9}$ | $7.84 \times 10^{-10}$ |
| | FSIM | 0.9999 | 1.0000 | 0.9998 |
| | Time(s) | 2.01 | 28.43 | 2.12 |
| (b) | MSE | 0.1711 | 0.0640 | 0.0136 |
| | FSIM | 0.9894 | 1.0000 | 0.9973 |
| | Time(s) | 1.67 | 77.33 | 2.03 |

Quantitative evaluations show that PM-FPM phase reconstruction is excellent in terms of similarity, but its MSE is higher than that of KKSAI. Also, the PM-FPM iterative computation is quite time consuming. On a computer (i7-7700k) with 64GB RAM, its run time is one order of magnitude higher than the one for the other two non-iterative methods in MATLAB R2018b.

The PM-DPC method provides reasonable phase rendering for the sample with small phase variations, but it fails when the phase variations are large. This is because its reconstruction

algorithm demands that the sample adheres to the weak sample assumption, which is necessary to first-order Taylor expansion approximation.
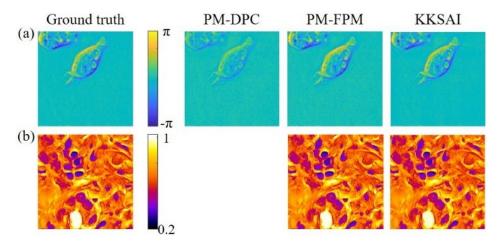


Figure 3.6: Reconstruction of complex-valued sample by two existing methods and our proposed method. (a) Phase; (b) Amplitude.

Table 3.2: Quantitative evaluation of reconstructions in Fig. 3.6.

| Metric | | PM-DPC | PM-FPM | KKSAI |
|---|---|---|---|---|
| Phase | MSE | 0.0531 | 0.0120 | 0.0037 |
| | FSIM | 0.9934 | 0.9997 | 0.9976 |
| Amplitude | MSE | / | $1.55 \times 10^{-8}$ | $3.78 \times 10^{-4}$ |
| | FSIM | / | 1.0000 | 0.9965 |
| Time(s) | | 3.83 | 109.37 | 2.82 |

Table 3.3: Similarity evaluation of overlapping spectrum regions in Fig. 3.6.



| Overlapping region | | | | |
|---|---|---|---|---|
| Phase | 0.9940 | 0.9967 | 0.9990 | 0.9926 |
| Amplitude | 0.9997 | 0.9998 | 0.9999 | 0.9998 |

Next, we simulated a complex-valued sample. Its amplitude and phase are displayed in Fig. 3.6 (a) and (b). The quantitative comparison is summarized in Table 3.2. For this sample, PM-DPC fails badly as it is incapable of recovering amplitude. Moreover, the large sample phase variations violate the weak sample assumption required for PM-DPC reconstruction.

Not surprisingly, it provides a poor phase rendering in this case. PM-FPM and KKSAI obtained phase and amplitude reconstruction of similar quality. However, PM-FPM requires good initialization and careful parameter tuning to arrive at the correct result. Furthermore, it takes 40 times longer to converge to the current result when compared to KKSAI.

In this simulation study, we conducted an additional evaluation for KKSAI. As the scanning scheme used here collects redundant information in the overlaps of the spectrum (see Fig. 3.4 (a)), we can also compare the reconstruction within the overlap regions using the FSIM metric. The results are summarized in Table 3.3. We can clearly see that the phase and amplitude recovery from KKSAI is highly consistent, thereby further reinforcing our confidence that KKSAI is providing us with the correct phase estimate.

To conclude, we can see that KKSAI outperforms PM-DPC and PM-FPM when we take both reconstruction accuracy and computational load into account.
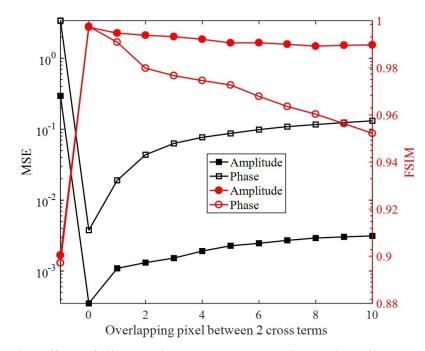


Figure 3.7: The effect of distance between aperture edge and pupil center on the final reconstruction accuracy.

Our simulation study allows us to investigate the potential impact of experimental imperfection when KKSAI is implemented practically. Specifically, we are concerned at the

way the aperture edge crosses the pupil center. Ideally, we want the aperture edge to cross the pupil center exactly ('zero' overlap). If the aperture includes the pupil center (positive overlap), the analyticity condition would be broken. On the other hand, if the aperture excludes the pupil center (negative overlap), $I_i(x, y)$ would not contain the ballistic planar wavefront and no cross-interference would occur, resulting in complete breakdown of the KKSAI processing. We simulated the impact of this overlap issue and showed our results in Fig. 3.7. We can clearly see that negative overlap results in a catastrophic failure of our KKSAI method, but positive overlap is somewhat tolerable as long as the overlap is small.
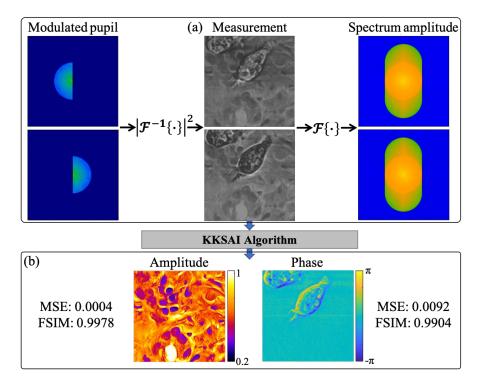


Figure 3.8: KKSAI based on the scanning scheme with only two measurements. (a) Measurements; (b) Reconstruction.

Our final simulation is focused on the feasibility of the scanning scheme in Fig. 3.4 (b). Here we use an aperture that alternates between two halves of the pupil as shown in Fig. 3.8 (a). We can see from the results reported in Fig. 3.8 (b) that the phase and amplitude recovery result is highly consistent with the ground truth in Fig. 3.6, as indicated by the quantitative metrics.

### 3.6 Experiment results and discussion

With the encouraging simulation results, we next conducted a series of experiments with the setup shown in Fig. 3.1 (a).
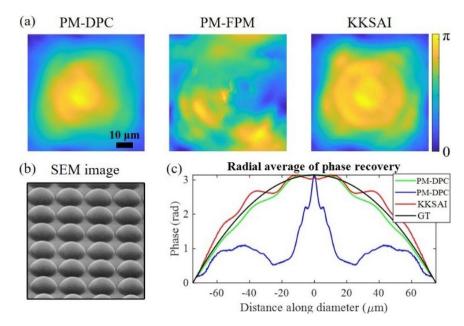


Figure 3.9: Experimental results for a microlens array. (a) Reconstructions of a single microlens by PM-DPC, PM-FPM and KKSAI using four measurements; (b) A close-up view of SEM image of microlens array (adapted from Thorlabs website); (c) Radial average profile of three phase recoveries compared with the ground truth (GT).

In the first experiment, a plano-convex microlens array with 150 μm pitch (Thorlabs MLA150-7AR-M) was used. It can be regarded as a purely phase test sample. Under the illumination wavelength of 532 nm, four pupil modulated images were acquired and processed by KKSAI reconstruction algorithm. The reconstructed phase map of a single microlens by KKSAI is shown in Fig. 3.9 (a), together with the reconstructions from PM-DPC and PM-FPM algorithm using the same measurements. We can clearly see that PM-DPC and KKSAI achieve better recovery of the microlens shape, providing the scanning electron microscope (SEM) image in Fig. 3.9 (b) as reference. To quantify their accuracy, we plotted the radially averaged cross-section profile from each result, labeled by green, blue, and red lines in Fig. 3.9 (c) respectively. According to the spec sheet of microlens array, the lens nominal profile is computed and labeled as a black line here. It can be seen that the phase profiles obtained by KKSAI and PM-DPC are in good agreement with the ground truth. The

FPM algorithm suffered from the low overlapping rate in Fourier domain and insufficient redundancy to be noise robust, thus resulting in a poor phase recovery.
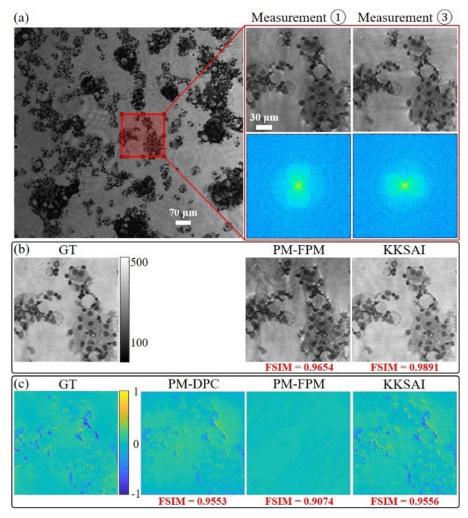


Figure 3.10: Experimental results for a thyroid carcinoma pap smear slide. (a) Two out of four measurements acquired by KKSAI and their Fourier amplitude spectrum; (b) Amplitude reconstruction by PM-FPM and KKSAI compared with ground truth; (c) Phase reconstruction by PM-DPC, PM-FPM, and KKSAI compared with ground truth.

Next, we imaged a papillary thyroid carcinoma pap smear slide with papanicolaou stain. It is a complex-valued sample. Figure 3.10 (a) displays two raw images acquired in the experiment. Due to the pupil amplitude modulation, the shade effect can be seen in the measurements and it is clear that their spectrum contains two cross-interference terms similar to those of an off-axis hologram, just as analyzed before. SLM allows us to carefully align

the aperture, such that these two cross terms are tangential to each other. Feeding the data into the KKSAI algorithm, we can then perform the complex field reconstruction and present the results in Fig. 3.10 (b) and (c).

To obtain the ground truth for this complex valued sample, we perform a separate PM-FPM experiment where 47 pupil modulation images with an overlapping rate of about 85% in Fourier domain were acquired. Its high-resolution reconstruction is taken as the ground truth to evaluate the performance of three methods. As we can see in Fig. 3.10 (b) and (c), PM-DPC cannot recover amplitude and PM-FPM reconstructions based on four measurements are of poor quality. By directly viewing the images, we can see that both amplitude and phase recovery of KKSAI are the closest to ground truth. To quantify the results, FSIM metrics are calculated and labeled under each reconstruction in red. From the quantifications, we can see that KKSAI clearly outperforms the other two methods.

Table 3.4: Similarity evaluation of overlapping spectrum regions in Fig. 3.10.

| Overlapping region |  |  |  |  |
|---|---|---|---|---|
| Phase | 0.9872 | 0.9927 | 0.9864 | 0.9691 |
| Amplitude | 0.9991 | 0.9998 | 0.9986 | 0.9991 |

Similar to the simulation, we can further evaluate KKSAI's performance by checking the consistency between reconstructions within the overlapping spectrum region from different scanning apertures. Here the FSIM metric is calculated and summarized in Table 3.4. As we can see, the recovery from different measurements matches very well in the overlapping regions.

Since the proposed KKSAI method can reconstruct the complex light field, we may apply digital post-processing operations to it. Here we show one possible application based on digital refocusing, axial chromatic aberration correction of imaging system. To demonstrate this, we image the same pap smear slide under the illumination wavelength of 405 nm (Thorlabs DL5146-101S), 532 nm and 638 nm (Thorlabs L638P040) sequentially.
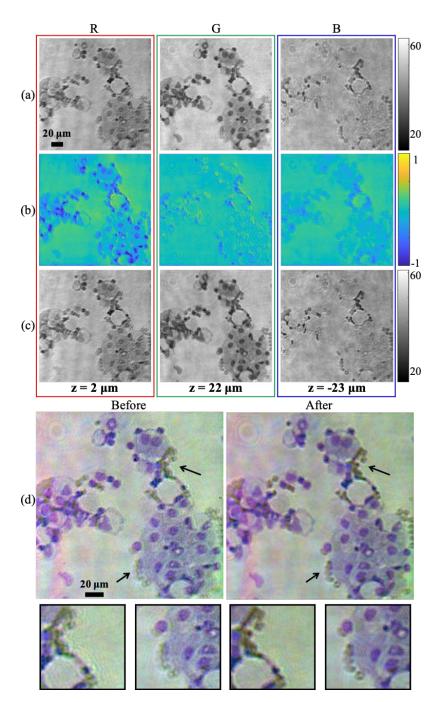
Figure 3.11: Chromatic aberration correction by digital refocusing ability of KKSAI. (a) Reconstructed amplitudes by KKSAI of three channels; (b) Reconstructed phases by KKSAI of three channels; (c) Digitally refocused amplitudes with the corresponding refocusing distance labeled at the bottom; (d) Color composite of three channels before and after digital refocusing with the enlargements showing improved image quality. R, red (638 nm); G, green (532 nm); B, blue (405 nm).

As shown in Fig. 3.11 (d), the chromatic focus shift causes rainbow fringes along boundaries in color composite image. It implies that there are still some residual chromatic aberrations associated with the objective. With the phase retrieved by KKSAI in Fig. 3.11 (b), we can utilize the angular spectrum method [7] to digitally propagate the complex wave-field in individual channels and an autofocusing metric based on edge sparsity criterion [96] to find the best focus location. Finally, we can obtain the optimal refocusing distance (labeled under each channel in Fig. 3.11 (c)). Comparing color composite image before and after digital refocusing in Fig. 3.11 (d), we can see that chromatic aberration artifacts are well corrected.

Finally, we conducted an experiment to prove the feasibility of reconstructing complex wave-field from only two measurements by KKSAI. The scanning scheme follows Fig. 3.4 (b). As shown in Fig. 3.12 (a), the spectrum amplitude of two raw images is in agreement with the simulation in Fig. 3.8 (a), although here the two cross terms are much stronger than the auto-correlation term, such that the auto-correlation term is barely discernible in the spectrum.

We can perform KKSAI reconstructions based on these two measurements. As shown in Fig. 3.12 (b) and (c), its result is compared to the KKSAI result based on four measurements of scanning scheme in Fig. 3.4 (a) and the aforementioned PM-FPM retrieval with 47 measurements. Taking PM-FPM result as the reference, the FSIM metric for two KKSAI reconstructions is computed and labeled in red at the bottom of each image. We can see that KKSAI reconstructions with only two measurements shows high conformity with PM-FPM ground truth. To conclude, our proposed KKSAI method can achieve complex wave-field reconstruction with almost the same quality as PM-FPM based on 47 measurements. Furthermore, it is iteration-free. As such, it is an attractive potential replacement for the current PM-FPM method.

To be clear, PM-FPM is distinct from the standard FPM [26]. PM-FPM does not have the capability to increase resolution beyond the limit of the objective NA, which is a key feature

of the standard FPM [26]. By extension, KKSAI with the current setup is not capable of increasing resolution beyond the objective NA.
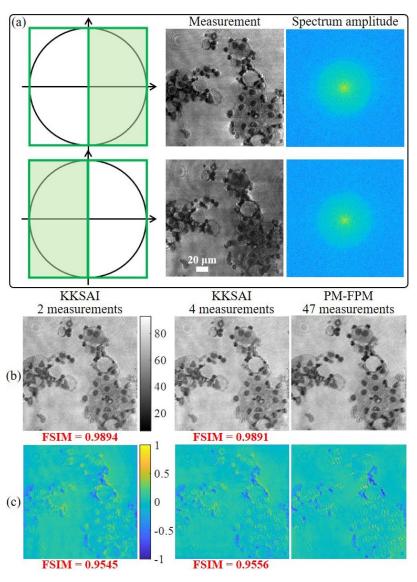


Figure 3.12: Complex wave-field reconstruction by KKSAI based on only two measurements. (a) Scanning scheme, raw measurements, and their spectrum amplitude; (b) and (c) Reconstructed amplitude and phase respectively, by KKSAI from two measurements, four measurements and PM-FPM with 47 measurements. Here, PM-FPM reconstruction is taken as the reference to calculate FSIM metric for KKSAI reconstruction.

In terms of reconstruction accuracy, KKSAI should be similar to Ref. [90] because they share the nearly identical mathematics. However, our KKSAI method is distinct from off-axis holographic microscopy [90] as it does not actually need a real reference arm. Compared to

Ref. [90], our system is simpler to implement. Also, SLM is not necessary for KKSAI. The SLM-based pupil modulation module can be replaced by a simple physical iris mounted and controlled by the motorized stage once the scanning scheme has been pinned down. By its simplicity, KKSAI is more robust to operate in practical situations. Besides, there actually exists a subtle difference between these two methods.

Mathematically, the off-axis holography experiment can be interpreted as the addition of a band-limited sample spectrum and a delta function in frequency domain. Their relative offset will decide the distance between two cross terms. In both KKSAI and off-axis holography [90], the delta function lies exactly on the edge of sample spectrum. But a key difference between the two methods is that for KKSAI there is a global offset for both sample spectrum and delta function such that the delta function happens to be located at the origin and is exactly the DC value. This explains the absence of interference fringes in the KKSAI measurements, unlike in off-axis hologram.

It is also worth mentioning that our pupil modulation mode can be easily adapted into its reciprocal mode, tilted illumination [97]. The multiplication with a phase ramp in spatial domain by tilted illumination is equivalent to the offset modulation in frequency domain. However, the advantage of our pupil modulation mode is that the thin sample requirement is circumvented, because it only focuses on the exit wave from sample.

An interesting point is that under tilted illumination mode, our method can be realized by only lighting up LED elements which are located on a ring with the illumination NA matching the objective NA. This is known as annular illumination scheme. It has been shown that only under this particular condition low-frequency phase components can be transferred into intensity measurements [94]. Although this conclusion is derived from the weak sample assumption, its coincidence with our findings indicates that matching illumination and objective NA or keeping the DC component just at the modulation aperture edge is an excellent operating point for computational imaging. This observation may inspire other new computational imaging methods. For example, the recovery of any coherent signal with a strong DC peak may benefit from the KKSAI concept.

## 3.7 Summary

In this chapter, a computational imaging method, synthetic aperture imaging based on Kramers-Kronig relations (KKSAI), for reconstructing the complex wave-field is reported. Its experimental setup, data acquisition and reconstruction algorithm are described in detail. By establishing the analogy between KKSAI measurement and off-axis hologram, a recent advance for off-axis holography based on KK relations [90] is adapted here to recover phase from intensity images in a non-interferometric way. The method leverages the analyticity of band-limited signals under pupil modulation to compute phase information from intensity measurements.

As a computational imaging modality, our KKSAI method co-designs the sensing part and the reconstruction algorithm to optimally operate together. From the perspective of sensing, it requires much fewer measurements than pupil modulation FPM. From the perspective of reconstruction, it is iteration-free and does not require any sample priors and is thus more generally usable than PM-DPC. Our simulation and experiment results demonstrate that KKSAI has clear advantages over the other two methods.

*Chapter 4*

# HIGH-RESOLUTION, LARGE FIELD-OF-VIEW LABEL-FREE IMAGING VIA ABERRATION-CORRECTED, CLOSED-FORM COMPLEX FIELD RECONSTRUCTION

*This chapter is adapted from the manuscript in preparation, R. Cao, C. Shen, and C. Yang, "High-resolution, large field-of-view label-free imaging via aberration-corrected, closed-form complex field reconstruction."*

In the previous chapter, we demonstrated that our proposed KKSAI has the potential to completely replace the pupil modulation FPM. It can seamlessly adapt to the tilted illumination mode by aligning the LED illumination NA with the objective NA. However, despite these advantages, it lacks two crucial features present in the tilted illumination mode FPM: aberration correction ability and optical super-resolution capability. These limitations restrict its universal applicability. To address this, in this chapter, we introduce two additional algorithms that complement KKSAI. The whole package method, known as analytical multi-angle illumination microscopy (AMIM), not only replicates the abilities of FPM but also outperforms it in terms of raw data volume, computational speed, reconstruction quality and robustness to aberrations. Our method achieves the same optical resolution as FPM with a reduced amount of raw data. Additionally, it provides analytical solutions for both sample spectrum and pupil function. Through various applications, ranging from aberration metrology to digital pathology, we demonstrate that our proposed method can effectively replace FPM.

## 4.1 Introduction

In Chapter 2, we provided a brief overview of the hardware development and application of FPM [98]. However, it is important to note that the heart of FPM lies in its reconstruction algorithm. A comprehensive summary of the FPM reconstruction algorithm has been provided from an optimization perspective [54].

The original Gerchberg-Saxton algorithm, which is based on alternating projection, can be viewed as a type of gradient descent method [26]. However, its convergence to the ground truth is relatively slow as it is a sequential first-order method. An alternative approach is the global version based on Wirtinger flow [99], which achieves faster convergence but at the expense of significant computational costs. Considering the need for faster convergence and stability compared to first-order methods, second-order methods are more favorable, specifically Newton methods. In our problem setting, two second-order methods, the sequential Gauss-Newton method [100] and the global Newton method [54], have been proposed. While the global Newton method slightly outperformed the former in terms of performance, it came with a considerable computational cost. Taking into account the considerations of convergence speed, algorithm stability, and computational cost, we ultimately selected the sequential Gauss-Newton method for our FPM discussion. This choice strikes a balance between convergence efficiency and computational feasibility.

Proposed as the embedded pupil function recovery (EPRY) algorithm [53], the pupil function $\mathbf{H}(\mathbf{k})$ can be reconstructed simultaneously with the sample spectrum $\hat{\mathbf{S}}(\mathbf{k})$, where $\mathbf{k}$ is the coordinate in frequency domain. Under the sequential Gauss-Newton framework, their update formula will be

$$\hat{\mathbf{S}}^{(m+1)}(\mathbf{k}) = \hat{\mathbf{S}}^{(m)}(\mathbf{k}) + \frac{\alpha}{\left|\mathbf{H}^{(m)}(\mathbf{k})\right|_{max}\left(\sum_i \left|\mathbf{H}^{(m)}(\mathbf{k}+\mathbf{k}_i)\right|^2 + \delta_1\right)}$$
$$\sum_i \left|\mathbf{H}^{(m)}(\mathbf{k}+\mathbf{k}_i)\right|\left[\mathbf{H}^{(m)}(\mathbf{k}+\mathbf{k}_i)\right]^*\left[\Phi_i^{(m)}(\mathbf{k}+\mathbf{k}_i) - \hat{\mathbf{S}}^{(m)}(\mathbf{k})\mathbf{H}^{(m)}(\mathbf{k}+\mathbf{k}_i)\right]$$

$$(4.1)$$

and

$$\mathbf{H}^{(m+1)}(\mathbf{k}) = \mathbf{H}^{(m)}(\mathbf{k}) + \frac{\beta}{\left|\hat{\mathbf{S}}^{(m)}(\mathbf{k})\right|_{max}\left(\sum_i \left|\hat{\mathbf{S}}^{(m)}(\mathbf{k}-\mathbf{k}_i)\right|^2 + \delta_2\right)}$$
$$\sum_i \left|\hat{\mathbf{S}}^{(m)}(\mathbf{k}-\mathbf{k}_i)\right|\left[\hat{\mathbf{S}}^{(m)}(\mathbf{k}-\mathbf{k}_i)\right]^*\left[\Phi_i^{(m)}(\mathbf{k}) - \hat{\mathbf{S}}^{(m)}(\mathbf{k}-\mathbf{k}_i)\mathbf{H}^{(m)}(\mathbf{k})\right]$$

$$(4.2)$$

Where $m$ is the iteration number, $\mathbf{k}_i$ is the frequency vector of corresponding LED with $i$ the LED index, $\Phi_i(\mathbf{k})$ is the multiplication of $\hat{\mathbf{S}}(\mathbf{k}-\mathbf{k}_i)$ and $\mathbf{H}(\mathbf{k})$ after updating with the $i$th measurement, $\alpha$ and $\beta$ are the step size for sample spectrum and pupil function update during iteration, $\delta_1$ and $\delta_2$ are regularization constants to ensure numerical stability.

Although this formulation grants aberration correction ability to FPM, it is accompanied by several challenges. Firstly, these two equations introduce multiple parameters for optimization. In its original proposal, these parameters were set as empirical values. However, in practice, it is necessary to adjust them based on different samples. Failure to do so may lead to convergence towards a local minimum, resulting in a low-quality reconstruction that deviates from the ground truth. This issue becomes more pronounced when dealing with severe aberrations. Unless a reliable initial guess for the pupil function is provided, the aforementioned formulation may prove to be ineffective [101]. Secondly, determining the termination condition for the iterative process is a delicate task that relies on human parameter selection. It requires careful consideration to strike the right balance between convergence and accuracy. Thirdly, the entire iteration process is time-consuming, particularly when applied to digital refocusing tasks. Multiple defocusing pupil functions need to be fed into the reconstruction algorithm, necessitating the repetition of the entire iteration cycle [102]. The time complexity of FPM poses a challenge for widespread adoption among general users.

The analytical solution to the pupil function and sample spectrum directly from the measured raw data would address the aforementioned challenges. In Chapter 3, we have made progress towards this goal. The chapter concludes that by aligning the LEDs such that their illumination NA matches the objective NA, we can utilize KK relations to analytically reconstruct the spectrum within the pupil [103].

However, KKSAI does have its limitations. When the illumination NA exceeds the objective NA, the DC term fails to enter the pupil, rendering its analogy to the hologram invalid. Consequently, we cannot apply KK relations for direct reconstruction of the corresponding spectrum in dark field data. This restricts KKSAI to achieving only double resolution of the original objective NA. Unlike the tilted illumination mode FPM, KKSAI lacks optical super-resolution capability. Another drawback is that KKSAI does not separate the pupil function and sample spectrum. Essentially, it is a phase retrieval method, where the reconstructed complex field is a combination of the pupil function and sample spectrum. While we can

apply digital refocusing to correct simple defocusing aberrations in the reconstructed field, estimating complex aberrations composed of various orders of Zernike modes, as demonstrated in Chapter 2 [104], becomes challenging. These limitations underscore the need for further advancements in order to overcome the restrictions of KKSAI.

Hence, in this chapter, we present two additional algorithms as complements to KKSAI, leveraging the insights gained from the relative locations of different parts of the spectrum. The whole package method, called analytical multi-angle illumination microscopy (AMIM), introduces a groundbreaking approach to computational imaging. AMIM starts from our KKSAI method for complex field reconstruction and employs analytical techniques for the retrieval of pupil function and high spatial frequency spectrum. This method demonstrates remarkable robustness against aberrations and provides an analytical solution for the complex field using fewer measurements compared to FPM. We showcase that our closed-form approach is capable of recovering significant aberrations that are typically challenging to address with FPM. Moreover, AMIM extensively utilizes darkfield measurements and achieves the same theoretical resolution enhancement as FPM, thus overcoming the limitations of KKSAI in terms of optical super-resolution.

## 4.2 Experimental setup

The experimental setup was based on the Olympus IX51 microscope. For our experiment, we utilized a 10X objective (Olympus Plan Achromat) with a NA of 0.25. The camera employed was a CMOS camera (Allied Vision Prosilica GT6400) with a pixel size of 3.45 μm. In terms of illumination, we customized it with an RGB LED matrix panel (32x32, 4mm pitch, Adafruit) and an RGB LED ring (16x, Adafruit NeoPixel), separately controlled by Arduino boards (Arduino Uno). The LED ring was first glued to the center of the LED matrix panel. Then this LED bundle was mounted on a six-degrees-of-freedom stage for alignment so that the illumination NA of ring LEDs matches well with the objective NA. The schematic of the imaging system can be seen in Fig. 4.1.

During the data acquisition process, we sequentially lit up a single LED and simultaneously triggered the camera to capture an intensity image when the LED was on. This process looped

through all LEDs used from the LED ring and matrix. This allowed us to capture data from different illumination angles. For demonstration, we utilized all LEDs with an illumination NA smaller than 0.5. According to the illumination NA of LEDs, we categorized them into three groups: bright field (BF) measurements, annular illumination (ANI) measurements, and dark field (DF) measurements, as depicted in Figure 4.1. In our set-up, the number of BF, ANI and DF images is 52, 16 and 248, respectively. In the following discussion, AMIM only utilizes ANI and DF measurements, whereas FPM requires all three measurements for its reconstruction process.
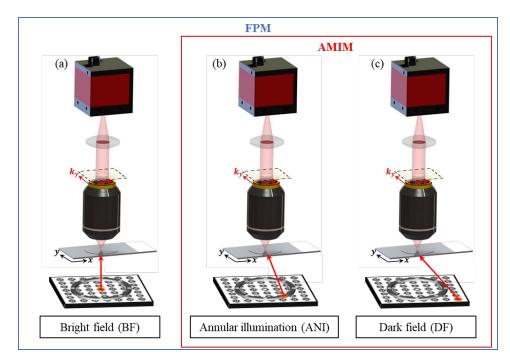


Figure 4.1: Data acquisition process of AMIM compared with FPM. (a) bright field (BF) measurement when the LED illumination NA is smaller than the objective NA. (b) annular illumination (ANI) measurement when the LED illumination NA matches the objective NA. (c) dark field (DF) measurement when the LED illumination NA is larger than the objective NA. AMIM only utilizes the ANI and DF measurements while FPM needs all three measurements.

## 4.3 Analytical reconstruction algorithm

The reconstruction process of AMIM begins with applying KK relations to recover the complex-valued spectrum within the scanning pupil aperture from the ANI measurements, followed by pupil phase estimation, which describes the optical aberrations. Then based on

the retrieved aberration-free sample spectrum, we further successively expand it by reconstructing the high spatial frequency spectrum part using DF measurements. Its entire reconstruction pipeline is presented in Fig. 4.2 (a).
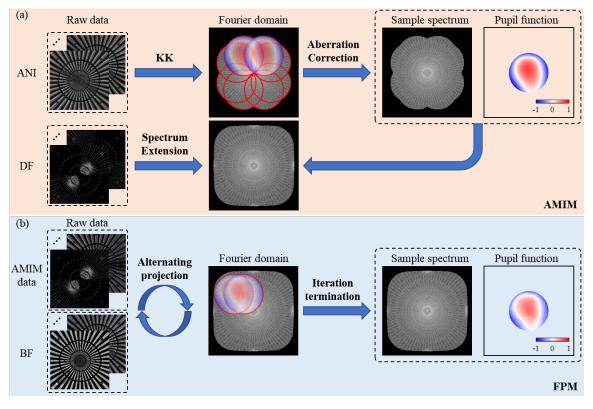


Figure 4.2: Reconstruction pipeline of AMIM compared with FPM. (a) AMIM first reconstructed the complex field corresponding to ANI measurements using KK relations, followed by analytical pupil function reconstruction to get the aberration-free sample spectrum and pupil function. Then, from DF measurements, the sample spectrum got further extended. (b) FPM reconstruction algorithm treated all three measurements as the same. It iterated between spatial and frequency domain to update the sample spectrum and pupil function estimate. In the spatial domain, the raw intensity measurements were used as the constraint while in the frequency domain, the pupil support was a constraint. Finally, when the preset termination condition was met, the iteration stopped, and the converged solution was taken as the final reconstruction results.

By comparison, the FPM reconstruction algorithm considers all three measurements to be of equal importance. It involves an iterative process that alternated between the spatial and frequency domains to update the estimates of the sample spectrum and pupil function. In the spatial domain, the raw intensity measurements serve as the constraint, while in the

frequency domain, the pupil support is used as a constraint. The algorithm iteratively refines these estimates until a predetermined termination condition is met. At that point, the iteration process stops, and the converged solution is considered the final reconstruction result, as depicted in Fig. 4.2 (b).

From Fig. 4.2, we can readily draw two key advantages of AMIM over FPM. Firstly, AMIM requires fewer measurements as it eliminates the need for BF images. This reduction in the number of measurements simplifies the data acquisition process, potentially saving time and resources. Secondly, AMIM stands out as an analytical method, eliminating the need for an iterative cycle. This feature allows for faster and more accurate results compared to FPM. By circumventing the iterative process, AMIM avoids convergence issues and tricky parameter selection associated with iteration, resulting in improved reconstruction speed and enhanced accuracy.

Given the detailed explanation of phase retrieval based on KK relations in Chapter 3, there is no need to delve into it further in this section. Instead, our focus will be on introducing two novel algorithms that specifically address the reconstruction of the pupil function and the DF spectrum, respectively.

To begin, we assume that the complex-valued spectrum within the scanning apertures corresponding to the ANI measurements has already been reconstructed using KK relations, as depicted in Fig. 4.3. Within each aperture, the spectrum represents the multiplication of the sample spectrum $\hat{\mathbf{S}}(\mathbf{k})$ and the pupil function $\mathbf{H}(\mathbf{k})$. Next, within the overlapping region of adjacent apertures, we observe that the sample spectrum remains identical, while the pupil function differs due to the shifting by tilted illumination, which can be denoted as $\mathbf{H}(\mathbf{k} + \mathbf{k}_i)$ and $\mathbf{H}(\mathbf{k} + \mathbf{k}_{i+1})$. As a result, a simple phase subtraction within the overlap will cancel out the phase of sample spectrum and only reveals the phase discrepancy between different parts of the pupil function. Building upon this insight, we can formulate a linear equation with respect to the vectorized pupil function from all overlaps. By solving this linear equation analytically, the imaging system aberration can be retrieved as the form of pupil function. It can be compensated in each scanning aperture so that we will obtain clean sample spectrums.

The corrected sample spectrums are then stitched together to obtain an aberration-free, two-fold resolution-enhanced sample image. This algorithm empowers the tilted illumination mode KKSAI with the ability to correct aberrations.
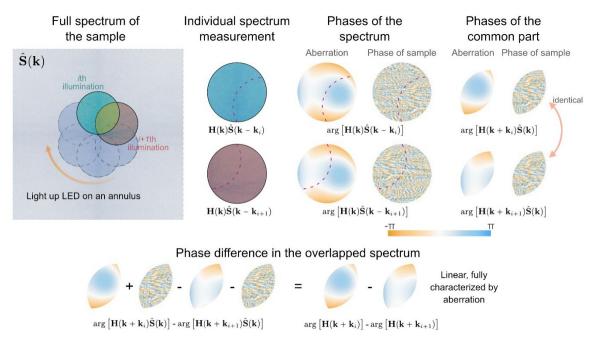


Figure 4.3: Analytical pupil function reconstruction. Within the overlapping region of adjacent apertures reconstructed from ANI measurements by KK relations, the sample spectrum is identical while the pupil part is different because the pupil function gets shifted by tilted illumination. Thus, the simple phase subtraction within the overlap will remain the difference between different parts of pupil function. Based on this insight, we can construct the linear equation on the pupil function and analytically solve it.

Next, we will outline the process of sequentially expanding the sample spectrum by retrieving the high-frequency components from the DF measurements. In this step, the reconstructed sample spectrum and pupil function obtained in the previous step act as initial knowledge. The overall concept is depicted in Fig. 4.4, providing a visual representation of the approach.

Spectrum of the E-field
(not directly measured)

$\hat{u}$  $\hat{p}$

Imaging system

Intensity measurement
(directly measured)

$\mathcal{F}$
$\mathcal{F}^{-1}$

FT of the measurement

$\hat{u} = \mathcal{F}(u)$    $\hat{p} = \mathcal{F}(p)$
(unknown)    (known)

$\mathbf{I} = |u + p|^2$

$\mathcal{F}(\mathbf{I}) = \hat{u} \star \hat{u} + \hat{u} \star \hat{p} + \hat{p} \star \hat{u} + \hat{p} \star \hat{p}$

Fourier transforms of the four terms of $\mathcal{F}(\mathbf{I})$

$\hat{u} \star \hat{u}$    $\hat{u} \star \hat{p}$    $\hat{p} \star \hat{u}$    $\hat{p} \star \hat{p}$ (known)

Extract the spectrum that is linear in the unknown part

Reconstruction

$\mathcal{F}(\mathbf{I})$ (measured) $-$ $\hat{p} \star \hat{p}$ (known) $=$ $\hat{u} \star \hat{u} + \hat{u} \star \hat{p} + \hat{p} \star \hat{u}$ → Part of $\hat{p} \star \hat{u}$ (linear in $\hat{u}$) → $\hat{u}$
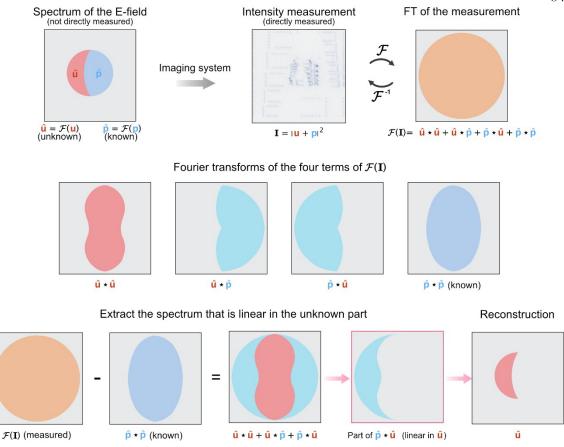
Figure 4.4: Analytical dark field spectrum reconstruction. Using the reconstructed spectrum from the previous step as prior knowledge, we can hypothetically separate the dark field spectrum into two components: an unknown part $\hat{u}$ and a known part $\hat{p}$ coming from the overlapping. Additionally, we have information about the intensity measurement of the dark field spectrum in the spatial domain. By expressing this intensity measurement in the Fourier domain using $\hat{p}$ and $\hat{u}$, it can be expanded into four terms. One of them is entirely dependent on $\hat{p}$ and is therefore known. By subtracting this known term and cropping out the non-overlapping portion of the remaining three terms, we obtain a region that is linearly dependent on $\hat{u}$. Based on this insight, we can construct a linear equation about $\hat{u}$ and solve it analytically.

We first choose a DF measurement whose spectrum overlaps with the known sample spectrum and hypothetically separate the DF spectrum into the known spectrum part $\hat{p}$ and the unknown spectrum part $\hat{u}$. Consequently, the problem transforms into the reconstruction of $\hat{u}$. An additional piece of information we have regarding the DF spectrum is the intensity measurement after it is transformed to the spatial domain. When expanded in the frequency domain using the notations $\hat{p}$ and $\hat{u}$, it consists of four terms: two autocorrelation terms of $\hat{p}$

and û, and two cross-correlation terms of p̂ and û, as illustrated in Fig. 4.4. Given that p̂ is known, its autocorrelation is also known. By subtracting the autocorrelation of p̂ from the Fourier transform of the intensity measurement, the remaining part comprises the summation of the autocorrelation of û and the two cross-correlation terms. Notably, we observe that there exists a region of the cross-correlation terms that does not overlap with the autocorrelation term. Therefore, we can isolate and crop out this region, which is linearly dependent on the unknown spectrum û. By vectorizing it, we can formulate a linear equation involving û and solve it analytically. Then, we can extend the known sample spectrum by stitching the retrieved û.

By repeating this process for each DF measurement in a sequential manner, following the ascending order of the illumination NA magnitude, we can achieve a high-resolution reconstruction of the sample spectrum. An important aspect to consider during this process is the presence of aberrations. Aberrations should first be introduced to the known sample spectrum p̂ to align it with the raw measurement. After recovering the unknown spectrum û, aberrations need to be removed again before we stitch û to the known aberration-corrected sample spectrum.

Once all the aforementioned steps are completed, we are able to obtain a high-resolution sample image that is free from aberrations. The theoretical resolution of this image is determined by the sum of the maximum illumination NA and the objective NA, which is equivalent to the theoretical resolution of FPM. However, it is important to note that FPM requires an iterative process to recover the spectrum and is sensitive to the choice of optimization parameters. In contrast, AMIM directly recovers the actual spectrum without any convergence issues. Furthermore, we will demonstrate that AMIM performs well even with significant aberrations, whereas FPM struggles due to the increased complexity of its optimization problem.

## 4.4 Experimental results and discussion

We performed a series of experiments to evaluate the effectiveness of AMIM and compare it with FPM.

In our first experiment, we used a Siemens star standard sample placed at the center of the camera FOV to assess the practical optical resolution achievable by AMIM and FPM [65]. The reconstruction results are presented in Fig. 4.5. As observed, both methods successfully resolved the Siemens star center. To better visualize the resolution limit, we plotted a line profile along a circle near the theoretical optical limit, as shown in Fig. 4.5(c). The line profiles show that both methods achieved isotropic contrast, indicating effective aberration correction, and reaching diffraction-limited resolution. However, we noticed that FPM reconstruction exhibited peculiar spurs at the trough locations, which were absent in our AMIM reconstruction. This artifact is attributed to the fact that the FPM reconstruction algorithm aims to obtain a self-consistent solution, sometimes leading to convergence to a local minimum near the true global minimum.
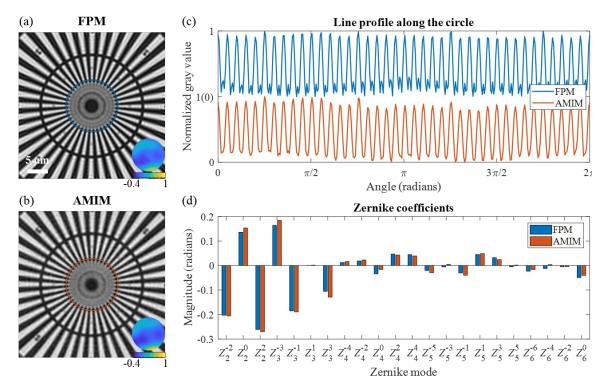
Figure 4.5: Resolution verification using Siemens star. (a) FPM reconstruction with its recovered pupil phase function. (b) AMIM reconstruction with its recovered pupil function. (c) The normalized gray value line profile along the circle in reconstructions. (d) Zernike decomposition of recovered pupil phase function.

The second aspect we investigated was the aberration correction capability of AMIM. By examining the recovered pupil phase maps depicted in Fig. 4.5 (a) and (b), it is evident that the aberration estimates obtained by both methods closely align with each other. To perform a quantitative comparison, we decomposed the aberration estimates into Zernike modes and plotted the magnitude of these modes. The matching between the two methods can be observed in Fig. 4.5 (d). To sum up, AMIM exhibits equivalent optical super-resolution and aberration correction capabilities to FPM, while requiring fewer raw measurements and exhibiting fewer reconstruction artifacts.
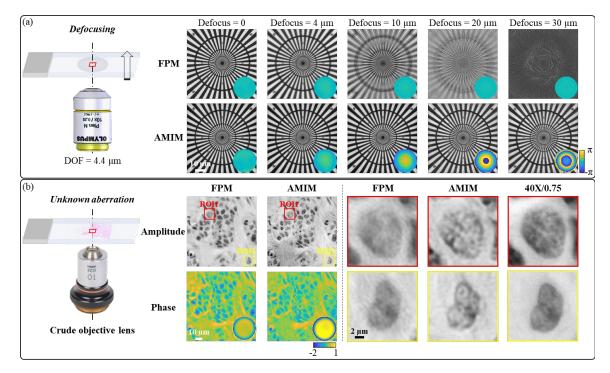


Figure 4.6: Aberration correction ability test. (a) FPM and AMIM reconstruction under different levels of defocusing. (b) FPM and AMIM reconstruction using a crude objective lens.

Next, we aimed to assess the limit of aberration correction ability of AMIM in comparison with FPM. We first focused on studying defocusing as a representative sample-dependent aberration, which is commonly encountered in various application scenarios such as slide scanning in pathology [105]. In terms of the pupil function, defocusing can be characterized by a quadratic phase term that varies quadratically with the radial distance from the center of the pupil. To simulate defocusing, we employed a $z$-axis piezo stage (Thorlabs MZS500-E)

to control the axial position of the Siemens star sample, intentionally introducing different levels of defocus. At each axial position, we collected data and fed it into both AMIM and FPM for reconstruction. The results are presented in Fig. 4.6 (a). Within the nominal depth of field (DOF) of the objective, both FPM and AMIM achieved successful reconstructions after accurately estimating the pupil function. However, beyond the DOF, FPM reconstruction significantly deteriorated, while AMIM continued to deliver satisfactory reconstructions by accurately estimating the defocused pupil function, even up to a defocusing of 30 μm, which corresponds to approximately 7 times the nominal DOF. It is important to note that in this comparison, neither FPM nor AMIM had prior knowledge of the defocus. FPM initiated its pupil estimation from a random guess. Therefore, we can conclude that in the case of high aberration, FPM is more likely to fail in reconstruction without a good initial guess.

The second case involved using a crude objective lens from Olympus that was designed many years ago and exhibits severe aberrations. In this experiment, we imaged a cancer pathology slide stained with hematoxylin and eosin (H&E). The results are presented in Fig. 4.6 (b). As observed, the aberration pattern appears as a peculiar asymmetric donut shape. To establish a ground truth for comparison, we also imaged the same slide under a 40X objective with an NA of 0.75. Upon zooming in on various small regions, we can clearly observe that the amplitude reconstructions obtained by AMIM closely match the ground truth, showcasing finer details and fewer artifacts. This comparison highlights the superior aberration correction ability of AMIM over FPM, as it consistently yields more accurate and high-quality reconstructions.

Lastly, we would like to highlight an important application of AMIM: digital pathology. When FPM was initially introduced, it was regarded as a highly promising technique for digital pathology [85]. However, in practical usage, pathologists have expressed certain concerns about FPM. Aside from the long computational time, FPM often produces color images that exhibit discrepancies compared to conventional wide-field microscope color images captured under white light [106]. Specifically, there is a noticeable color shift. As

shown in Fig. 4.7, a normal H&E-stained pathology slide appears purplish when observed using incoherent illumination and a wide-field microscope. However, color images generated by FPM tend to have a reddish tone, even after applying white balancing. This indicates that FPM alters the ratio between the background intensity and the higher frequency components in each color channel, while still maintaining well-reconstructed contrast.
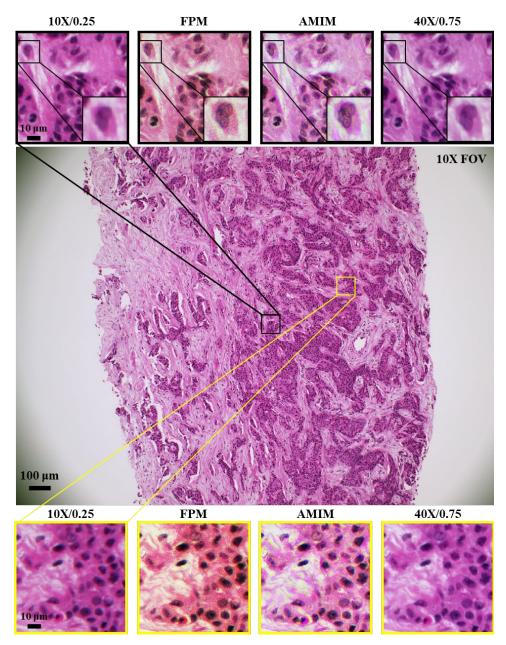


Figure 4.7: Demonstration of digital pathology application by AMIM.

In contrast, our AMIM reconstructions demonstrate more accurate color rendering at different locations within the FOV. Additionally, upon closer examination, we observed artifacts in the FPM reconstructions. We specifically investigated the reconstruction results for each color channel, as depicted in Fig. 4.8. Since we focused on the sample using the green channel and there were residual chromatic aberrations associated with the objective lens used, the sample in the blue and red channels experienced slight defocusing. From Fig. 4.8, it is evident that AMIM successfully corrected these aberrations without any adverse effects, while FPM failed to correct the aberration in the blue channel, resulting in strong artifacts. These findings strongly suggest that our proposed AMIM algorithm offers a superior solution for pathology imaging applications.
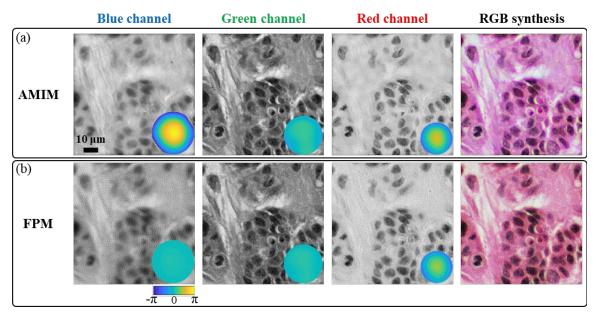


Figure 4.8: Investigation of reconstruction by AMIM and FPM for each channel. (a) AMIM reconstruction with the recovered pupil function in each channel. (b) FPM reconstruction with the recovered pupil function in each channel.

## 4.5 Summary

In summary, our study demonstrates the superiority of AMIM over FPM in terms of extracting large aberrations and synthesizing high-resolution complex-valued images with a large field of view using low magnification objectives. AMIM proves to be robust against aberrations, while FPM struggles to produce aberration-free, high-resolution reconstructions

when confronted with highly aberrated measurements. AMIM also overcomes the drawbacks of conventional phase retrieval algorithms, such as parameter sensitivity and convergence issues. It provides accurate results without the need for iterative processes or parameter selection, making it a more reliable and efficient method. However, proper calibration is essential for AMIM to account for positional errors of LEDs and ensure accurate solutions. Compared to FPM, AMIM offers computational advantages by directly solving for the complex field and reducing processing time. With its robustness, efficiency, and versatility, AMIM opens up new possibilities for label-free computational microscopy and has the potential to revolutionize micro-scale research and the design of optical instruments.

*C h a p t e r   5*

# AUTOMATIC DETECTION OF CIRCULATING TUMOR CELLS AND CANCER ASSOCIATED FIBROBLASTS USING DEEP LEARNING

Beginning with this chapter, we transition from model-based computational imaging to data-driven computational imaging. We will utilize deep learning techniques to solve various image analysis tasks for biomedical applications. This chapter concentrates on a specific task of merging optical system design with deep learning-based algorithms to fully automate the cell detection process.

Circulating tumor cells (CTCs) and cancer-associated fibroblasts (CAFs) from whole blood are emerging as important biomarkers that potentially aid in cancer diagnosis and prognosis. The microfilter technology provides an efficient capture platform for them but is confounded by two challenges. First, uneven microfilter surfaces makes it hard for commercial scanners to obtain images with all cells in-focus. Second, current analysis is labor-intensive with long turnaround time and user-to-user variability. In this chapter, we addressed the first challenge through developing a customized imaging system and data pre-processing algorithms. Utilizing cultured cancer and CAF cells captured by microfilters, we showed that images from our custom system are 99.3% in-focus compared to 89.9% from a top-of-the-line commercial scanner. Then we developed a deep-learning-based method to automatically identify tumor cells serving to mimic CTC (mCTC) and CAFs. Our deep learning method achieved precision and recall of 94% (±0.2%) and 96% (±0.2%) for mCTC detection, and 93% (±1.7%) and 84% (±3.1%) for CAF detection, significantly better than a conventional computer vision method, whose numbers are 92% (±0.2%) and 78% (±0.3%) for mCTC and 58% (±3.9%) and 56% (±3.5%) for CAF. Our custom imaging system combined with deep

learning cell identification method represents an important advance on CTC and CAF analysis.

## 5.1 Introduction

Approximately 90% of all cancer-related deaths are due to the process of metastatic spread [107, 108]. During metastasis, tumor cells detach from the primary tumor, migrate into blood vessels and these circulating tumor cells (CTCs) travel to distal organs through the circulatory system. CTCs tend to survive even in the harsh circulatory environment and retain the ability to proliferate and create secondary tumor sites in distant organs [109, 110]. Over the past decades, there has been increasing interest in understanding the metastatic cascade with respect to how CTCs gain access to the circulatory system, survive in circulation, and thereafter create secondary metastatic sites, as well the biology behind tropism towards distant organ sites for different tumor types [111–115]. This has brought CTCs to the spotlight in a rapidly expanding field of liquid biopsy and cancer diagnostics.

CTCs can be found as either single cells or clusters, with the latter being formed by tumor cells alone or more significantly, as clusters of multiple cell types such as peripheral blood mononuclear cells (PBMCs) and circulating cancer associated fibroblasts [116–118]. It has become increasingly clear that in addition to CTCs, the circulating cells in the tumor microenvironment (TME) also play an important role in metastasis. Cancer-associated fibroblasts (CAFs) are a group of activated fibroblasts with significant heterogeneity and plasticity in the TME [119–122]. It has been shown that CAFs also gain access to the circulatory system and play a significant role in the process of metastasis. They modulate cancer metastasis through synthesis and remodeling of the extracellular matrix (ECM) and production of growth factors. Thus, targeting CAFs has gained considerable interest and is being explored as an avenue for novel cancer therapies.

In general, CTCs are extremely rare in blood samples and range from one to a hundred cells in a 7.5 mL tube of human blood, depending on the stage of the disease [115]. CTC clusters are even rarer and constitute only 2-5% of all CTCs [123]. However, CTC clusters are disproportionately efficient at seeding metastases. Therefore, CTCs and CTC clusters are

valuable biomarkers to determine prognosis, monitor therapy, assess risk of recurrence, and possibly in the future to aid in early cancer detection and screening [109–118]. Over the past decade, numerous CTC capturing techniques have been developed [123–126]. The different enrichment and isolation processes either leverage the morphological characteristics of CTCs which include size/deformability-based separation [127–129] and density-gradient centrifugation [130, 131], or rely on immunoaffinity-based separation through targeting specific cell surface epitope expression [132, 133]. The identification and characterization of CTCs can be performed in several ways: immunofluorescence [134], fluorescence in situ hybridization (FISH) [135], real-time reverse transcription polymerase chain reaction (qRT-PCR) [136], genomic analysis [128], and RNA sequencing [137]. Capturing of CAFs can also be achieved through similar methods used for CTC [120, 138, 139]. Our team has developed a size-based enrichment technology using a membrane microfilter device with defined pore size to isolate CTCs and CAFs from whole blood as well as other bio-fluids [140]. Exploiting the size characteristics of CTCs and CAFs (being larger than the vast majority of normal blood cells), the pores allow the smaller blood cells to be filtered through while capturing the larger CTCs and CAFs along with some extraneous, larger sized mononuclear cell lineages. Among the methods discussed above, size-based isolation techniques combined with immunofluorescence labelling have gained significant traction due to the simplicity of this technique, its unmatched capture and enrichment efficiency, and its ability to capture and enrich CTC of different embryologic origin (e.g., epithelial vs. mesenchymal), because size-based filtration of CTC does not rely on the immunoaffinity methods.

This approach has two technical challenges that are the focus of our present study. The first challenge is that while the microfilter membranes (approximately 15 microns thick) appear to be flat and uniform, they are flexible and have slightly wavy surfaces, thus presenting widely varying focal planes. The small height variations of these surfaces will be greatly magnified under the high powers of microscopy needed to identify and characterize CTC and other captured cells. Since the captured cells are randomly distributed across the wavy microfilter surface, they are, in turn, distributed across different focal planes during

microscopic imaging. This makes it difficult to utilize commercially available slide scanners to collect whole filter scans with all cells in-focus and requires robust auto-focusing ability of the imaging system across not only the different fields-of-view (FOVs) but also even within the same FOV. Commercial scanners are ill-suited to handle this degree of focal plane heterogeneity [105], especially within the same FOV. Currently, the only way to address this is to acquire multi-focal-plane data ('$z$-stacking') for each FOV so that all cells within each FOV can be digitally focused post-imaging.

Second, detection of CTCs and CAFs on the microfilter is currently a manual process. The user needs to evaluate a large number of captured cells to identify the very few that are of interest (CTCs, CAFs), and these cells are often morphologically similar to other captured cells, making the analysis of these filters laborious, inefficient, and subjective. Automating the cell localization and classification process has been attempted by employing image processing and machine learning (ML) algorithms on fluorescent microscope images of CTCs [141–144]. The general workflow for such a process requires several steps. Initially, a region of interest (ROI) identification algorithm is designed to segment all possible CTC or CAF events out of the whole slide images [144]. Then the pre-screened image patches are normalized and transformed to a feature space by human engineering. For example, the features could be pixel gray value histograms in each red-green-blue (RGB) channel [141] or vector arrays in hue-saturation-value (HSV) color space [142] or morphological characteristics including area, perimeter, and eccentricity [143]. Finally, these feature representations of events are classified by conventional machine learning models, such as naive Bayesian classifier [141, 143], support vector machine [141–143] or random forest [142, 143]. This type of cell detection workflow is not holistic, and the accuracy of upstream steps will inevitably affect the performance of downstream ones as detection errors accumulate. In addition, these methods are, at best, semi-automated, as they rely heavily on manual input, such as the parameter tuning for segmentation process and the manual decision on feature engineering. We believe that deep learning (DL) is a better method, as it has the advantages of enabling self-learned feature engineering and end-to-end training. To date,

only a few studies [145, 146] have explored applying DL for CTC analysis, and there are none so far on the use of DL for CAF analysis.

In this chapter, we have addressed both challenges and present a comprehensive scheme to automatically detect and analyze tumor cells and CAFs that are captured by our microfilter device. We utilized a model system of cultured cancer cells to mimic the processing of a clinical sample but in a controlled fashion. We refer to these cultured tumor cells as mimicked CTCs (mCTCs). First, we have developed a hardware imaging solution that can collect high quality, uniformly in-focus images. Second, we have used these high-quality in-focus images for deep learning model training and subsequent analysis of new samples using DL.
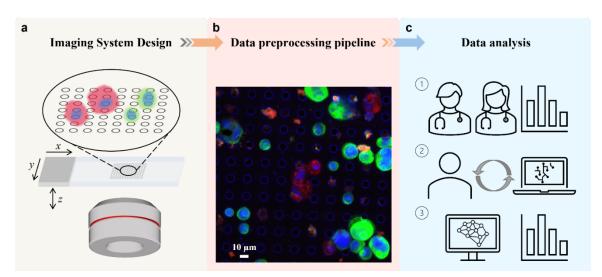


Figure 5.1: Schematic of overall design. a. Multi-channel epifluorescence microscope imaging system. Since our target cells are distributed on the micro-filter at varied heights, the sample is three-dimensional in nature. They are scanned axially under four channels to fully capture the cell-specific biomarker expression. b. Data preprocessing pipeline. The raw image data are synthesized into a single multi-color all-in-focus whole slide image for further analysis. c. Data analysis. The classical way to detect CTCs and CAFs relies on human experts. ① First, the experienced pathologists review the whole slide, annotate cells of interest, and count their number. ② Then this annotation paired with fluorescence images is used to train a deep learning model. Because of inherent human observer bias in calling or ignoring positive cells, the prediction from the pre-trained deep learning model is used to cross-validate human expert annotation. ③ Finally, the well-trained deep learning model can independently conduct the cell detection and analysis task.

The overall approach of this work consists of three stages depicted in Fig. 5.1: (a) Autofocusing and axial scanning design of our customized microscope system to ensure that information of all cells located at different focal planes can be acquired. (b) An optimized data preprocessing pipeline to synthesize raw multi-channel multi-slice images into a single all-in-focus (AIF) RGB whole slide image for each sample. (c) A collection of these AIF slide images that are annotated by human experts and subsequently used to train deep-learning-based mCTC and CAF detection systems. The trained systems thereafter can be used to automate the analysis of new samples.
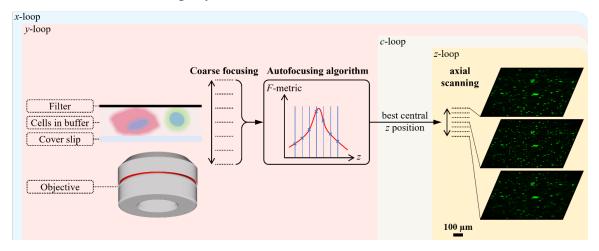
## 5.2 Sample preparation

We did not use any clinical blood samples from patients. Instead, we used cell lines to mimic the circulating tumor cells and cancer associated fibroblast cells from blood. To create controlled, synthetic cell samples that would roughly mimic the target human cell types expected to be found in clinical blood samples from breast cancer patients, human breast cancer (SKBR3) cells were obtained from the American Type Culture Collection (ATCC) and cultured using phenol red McCoy's 5A medium supplemented with heat-inactivated 10% fetal bovine serum (FBS). The primary CAF23 cell lines were acquired from our collaborator lab (Dr. Dorraya El-Ashry) and have been previously isolated and characterized by her group [147], which were maintained in phenol red Gibco's Improved Minimum Essential Medium (IMEM) supplemented with heat inactivated 10% FBS. All the cells were grown in Heracell VIOS cell incubators at 37°C with 5% $CO_2$. Cells were passaged continuously by detachment using TrypLE.

Model systems were created for the development of the imaging scanner and cell identification algorithm. SKBR3 and CAF23 cells were cocktailed together and spiked into 7mL 1X phosphate buffered saline (PBS). The cells were spiked within a range of 500-1000 of each cell type to create a variety of model systems during testing. The spiked cells in 1X PBS were processed using the Circulogix faCTChecker instrument. Post processing the microfilter slides were retrieved from the instrument with the fixed cells captured on the microfilter membrane, ready for downstream immunofluorescence staining and imaging.

SKBR3 and CAF23 cells were detected by double immunostaining using markers that would allow for identification for each cell type, namely pan-cytokeratin (CK) for SKBR3 and fibroblast activation protein (FAP) for CAFs. Samples were blocked for 1 hour with blocking buffer consisting of 5% normal goat serum (Life Technologies) and 0.3% Triton X-100 (Sigma Aldrich) in 1X PBS and then incubated overnight at 4ºC with rabbit anti-human FAP (Millipore). Samples were then incubated with a goat anti-rabbit Alexa 594 labeled secondary antibody (Life Technologies) at room temperature for 1 hour. Samples were then incubated with Alexa 488 pre-conjugated CK antibody (eBioscience). Samples were counterstained with a combination of DAPI for twelve minutes followed by Hoechst for five minutes, to optimize signal and minimize quenching in the DAPI channel (as it is used as the focus channel), which was experienced during initial testing. Finally, the samples were mounted on coverslips with ProLong Diamond Antifade Mountant (Life Technologies).

## 5.3 Customized microscope system



Figure 5.2: Auto-focusing principle during scanning. First, a coarse scanning with large step size over a wide *z* range is performed. Then, the image at each *z* position is used to calculate the focus measure (*F*-metric). The best focus *z* position is then estimated as the peak location by fitting a Gaussian function to discrete *F*-metrics. Centered on this estimated best focus *z* position, a fine axial scanning with small step size is performed to capture the whole 3D information. Autofocusing is repeated for every lateral *x-y* scanning position and executed only in DAPI channel. The estimated best focus *z* position will be used across all channels. Chromatic aberration can be compensated by axial scanning.

Commercially available slide scanners that host autofocusing functions have been unable to provide a uniform in-focus whole scan of our microfilter. Without the ability to acquire AIF whole scans, it is difficult to analyze clinical samples with high accuracy and confidence and at times unable to analyze them at all. Therefore, we set out to emulate the manual focusing process with a customized scanning strategy for an epifluorescence microscope.

The whole imaging system is composed of an inverted epifluorescence microscope (Olympus IX83) with an achromatic objective lens (Olympus PLN 20X, NA 0.4), a motorized lateral scanning stage (Thorlabs MLS203-1 and BBD302) and a high-performance CMOS camera (ISVI IC-M50T-CXP). The control of microscope is through µManager [148] and a Python wrapper was written on top of it to interface with the microscope. More specifically, it controlled the motorized fluorescent mirror turret, epi-shutter and motorized focus controller of microscope. The lateral scanning stage is controlled by a Python module wrapping Thorlabs APT library [149] and the camera is communicated with the computer via a Python image acquisition library, Harvester [150]. Finally, all communication functions are assembled in a Python environment.

As shown in Fig. 5.2, we started this process by performing coarse focusing with a large $z$ step size over a large range. Next, we calculated the focus measure metric ($F$-metric) based on this $z$ stack. There have been extensive studies on different types of autofocus metric functions for fluorescence microscopes [151]. For our purpose, we used the Tenegrad function, a robust and simple $F$-metric. To achieve an accurate estimate of the best focus position, we fitted a Gaussian function to the discrete $F$-metric sequence from $z$-stack and took its peak as the final best focus position, with which we further conducted the fine axial scanning with small $z$ step size. Through fine scanning, cells located at different focus planes can be digitally revisited afterwards or synthesized into a single 2D AIF image. For the objective lens used in our system, its depth of field was 1.72 µm. The coarse $z$-axis focusing step size was 20 µm with a scan range of 200 µm and the fine scanning $z$ step size was 5 µm with a scan range of 50 µm.

About the selection of *F*-metric, we investigated different kinds of autofocus metric functions. From Fig. 5.3, we can see from the central slice that a single *z*-slice is not capable of presenting all cells in good focus. Maximum intensity projection, where we project the maximum pixel value along the *z*-axis, is one possible way to resolve this. However, as can see from the image, this simply creates an image with significant haze artifacts. Local-level patch selection AIF methods can avoid this problem as long as the focus measure is correctly chosen. Here, Tenegrad function is chosen as our metric because it can find the correct focus *z* position with excellent sensitivity.
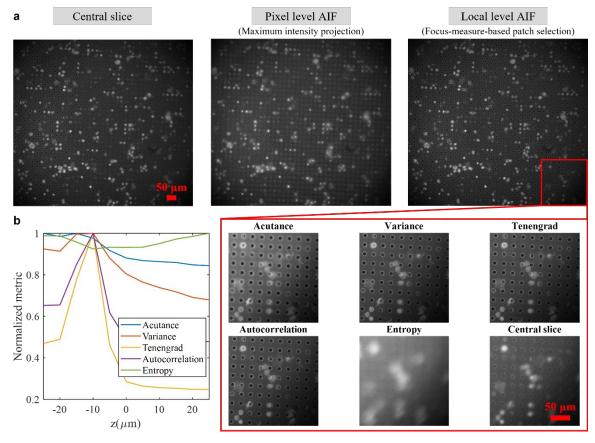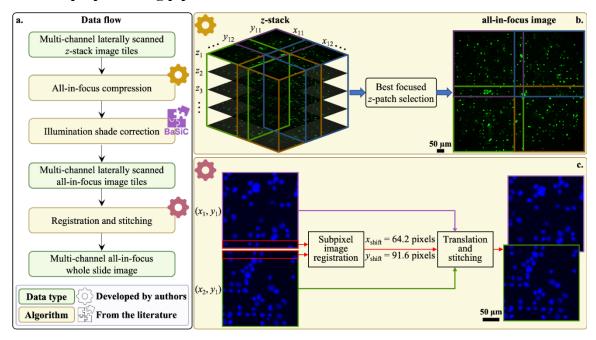


Figure 5.3: Comparison of different all-in-focus (AIF) compression methods. a. Central slice of a *z*-stack from DAPI channel, compared with pixel level AIF compression result based on maximum intensity projection and local patch level AIF compression results based on different *F*-metrics. b. Normalized curve of different focus measures versus *z* positions.

We can apply this strategy to image a large sample by repeating the process for overlapping FOVs across the entire sample by lateral scanning. The best *z*-focus estimation only needs to

be run once for each lateral position and thereafter is applied across all the fluorescence channels. The number of fluorescence channels utilized is dependent on the immunofluorescence staining protocol for the biomarkers used for mCTC and CAF identification. Residual axial chromatic aberration from the objective can be addressed when we process the fine-step $z$-stack data collected in the second step.

## 5.4 Data preprocessing pipeline



Figure 5.4: Data preprocessing pipeline. a. Data flow starting from raw measurement and ending with a multi-channel all-in-focus whole slide image. Preprocessing consists of three algorithms, among which two are developed by authors and the other one is adapted from an existing work. b. Principle of all-in-focus compression. $Z$-stack at each $x$-$y$ location is split into smaller patches and the best focused $z$-patch is selected with focus measure. Finally, $z$-patches are fused into an all-in-focus $x$-$y$ tile. c. Principle of registration and stitching. There is overlap between adjacent $x$-$y$ tiles due to the tilt between scanner lateral movement coordinates and camera frame coordinates. Subpixel image registration algorithm relies on the overlapping region to find the subpixel shift between two adjacent $x$-$y$ tiles. Taking the upper left corner tile $(x_1, y_1)$ as the anchor for the final mosaic, all other $x$-$y$ tiles are translated and stitched to it by blending based on distance transform.

For each microfilter slide, we acquired multi-channel laterally scanned $z$-stack image data by the above process. Our ultimate purpose was to render a single AIF RGB whole slide image so that it can be effectively processed by the subsequent DL model without defocus-

associated inaccuracies. As a sidenote, this single AIF whole slide image is also a good fit for pathologist's workflow as it eliminates the need to manually retune the image focus during analysis. In our experiment, this rendering process consisted of three steps, summarized in Fig. 5.4.

First, we compressed the images' $z$-dimension by cropping the image tile for a single FOV into smaller patches with the size of 600 x 600 pixels and selecting the $z$-slice with highest focus quality for each patch. (The patches were partitioned in such a way that there is some overlap with neighboring patches.) This allowed us to flexibly put cells in the single FOV, which may be located at different $z$-planes, into focus. We again adopted the Tenegrad function in selecting the best focus quality. As the collected fluorescence image can be affected by non-uniform excitation light profile due to vignetting effect and temporal background variation, we next normalized the image brightness spatially. To accomplish this, we used an established shade correction method, BaSiC [152]. Finally, we used a customized registration algorithm and leveraged the overlap between neighboring patches to stitch these patches together into the entire whole slide FOV. This customized stitching algorithm was necessary to accommodate for tilts between the lateral movement coordinates of the motorized scanner and the camera frame, which prevented a simple direct stitching of the patches. For our experiment, we typically acquired a whole slide FOV of size 6.9 mm x 6.9 mm. Our objective allowed us to acquire single image tile FOVs that were 0.92 mm x 0.85 mm, corresponding to 4000 x 3700 pixels. Between two neighboring patches, the overlap was 310 pixels.

This combination of optimized auto-focusing and image processing algorithm created a whole slide image of the microfilter with high focus quality. To verify its performance, we compared our system with a commercial slide scanner, Olympus VS120, which we have used in the past to perform whole microfilter scans. Results are summarized in Fig. 5.5. The same sample was scanned using both our custom developed scanner and the Olympus VS120 scanner. Two whole slide images were registered, and the maximum overlapping region was displayed.
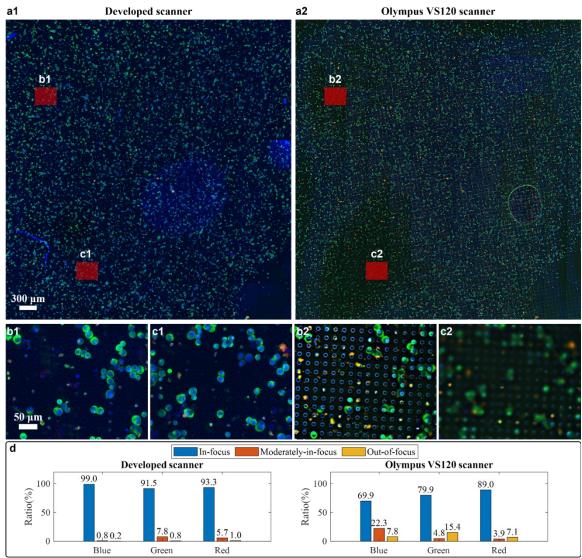
Figure 5.5: Comparison of the whole slide image focus quality by our developed scanner and Olympus VS120 scanner. a1. Whole slide image (WSI) of a model sample under 20X objective from our developed scanner. a2. WSI of the same model sample under 20X objective from Olympus VS120 scanner. b1, b2 and c1, c2 are the zoom-in on the same regions from two WSIs. Their area size is the same as the image tile from VS120 scanner, 366 μm x 287 μm. d. Quantitative analysis of focus quality of WSI from both scanners in blue, green, and red channel.

From Fig. 5.5 (a), we can clearly see that our developed system was able to achieve a more uniform focus quality than the Olympus scanner. The commercial scanner only performs focus selection at the image tile level (nominally at an area size of 366 μm x 287 μm) while our system performs focus selection at a smaller patch level (nominally at an area size of 138

µm x 138 µm). Moreover, the commercial scanner was unable to correctly focus over a significant fraction of the sample, as can be seen in areas noted in Fig. 5.5 (b) and (c).

To quantify the AIF performance difference in this characterization experiment, we utilized a deep neural network model that predicts an absolute measure of focus quality at a patch level for fluorescence images [153]. The model was originally trained on wide field fluorescence images of U2OS cells with Hoechst-stained nuclei. It has demonstrated robust generalization capabilities to other stains, such as Tubulin and Phalloidin, as well as the MCF-7 cancer cell type, despite not being explicitly trained on them. Given that our images also involve wide field fluorescence imaging of cancer cells, we anticipate that this method should generalize well to our dataset as well. This model outputs a probability distribution over 11 discrete focus levels for a patch with 84 x 84 pixels; the focus level with highest probability is taken as the final prediction. We ran this model, patch by patch, over each channel of two whole slide images and generated a focus quality map with the granularity of 84 x 84 pixels. For better visualization, a bar chart to summarize the portion of patches with different focus quality for two WSIs is shown in Fig. 5.5 (d). Here, the first three focus levels are binned into 'In-focus' class, the middle four focus levels are binned into 'Moderately-in-focus' class and the last four focus levels are taken as 'Out-of-focus' class.
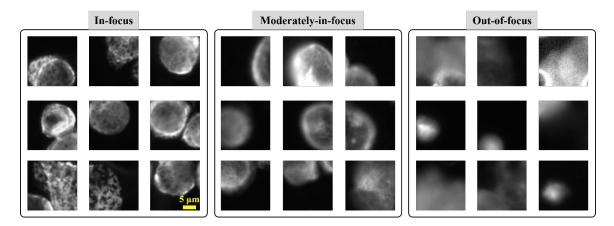


Figure 5.6: Three focus quality classes derived from a reported absolute measure method.

In order to ensure that the predicted focus quality measures align with human vision evaluation, we conducted a manual verification. Our visual evaluation indicated good

agreement between the predicted focus quality measures and the actual images. Specifically, the patch images classified as 'In-focus' showed clear cell cytokeratin details, while the 'Moderately-in-focus' images still display recognizable cell cytokeratin features. In contrast, the 'Out-of-focus' images barely resembled cells and often resembled precipitates. We have provided several examples of each class in Fig. 5.6. Our custom scanner performed with an average 99.3% of regions across different channels in the WSI either in-focus or moderately in-focus. In contrast, the in-focus or moderately-in-focus portion only occupies an average of 89.9% area of WSI from the Olympus scanner. This is especially problematic for our application because almost 10% of the Olympus scanner image was out-of-focus, which is a data loss undesirable for our purpose.

As a sidenote, the defocus output from the model can be associated with the physical defocus of our system. In [153], it was concluded that the trained model can generalize on previously unseen fluorescence images, identifying the absolute image focus to within one defocus level (approximately 3-pixel blur diameter difference) with 95% accuracy. According to our system parameters and considering the size of mCTCs is larger than 8 μm, we established that our 'in-focus' class corresponds to target objects within ±2.5 μm of the focus, the 'moderately-in-focus' class corresponds to target objects between ±2.5 μm and ±7 μm of the focus, and the 'out-of-focus' class corresponds to all distances beyond.

**5.5 Cell detection based on conventional computer vision method**

First, a conventional computer vision (CV) method based on pathologist screening protocol and image processing algorithms was developed to detect cells of interest.

The screening criteria to identify mCTCs and CAFs for pathologists were subjectively based on the fluorescence signal intensity and localization, cell size and shape [134]. The main goal is to distinguish mCTCs and CAFs from other cells, precipitate, and junk. A simple decision tree for the process is given in Fig. 5.7. In this project, mCTC clusters were not considered as an independent class but rather as a set of clumped single cells. Once all single mCTCs are detected, we could evaluate their pairwise distance to identify clusters further. Then, a conventional CV method to detect mCTCs and CAFs based on this pathology screening

protocol was developed and summarized in Fig. 5.8. This method does not rely on any conventional machine learning techniques such as support vector machine.
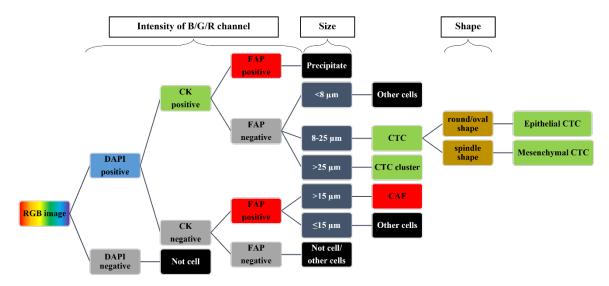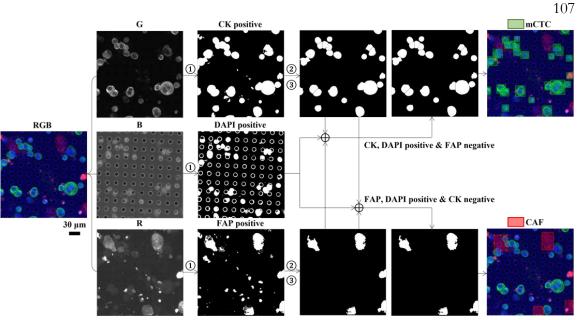


Figure 5.7: Pathology screening decision tree. CK, cytokeratin; FAP, fibroblast activation protein.

First, three channels were separately binarized to identify numerous positive events, DAPI positive (blue channel), CK (cytokeratin) positive (green channel), and FAP (fibroblast activation protein) positive (red channel). Then, CK and FAP positive events were segmented using watershed transform and cleaned by removing event regions with size below or beyond expectations. Finally, binary images from each of the three channels were cross-checked. The event regions which were DAPI and CK positive but FAP negative were accepted as CTCs and those which were DAPI and FAP positive but CK negative were identified as CAFs. Here, DAPI positive events got confirmed by calculating mean intensity value within the area of CK/FAP positive events. This approach helped reject most of the microfilter holes due to their hollow structure. During this image processing pipeline, the binarization threshold, segmentation parameters, and size threshold were all chosen and optimized by trained observer guidance.

Figure 5.8: Conventional computer vision method for mCTC and CAF detection using image processing algorithms.

## 5.6 Cell detection based on ensemble deep learning approach

Despite the capabilities of the conventional CV method to detect cells to some extent, it falls short of achieving the level of detection accuracy that we desire. Moreover, it requires human intervention to fine-tune various parameters, rendering it an incomplete automated solution. Hence, our goal is to implement a fully automated cell detection and identification system based on deep learning techniques, utilizing data acquired from our imaging system.

To obtain a large number of training instances, we prepared model sample slides using specific cell lines. Then, the WSIs from these model system slides were split into training and testing datasets at the level of individual slides. The information on the data description and splitting scheme is presented in Fig. 5.9. To reduce color variation, we manually adjusted the histogram of each color channel in every whole slide image so that cells of the same type displayed a consistent color pattern before inputting them into the deep learning model. Despite this, the background varied across different slides, with microfilter holes producing autofluorescence in the DAPI channel being particularly pronounced in slide 3. However,

the high precision values obtained from our trained deep learning model during testing shown later indicate that it was not affected by this background 'noise'.
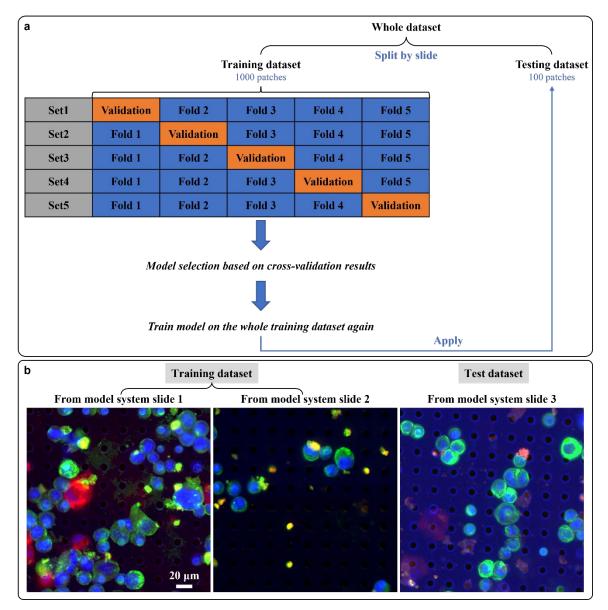


Figure 5.9: Overall plan of data splitting for deep learning. a. The dataset was initially split into a training dataset and a testing dataset by slide. Two out of the three slides from different batches were used for training, while the remaining slide was used for testing. During the deep learning model training, we conducted 5-fold cross-validation to select the best model architecture. Subsequently, the model was retrained using the entire training dataset for final testing. b. Sample images from three slides.

The scheme for cell detection via deep learning is summarized in Fig. 5.10. For the training image set, the cells of interest were manually annotated in QuPath [154] using the dot annotation function for fast screening. However, the dot format was insufficient to generate the bounding box for each cell. Therefore, the conventional CV method mentioned above was also applied. Results from both methods were cross validated by matching annotation dots and segmentation regions. Any region containing annotation dots was used to generate a bounding box and paired with the annotation label. For dots outside any region, bounding boxes that centered them were generated with the side length matching the empirical cell diameters reported in previous studies. Then, training images and their corresponding bounding boxes with class labels were used to train a generic object detection DL model. To reduce the training time and simplify the hyper-parameter tuning process, transfer learning was adopted by using weights pretrained on the COCO benchmark dataset [155]. Furthermore, to decrease model performance variation, we ensembled the predictions from five models with the same architecture that had been separately trained on different image batches. This process was executed for mCTCs and CAFs separately.
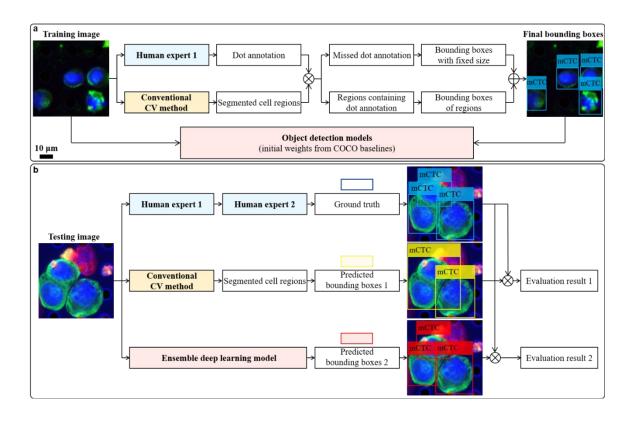
Figure 5.10: Cell detection via deep learning. a. Training pipeline. An experienced pathologist annotates the cells of interest in training images with dots and simultaneously these images are processed by a conventional computer vision (CV) method to segment cell regions. Results from both methods are cross validated by matching annotation dots and segmentation regions. Any region containing annotation dots is used to generate a bounding box and paired with the annotation label. For dots which do not lie in any region, a bounding box centered at each of them is generated with the size of empirical cell diameter. Then, training images and their corresponding bounding boxes with class labels are used to train a generic object detection deep learning model. Here, transfer learning is adopted by using weights pretrained on the COCO benchmark dataset. b. Testing pipeline. The unseen testing images are analyzed in three ways. First, the same experienced pathologist screens testing images by annotating the cells of interest with bounding boxes, which are sequentially double checked by another computational pathology researcher to make sure there is no oversight or mislabeling. This result is taken as ground truth. In parallel, testing images are segmented by the conventional CV method and then the prediction boxes with labels are generated from segmented regions. Finally, they are sent to our well-trained cell detection model and the predicted bounding boxes can be directly generated. Comparing results from the latter two methods with the ground truth, we find our trained deep learning model outperforms the conventional CV method.
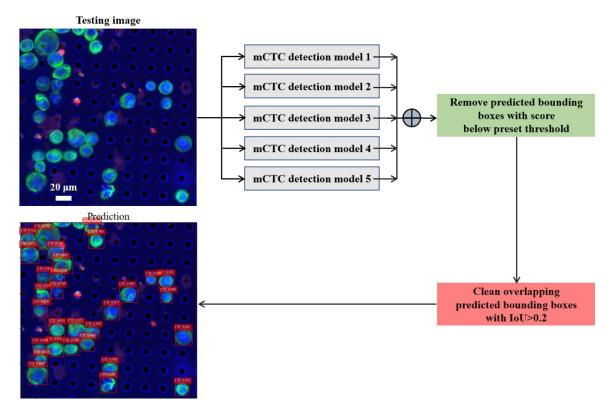


Figure 5.11: Ensemble of five object detection models and cleaning of prediction result.

The ensemble strategy is illustrated in Fig. 5.11. Five cell detection models have the same neural network architecture but are trained on randomized image batches. The ensembled prediction is a set of bounding boxes with confidence score. We first remove the predicted boxes with score lower than a chosen threshold. Then, the repeated boxes predicted for the same cell can be reduced to a single box by quantifying the pairwise overlapping ratio between boxes with intersection-over-union (IoU) metric, finding the clustered boxes and only keeping the one with the highest score.

Our choice of DL model was informed by the following process. The varieties of neural network architectures for object detection can be categorized into two groups: one-stage detectors and two-stage detectors [23, 156]. Their respective representatives are RetinaNet [157] and Faster R-CNN [24]. We paired them with different backbones, ResNet-50 [22], ResNet-101 [22] and ResNeXt-101 [158], as well as pre-trained weights on different benchmark datasets, including COCO [155] and Pascal VOC [159]. The precision-recall curves of different models for mCTC detection are summarized in Fig. 5.12 (a) using 5-fold cross validation. From that analysis, we saw that RetinaNet and Faster R-CNN with ResNet-101 backbone pretrained on COCO dataset achieved the best area under curve (AUC) in one-stage and two-stage detectors, respectively. In general, RetinaNet models worked better than Faster R-CNN models. Therefore, we finally chose the RetinaNet model for the CTC detection task. Based on this result, we only tested the RetinaNet and Faster R-CNN with ResNet-101 backbone pretrained on COCO dataset on the CAF detection task. We experimentally found that Faster R-CNN has better localization and recognition accuracy, as indicated by the higher AUC in Fig. 5.12 (b). As such, we chose Faster R-CNN for the CAF detection task.

The dataset and code were saved in Google Drive and the training session of all models was run in Google Colab. The assignment of GPU resource was random due to the subscription rules behind Colab. It was either NVIDIA A100 (SXM4, 40GB) or Tesla V100 (SXM2, 16GB). The object detection tasks were developed based on an open-source computer vision

library, Detectron2 [160], implemented in PyTorch. It provides a large set of baseline results and pre-trained models.

| | | Training dataset | | | |
|---|---|---|---|---|---|
| Set1 | Validation | Fold 2 | Fold 3 | Fold 4 | Fold 5 |
| Set2 | Fold 1 | Validation | Fold 3 | Fold 4 | Fold 5 |
| Set3 | Fold 1 | Fold 2 | Validation | Fold 4 | Fold 5 |
| Set4 | Fold 1 | Fold 2 | Fold 3 | Validation | Fold 5 |
| Set5 | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Validation |
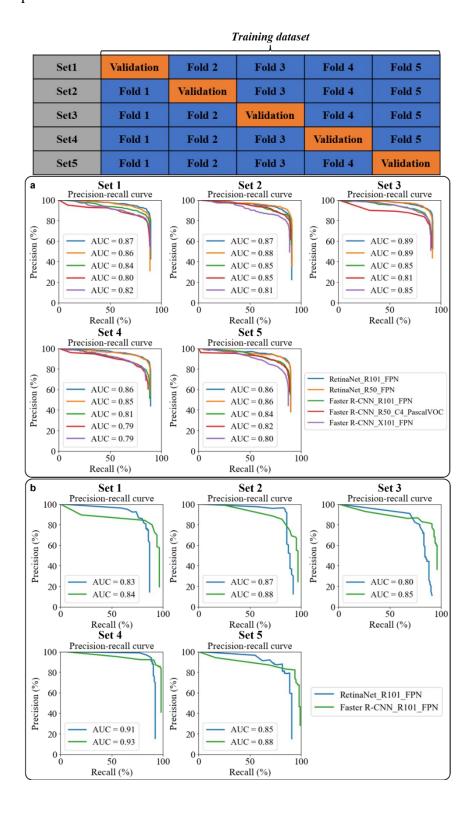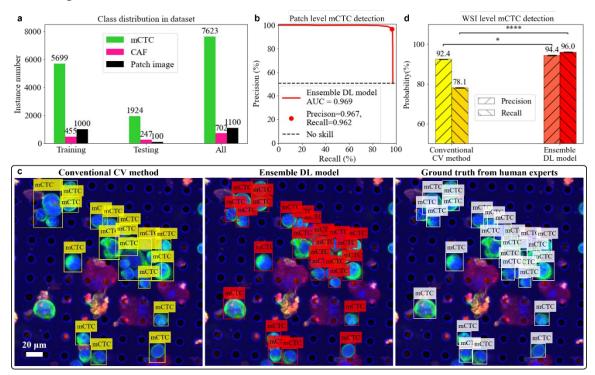
Figure 5.12: Comparison of different object detection baseline models by 5-fold cross validation. a. mCTC detection by five selected detectors differing in either network architecture, backbone or pretraining benchmark dataset. From the precision-recall curves, RetinaNet and Faster R-CNN with ResNet-101 backbone pretrained on COCO dataset achieved the best area under curve (AUC) in one-stage and two-stage detectors, respectively. b. Only the best one-stage and two-stage models were tested on the CAF detection task. The two-stage detector, Faster R-CNN, had better localization and recognition accuracy as indicated by higher AUC.

All cell detection models were trained for 1000 epochs and the learning rate was set as 0.00025. The number of output classes was always one and the class label was either mCTC or CAF. Other settings, including optimizer and loss function, were kept as default in the trainer. A training session of 1000 epochs took around 3.5 minutes to finish on NVIDIA A100 or Tesla V100. We did not observe a significant speed difference probably because our training was not computationally heavy.

## 5.7 Comparison between two methods



Figure 5.13: Evaluation of mCTC detection. a. Class distribution and the number of patches images in the training/testing/whole dataset. b. Precision-recall curve of the ensemble deep learning (DL) model to detect mCTCs in testing patch images. The red dot represents the final chosen operating point. c. Example of mCTC detection by conventional computer

vision (CV) method and ensemble DL model shown horizontally with ground truth from human annotation. d. Performance comparison between conventional CV method and ensemble DL model to detect mCTCs on the whole slide image level. Both precision and recall of ensemble DL model are significantly higher than the ones of conventional CV method. Statistical analysis uses the ensemble DL model result as the reference to test their difference significance, error bars show standard deviation of precisions and recalls by randomly sampling testing dataset 1000 times and the $p$-values are specified in the figure for *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, NS, not significant, two-sided $z$-test.

To evaluate the cell detection performance of two methods on unseen test slide samples, we designed the testing phase of the experiment as follows. First, the same experienced pathologist screened testing images by annotating the cells of interest with tight bounding boxes, which were consecutively double checked by a second reviewer to make sure there was no oversight or mislabeling. This result was taken as the ground truth. Then, testing images were segmented by the conventional CV method in the same way as training images. On a parallel track, we also used our trained ensemble DL model to directly predict bounding boxes on the testing images. Comparing the results from the two computational methods with the ground truth, we found our final ensemble DL model significantly outperformed the conventional CV method.

The output of the DL detection model for a testing image patch was a set of bounding boxes with confidence scores. As shown in Fig. 5.11, the bounding boxes with scores below the threshold were discarded. By adjusting the threshold to this score, we altered the precision and recall of our ensemble model, shown as the precision-recall curve in Fig. 5.13 (b). The AUC was 0.97 for mCTC detection. It is a good indication that our ensemble model performed well. To balance type I and type II errors, the final operating point was chosen to give precision of 97% and recall of 96%. It is worth noting that the input image size used for both training and testing in our study was 1000 x 1000 pixels, which was cropped from much larger whole slide images. In cases where mCTCs were located on the boundary of the FOV, only a portion of the cell would appear within the DL model's FOV at any given time. For instance, in some cases, only part of the CK signals would be visible without any DAPI signal. To avoid any ambiguity in our testing and ensure accurate cell counting, we only

annotated and tested cells that were completely visible within the FOV of patches and excluded those that were partially visible.

Figure 5.13 shows an example of mCTC detection by an experienced pathologist, by the ensemble DL model, and by the conventional CV method. We can see that conventional CV method missed a significant number of mCTCs for two major reasons. First, the conventional image segmentation method appeared to have difficulties segmenting clumps of cells. It mistakenly predicted several mCTCs as one. Second, mCTCs that were surrounded by precipitates with other stains appeared to be more likely to be eliminated after thresholding in other channels. Our ensemble DL model neither down-segmented the mCTCs nor was it strongly affected by precipitates attached to the mCTCs.
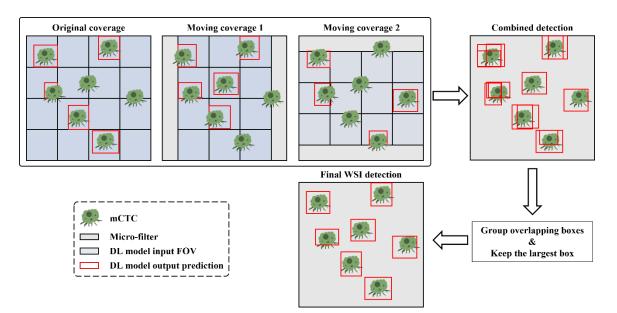


Figure 5.14: Cell detection at the whole slide image level.

When the ensemble model was directly applied to the WSI, it tended to miss the cells lying on the patch image boundary. If we were to naively apply our trained model, the precision and recall for detecting mCTCs at the WSI level would drop to 93% and 94%, respectively. To solve this problem, we proposed a whole slide image detection scheme described in Fig. 5.14. The patch coverage was shifted horizontally and vertically by the half size of DL model input FOV. Then repeated detections were removed by calculating the overlapping ratio

between each two of the predicted bounding boxes. After accounting for this, we obtained

a final precision and recall at the WSI level of 94% (±0.2%) and 96% (±0.2%), respectively.

In contrast, when we directly applied conventional CV method on the WSI, we obtained a

precision of 92% (±0.2%) and a recall of 78% (±0.3%). In general, the conventional CV

method can be seen as a conservative detector with precision close to our ensemble DL model

but is more likely to miss the mCTCs.
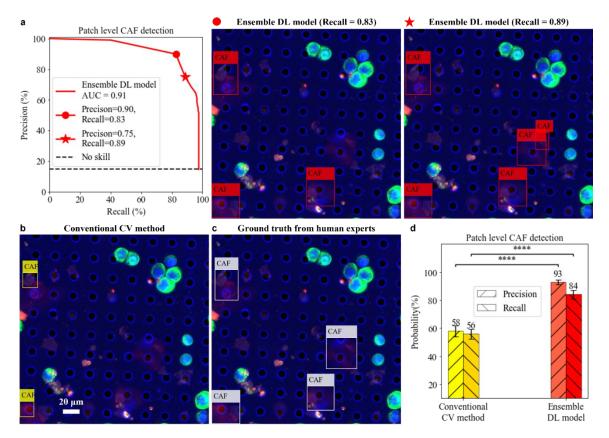


Figure 5.15: Evaluation of CAF detection. a. Precision-recall curve of the ensemble deep learning (DL) model to detect CAFs in testing patch images. The red dot represents the final chosen operating point. The red star represents another operating point with higher recall but lower precision. Any possible CAF event will be caught but it requires further human analysis to exclude the false alarms. b. CAF detection by conventional computer vision (CV) method. c. Ground truth from human expert annotation. d. Performance comparison between conventional CV method and ensemble DL model to detect CAFs on the patch image level. Both precision and recall of ensemble DL model are significantly higher than the ones of conventional CV method. Statistical analysis uses the ensemble DL model result as the reference to test their difference significance, error bars show standard deviation of precisions and recalls by randomly sampling testing dataset 1000 times and the *p*-values are

specified in the figure for $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$, NS, not significant, two-sided $z$-test.

We repeated the same training and analysis experiment on CAFs. The results are summarized in Fig. 5.15. In general, we found CAF detection to be more challenging than mCTC detection for the computational methods as there were more fluorophore precipitates in the specific fluorescence channel for CAF as well as the size and shape of CAFs vary more widely, and as some CAFs were found to be caught within the filter membrane pores and deformed more extensively than the cultured tumor cells. We also note that, within the context of our experiment, the number of CAFs was much lower than that for the mCTCs in our dataset. Under these circumstances, we achieved the final AUC of the precision-recall curve for our ensemble DL CAF detection model of value 0.91. When we seek a balanced precision and recall, this model yields a precision and recall of 90% and 83%, respectively. Following statistical analysis, which will be discussed later, we obtained a final precision and recall of 93% (±1.7%) and 84% (±3.1%), respectively, for our DL CAF detection model. In contrast, the conventional CV method gave a balanced precision and recall of 58% (±3.9%) and 56% (±3.5%), respectively.

We anticipate that by increasing the number of training instances we will see an improvement in the performance. Empirically, we saw that training instance number on the order of around 5000 resulted in an acceptably high AUC for mCTC analysis. This suggests that we should be able to boost the CAF's AUC quantity by increasing the training instance to that order of magnitude in our next phase of work. Alternately, we can still employ this existing ensemble model by pivoting its use as a screening tool to aid manual analysis. Specifically, we can choose the operating point to have a high recall (example: recall of 89% and precision of 75% shown in Fig. 5.15 (a)). This will result in a high identification rate for CAFs with a high incidence of false positives, which can then be subsequently ruled out by the human operator. While a model with slightly lower AUC would still be useful for pre-screening CAFs for skilled observers (i.e., cytopathologists) to make the final assessment, this finding strongly indicates that our clinical research phase should aim to collect more CAFs instances for DL training in order to boost the AUC.

Last but not least, cell detection results from our ensemble DL method and the conventional CV method were compared using a two-tailed $z$-test of two population proportions with significance classified for $p$-values as: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$, and NS, not significant.

To obtain statistically meaningful expectations and standard deviations of precisions and recalls for different detectors, we randomly sampled the testing dataset 1000 times and performed testing on each individual sample. The resulting mean was then reported and used for the two-tailed $z$-test, with the standard deviation being reflected as error bars in Figures 5.13 and 5.15. For patch-level testing, we conducted bootstrapping to sample the original testing patch images. For WSI-level testing, as we only had a single slide available for testing, we randomly cropped 90% of the whole slide area to obtain multiple samples.

## 5.8 Summary

Our microfilter technology provides a robust platform to capture and enrich CTCs and other rare cells of interest from the peripheral blood of cancer patients. Assessments of these events provide opportunities for advanced liquid biopsy diagnostics in a minimally invasive manner. To this end, the combined hardware and deep-learning-based analytical solution that we present here could significantly advance the practice of cell-based, liquid biopsy cancer diagnostics.

We have developed an imaging system that can adequately address the inherent imaging constraints imposed by capturing cells on micropore filters or other surfaces with focus plane variations that are ill-addressed by currently available commercial slide scanners. We showed that our hardware-software hybrid system can locate the correct focal plane within the objective's FOV at a patch size of 138 μm x 138 μm and it can render AIF whole slide images for our filter with only about 0.7% of the total area missing the target focus. We next showed that this uniform focus quality of the rendering is well suited for DL based object recognition for automated detection of mCTCs and CAFs on the filter. We further demonstrated that the DL approach outperforms conventional CV models in the recognition task. The positive findings form the foundation to begin our next phase of research: collecting

clinical CTCs and CAFs samples to train the DL models and apply the trained system for routine automatic identification of CTCs and CAFs in clinical blood samples.

Finally, we conclude by noting that a fully automatic CTC and CAF detection system based on DL techniques is useful for freeing human labor from the tedious cell identification and eliminating human subjectivity from the process. We anticipate that such a system can potentially find applications in clinical research and ultimately in the use of clinical cancer management.

*Chapter 6*

# STAIN-FREE DETECTION OF EMBRYO POLARIZATION USING DEEP LEARNING

In the previous chapter, we explored how deep learning methods can replace human labor. In this chapter, we will shift our focus to utilizing deep learning techniques to tackle tasks that humans struggle to handle effectively.

Polarization of the mammalian embryo at the right developmental time is critical for its development to term and would be valuable in assessing the potential of human embryos. However, tracking polarization requires invasive fluorescence staining, impermissible in the in vitro fertilization clinic. In this chapter, we report the use of artificial intelligence to detect polarization from unstained time-lapse movies of mouse embryos. We assembled a dataset of bright-field movie frames from 8-cell-stage embryos, side-by-side with corresponding images of fluorescent markers of cell polarization. We then used an ensemble learning model to detect whether any bright-field frame showed an embryo before or after onset of polarization. Our resulting model has an accuracy of 85% for detecting polarization, significantly outperforming human volunteers trained on the same data (61% accuracy). We discovered that our self-learning model focuses upon the angle between cells as one known cue for compaction, which precedes polarization, but it outperforms the use of this cue alone. By compressing three-dimensional time-lapsed image data into two-dimensions, we are able to reduce data to an easily manageable size for deep learning processing. In conclusion, we describe a method for detecting a key developmental feature of embryo development that avoids clinically impermissible fluorescence staining.

**6.1 Introduction**

Mammalian embryo polarization is the process by which all individual cells of the embryo establish an apical domain on the cell-cell contact-free surface. In the mouse embryo, this process occurs at the late 8-cell stage, on the third day of development after fertilization as shown in Fig. 6.1 (a), and in humans on the fourth day at the 8-16 cell stage [161–167]. The apical domain is composed of the PAR complex and ERM proteins (Ezrin, Radixin, Moesin), enclosed by an actomyosin ring [162, 168–172]. The cells which inherit this apical domain after division will become specified as trophectoderm (TE), which ultimately forms the placenta. In contrast, those cells that do not inherit the apical domain will form the inner cell mass (ICM), which will give rise to all fetal tissues and the yolk sac [164–166]. Thus, embryo polarization provides the first critical bifurcation of cell fates in the mammalian embryo, and establishment of cell lineages in the blastocyst, which is crucial for implantation and a successful pregnancy. In agreement with this, preventing cell polarization of the mouse and human embryo, prevents its successful development [163, 164, 171, 173].

Given the importance of polarization, an ability to detect this developmental feature non-invasively would be beneficial, for example, for the screening of viable human embryos for implantation. However, all current methods for detecting polarization are invasive as they rely on modifying embryos to express fluorescently tagged proteins that mark the apical domains [174, 175]. Such fluorescent tagging of human embryos meant for implantation is impermissible, which currently prevents clinical embryologists from utilizing polarization to evaluate the quality of human embryos for transfer to mothers-to-be.

Tracking polarization without the use of fluorescence could be solved using deep learning, which is able to discern salient features that may be unintuitive for humans [176, 177]. Indeed, deep learning has been recently used successfully to automate detection of an embryo's morphological features and applied on single time-point images to assess implantation potential of human embryos [176–181]. These deep learning approaches either provide a means of accurately counting cell numbers with developmental time [176, 177], relate embryo morphology to subjective developmental criteria assessed by embryologists

[178, 179], or provide independent assessment of morphological features [179, 180]. One study has related preimplantation morphology with successful development of fetal heartbeat [181]. However, the morphological features being assessed by the deep learning algorithms used to date are generally not clear. In addition, these current approaches do not rely upon known critical developmental milestones in the molecular cell biology of preimplantation development. In this chapter, we have used time lapse movies of fluorescent markers of polarization in the developing mouse embryo to train a deep learning system to recognize the polarization events in the corresponding bright field movie frames with a high degree of success. This is the first time that deep learning has been applied to recognize a specific molecular cell biological process in an embryo that is key for developmental success.
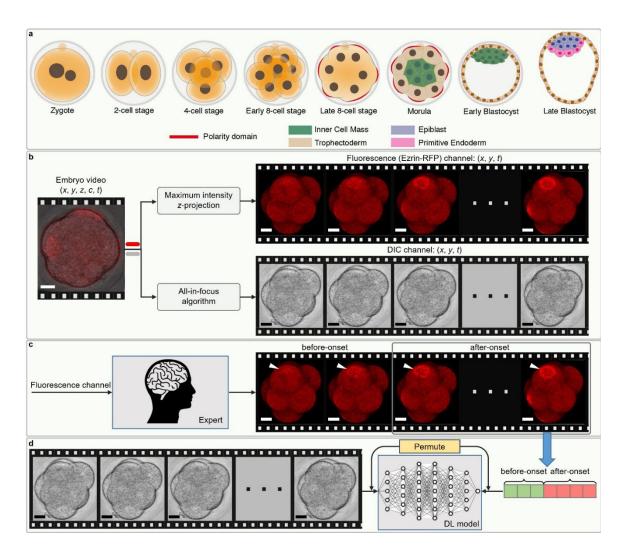
Figure 6.1: Method to track and annotate polarity. a. Overview of mouse pre-implantation development, from the zygote stage at embryonic day 0 to the late blastocyst stage at embryonic day 4.5. At the late 8-cell stage, polarization takes place, as each blastomere gains a defined apical-basal axis of polarity indicated by the presence of an apical domain (red). b. Data preprocessing of dual-channel 3D mouse embryo videos, each of which is a 5D tensor with the dimension of $x$, $y$, $z$, $c$ (channel), and $t$ (time). First, each video was split into a fluorescence (Ezrin-RFP) and DIC channel, visualized in red and gray respectively. Then, each channel was compressed along the $z$ dimension by different algorithms. The maximum intensity $z$-projection algorithm was applied for the fluorescence channel and DTCWT based all-in-focus algorithm for the DIC channel to get the frame sequences. c. Expert annotation on fluorescence frame sequences, where the time point of polarity onset is pinpointed. In the time sequence, the onset of polarization was defined as the frame in which the blastomere had a clear polarity ring or cap (closed) which took up at least 1/3 of the visible surface, or 1/3 of the cell surface curve if displayed side-on. Frames before this point were defined as before-onset, whilst frames including and after this point are defined as after-onset. d. Supervised learning of a single DCNN model. The DIC frame sequences paired with the class labels from fluorescence annotation were permuted and used as the input and target of the supervised learning. Transfer learning from pre-trained weights on ImageNet database and data augmentation are utilized in the training of all DCNN models. Scale bar = 30 $\mu$m.

## 6.2 Data collection, preprocessing and annotation

In order to develop our deep learning model for detecting the polarization status of live and unstained embryos, we first required a large dataset of DIC embryo frames for which the polarization is unambiguously evident. Generating this dataset required each DIC image to have a corresponding channel that uses fluorescently tagged proteins to indicate polarization for each embryo clearly. We built a large dataset of mouse embryo time-lapse recordings during polarization, each containing a fluorescence channel to indicate embryo polarization, and a synchronized DIC channel for model training and testing (Fig. 6.1 (b)).

Two channels were first compressed along the $z$ axis using different algorithms. The fluorescence $z$-stack images were converted into maximum intensity $z$ projection frames for annotation, and the DIC $z$-stack images were converted into all-in-focus frames for deep learning input. The polarization of a single blastomere in the embryo can be determined by the localization of apical proteins, which are enclosed by an actomyosin ring [169–171, 182]. For time-lapse recordings, embryos were injected at the 2-cell stage with synthetic mRNA for Ezrin tagged with red fluorescence protein (RFP), as previously [183], and cultured *in vitro* to the 16-cell stage. We used Ezrin as a marker for blastomere polarization, as Ezrin

localizes to the apical surface during the formation of an apical polarity domain [172, 184]. Based on the Ezrin-RFP ring indication, we determined the time point at which the first blastomere of the embryo polarized for each time-lapse recording, indicated by formation of a clear apical polarity cap (Fig. 6.1 (c)). Using this time point, each DIC frame was labelled as either before or after the onset of polarization (Fig. 6.1 (d)). The annotated DIC frames were then randomly split into a training and testing cohort by embryos.
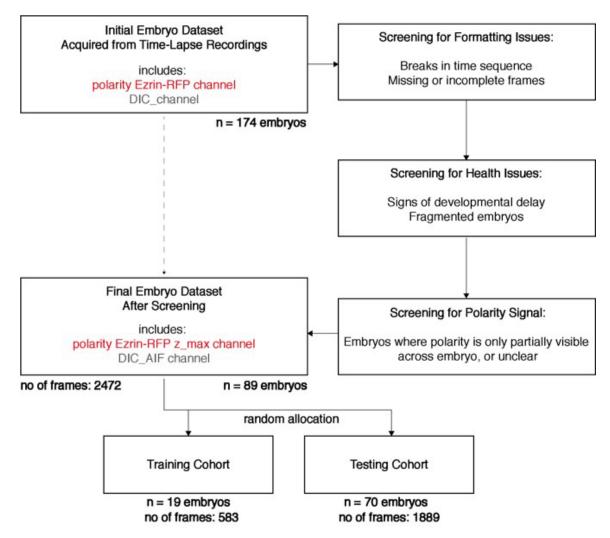


Figure 6.2: Flowchart indicating initial cleaning and pre-processing of data.

All mouse experimental data was obtained in accordance with the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012, under project license by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB). Reporting of animal

experiments follows ARRIVE guidelines. Embryos were collected at the 2-cell stage from F1 females (C57BI6xCBA) mated with F1 studs, following super ovulation of the female: injection of 7.5 IU of pregnant mares' serum gonadotropin (PMSG; Intervet), followed by injection of 7.5 IU of human chorionic gonadotropin (HCG; Intervet) after 48 hours and immediately prior to mating.

Images were collected on confocal Leica SP5 or SP8 microscopes. The dual-modal confocal microscope imaging allows synchronized acquisition of a DIC channel and a fluorescent polarity-indicating channel. The interval between each frame on the time ($t$) axis was 1200 s - 2400 s for each embryo, and $z$ frames were taken at 4 μm intervals on the $z$ axis. Time-lapse recordings were converted into TIFF files for analysis and processed on Fiji software. Recordings that were incorrectly formatted, visually unclear, or which showed grossly defective embryos were excluded. From an initial 174 mouse embryo time-lapse recordings from, 89 were used for deep learning and human testing (Fig. 6.2). Only 8-cell stage frames were included in deep learning and analysis (defined as frames from the first frame where 8 distinct blastomeres are visible, to the frame immediately prior to the moment at which the final blastomere starts dividing).

To obtain comprehensive three-dimensional (3D) information about the embryo, raw data were required in $z$-stack frames. However, current deep learning models are predominantly designed for two-dimensional (2D) image inputs. Therefore, in prior research, the model input was usually a single slice image along the $z$-axis [176–181]. A single $z$-slice image is inadequate in capturing the complete 3D structural information of the embryo. Direct analysis of a 3D image stack with deep learning requires a re-designed model architecture that dramatically increases the complexity and time required for model development [185, 186]. Moreover, adapting existing pre-trained deep learning networks for 3D analysis through transfer learning [187] would not be straightforward as these networks are predominantly designed for 2D image recognition tasks. To resolve this problem, we utilized a state-of-the-art all-in-focus (AIF) algorithm based on dual-tree complex wavelet transform (DTCWT) [188] to compress the optically sectioned $z$ stack of each DIC frame in our dataset.

The result was a single 2D AIF DIC image capturing the majority of relevant but sparsely distributed 3D embryo information at each time point.

We found that AIF images based on DTCWT could reveal all blastomeres of a 3D embryo in a single 2D image (Fig. 6.3). In contrast, the median $z$ slice typically contained several blastomeres that were optically out of focus, resulting in lost information. AIF images also resembled standard images, allowing for straightforward transfer learning using open-source 2D image classification models pre-trained on ImageNet [20] as initialization. In this way, we were able to circumvent 3D image stack analysis. Prior studies that apply deep learning to embryo development have used single $z$ slice DIC images as input, which obscures important 3D blastomere features. Our DTCWT-based method of compressing multiple $z$ slices into a single maximally informative 2D representation reduces data size, allowing a full range of published 2D neural network models to become accessible for analysis.
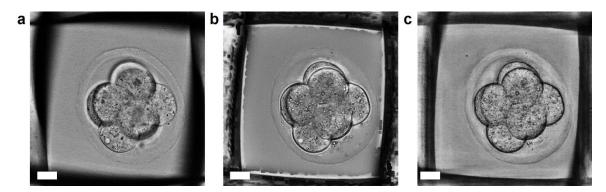


Figure 6.3: Comparison among different input image formats. Previous deep learning studies on embryo development used a single $z$ slice image, in most cases the middle plane, as seen in (a). However, this resulted in some blastomeres being highly defocused and blurred. The traditional all-in-focus algorithm based on variance metric (b) can bring all the blastomeres into focus in a single 2D image but also result in some artifacts. Thus, we proposed to utilize the all-in-focus algorithm based on dual tree complex wavelet transform (c). Scale bar = 20 $\mu$m.

Each embryo time-lapse recording was marked with a polarization onset time by a human expert annotator, corresponding to the first frame in which a polarized blastomere is clearly visible. This was achieved using the maximum intensity $z$ projection Ezrin-RFP frame: the polarization onset frame is one in which the first apical ring or cap is completely and clearly formed (closed) on any blastomere, and which takes up greater than or equal to 1/3 of the

surface of the cell as visible in the recording. All frames after and including this polarization onset point were classified as after-onset. All frames prior to this point were classified as before-onset.
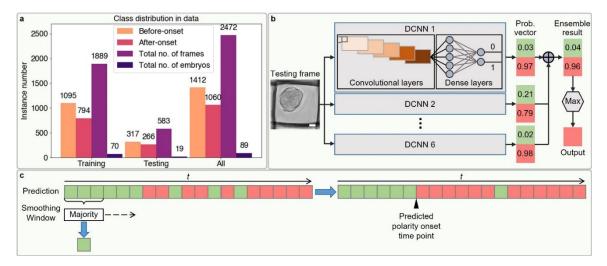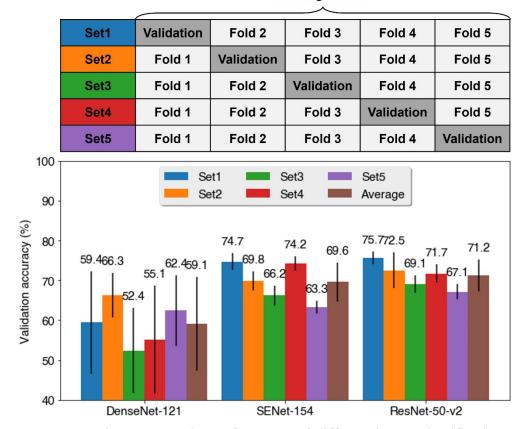


Figure 6.4: An ensemble deep learning approach to predict embryo polarization from DIC images. a. Class distribution in the training/testing/whole dataset. b. Ensemble learning on six DCNN models. The predicted probability vectors for two classes on a single testing frame by six DCNN models were averaged element-wisely and the class corresponding to the larger probability was used as the final predicted label. c. Temporal smoothing on the predicted labels for each testing embryo's DIC frame sequence. The majority voting based smoothing window slid over the chronologically ordered binary labels. The window length is 3 and we kept the label at both ends untouched. Finally, the time index of first after-onset prediction was taken as the final prediction of polarity onset time point. Scale bar = 20 $\mu$m.

## 6.3 Polarization prediction based on deep learning approach

The dataset consisting of AIF DIC images paired with corresponding annotated polarization labels was randomly split into a training cohort of 70 embryos (1889 frames) and a testing cohort of 19 embryos (583 frames) (Fig. 6.4 (a)). These were used as learning and evaluation datasets, respectively. To fully leverage the potential of deep learning techniques for predicting polarity onset, we combined two highly effective methods, deep convolutional neural network (DCNN) and ensemble learning. For supervised learning of DCNN models, we retained only information about whether a frame was before or after onset and stripped away other time information (Fig. 6.4 (b)). On individual testing frames, each DCNN model outputs whether polarization was detected as a vector containing two probabilities—one for

each class (before or after onset, Fig. 6.4 (b)). To reduce the prediction variation, multiple (6 here to match the number of recruited human volunteers discussed later for fair comparison) DCNNs learnt on the training cohort and then their output predictions were averaged to predict the class label of each testing image. The six DCNN models were trained using different initializations and different optimizers but over the same number of epochs. The final polarization status prediction for a single input image is the class (before or after onset) having the highest average probability across all six contributing models. Overall, our model accuracy increased from an average of 82.6% for a single DCNN to 85.2% with ensemble learning.

**Cross validation for model selection**

*Training dataset*

| | | | | | |
|---|---|---|---|---|---|
| Set1 | Validation | Fold 2 | Fold 3 | Fold 4 | Fold 5 |
| Set2 | Fold 1 | Validation | Fold 3 | Fold 4 | Fold 5 |
| Set3 | Fold 1 | Fold 2 | Validation | Fold 4 | Fold 5 |
| Set4 | Fold 1 | Fold 2 | Fold 3 | Validation | Fold 5 |
| Set5 | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Validation |



Figure 6.5: Comparison among the performance of different image classification DCNN backbones. Five-fold cross validation was adopted here to make model selection. The training dataset was evenly split into 5 folds. Then, three backbone models learnt on four folds and were validated on the remaining one. To minimize the variance brought by the optimization setting, we repeated each numerical experiment 5 times. We summarized the

validation accuracy of three models on each fold and the total average validation accuracy in the bar chart, where the error bars represent their standard deviation. Both validation accuracy and their standard deviation shows that ResNet-50-v2 is the optimal backbone for our task.

We considered several possibilities during the design of our model architecture. Many DCNN architectures have been published for image classification, including dense convolutional network (DenseNet) [189], squeeze-and-excitation network (SENet) [190], and residual neural network (ResNet) [22]. Based on the results of 5-fold cross validation (CV) experiments on these three model architectures (Fig. 6.5), we found ResNet has the highest validation accuracy with the lowest variance on average. Considering both prediction performance and computational load, ResNet model is the optimal choice for our application.

Table 6.1: Result summary of different ensemble techniques. 95% confidence intervals are included in square brackets. They are estimated by bootstrapping the testing dataset with 1000 replicates. The highest testing accuracy in each category is marked in bold.

| Varied element | Different techniques | Testing accuracy |
|---|---|---|
| Training data | raw training data | **85.2%** [82.2%, 88.2%] |
| | 6-fold cross-validation on training data | 83.4% [80.3%, 86.4%] |
| | bootstrapping training data 6 times | 82.2% [78.7%, 85.4%] |
| Model | varying optimization initials | **85.2%** [82.2%, 88.2%] |
| | 3 SENet-154 and 3 ResNet-50-v2 models | 83.4% [80.1%, 86.4%] |
| Combination method | averaging the output probability from ensemble members | 85.2% [82.2%, 88.2%] |
| | applying majority voting on the output labels from ensemble members* | **85.6%** [82.5%, 88.3%] |

\* If we meet a tie, the averaged probability will be used to make a judgement instead.

Specifically, the ResNet backbone was chosen as the main part of each DCNN model. A dense layer with two output nodes is added on top of the ResNet backbone. We used the pre-trained weights on ImageNet database as the initialization for each DCNN model. Three of them were trained with SGD optimizer and the other three were trained with Adam optimizer. All of them were trained for 40 epochs. At the end of 40 epochs, all the models converge to nearly 100% in terms of the training accuracy. Different training settings made the six trained CNNs a bit more diverse from each other, where the diversity among CNNs would improve the generalization ability of the ensemble model. Then, we tried different ensemble

techniques, and their results are summarized in Table 6.1. From this table, we can see that ensemble on six ResNet models using all the training data by varying the optimization initials and then applying majority voting on the output labels achieved the best testing accuracy. However, its advantage over averaging the output probability is only 0.4%. It is quite marginal, and the latter is more intuitive and common considering the number of ensemble members is even. We further note that there is a limit on ensemble learning improvement over a single classifier, which is mainly bottlenecked by the classifier architecture rather than ensemble techniques.

## 6.4 Comparison with human prediction and temporal analysis

In order to evaluate the performance of our DL model, comparative trials on human volunteers to identify polarity onset were conducted. We recruited six volunteers to compare their polarization prediction accuracy against our model. We aimed to recruit human volunteers from a STEM background, who would be motivated to benefit from the technology in a clinical setting and who might compare favorably with our machine learning system. Six human volunteers (three males and three females for gender equality) with a bachelor's degree in a STEM subject but without prior experience of mouse embryo development studies were recruited from the Caltech community, as representatives for competent STEM-trained but inexperienced volunteers who would benefit from the technology in a clinical setting.

Volunteers were sent an email with clear instructions and a link to the same training and testing data used by our model. Each was asked to learn on the training dataset first and then apply their learnt patterns to the testing images, to determine the polarization status for each test image (before or after onset) by filling in an Excel table with predicted labels. After the test, they each returned their Excel file for evaluation. All participants provided informed consent before taking part in our study. They consented to allow their data to be used in the final analysis and all individuals received reward for participation. The study was approved by the Caltech Institutional Review Board.
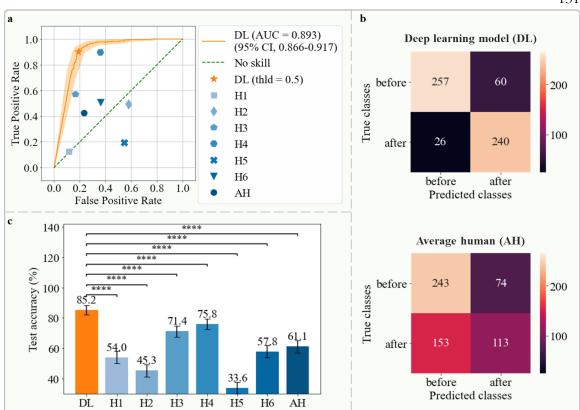
Figure 6.6: Results of image classification task by the ensemble deep learning model and the average human. a. The receiver operating characteristic (ROC) curve of the ensemble deep learning (DL) model on testing frames. The 95% confidence intervals (CIs) of the ROC curve are indicated by the orange shaded area. The orange solid star represents the performance of the ensemble DL model with the default probability threshold of 0.5 to binarize its output and blue signs represent the performance of 6 human volunteers with the darkest one the average human (AH), which is an aggregate result of 6 human volunteers' prediction. We applied majority voting to the six predictions on each testing frame to obtain the average human performance. If each prediction received three votes, we randomly assigned a prediction of before or after onset. b. Confusion matrix of image classification on testing frames by the ensemble DL model with the binarization threshold of 0.5 and the average human. c. Testing accuracy bar chart of the ensemble DL model, 6 human volunteers and the average human, where the error bars represent the 95% CI. The ensemble DL model significantly outperforms all 6 humans and the average human. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$, NS, not significant, two-sided $z$-test. All the 95% CIs are estimated by bootstrapping the testing dataset with 1000 replicates.

Results from the testing data for each of the model and human predictions were processed as follows: in classification analysis, classified frames from the model/prediction were automatically sorted into one of four categories visible in the confusion matrix (polarized or

non-polarized annotated true class, versus polarized or non-polarized predicted class).
Cases in which the true class matched the predicted class were scored as an accurate
prediction, and cases where the two classes did not match were scored as an inaccurate
prediction. Population proportions of accurate results represent the proportion of accurate
frames in the total population of frames.

Table 6.2: The performance of the ensemble deep learning (DL) model, six human
volunteers (H1-H6) and their average (AH) on the testing dataset. 95% confidence intervals
are included in brackets. They are estimated by bootstrapping the testing dataset with 1000
replicates. The highest value of each metric achieved by all predictors is marked in bold
red. PPV, positive predictive value; NPV negative predictive value. H1-H3 are females and
H4-H6 are males.

| Metrics | DL | H1 | H2 | H3 | H4 | H5 | H6 | AH |
|---|---|---|---|---|---|---|---|---|
| Sensitivity (%) | **90.2** (86.1, 93.8) | 12.8 (8.8, 16.8) | 49.2 (43.3, 55.0) | 57.1 (51.0, 63.1) | 89.8 (85.9, 93.4) | 19.5 (14.6, 24.3) | 50.8 (44.5, 56.7) | 42.5 (36.4, 48.5) |
| Specificity (%) | 81.1 (76.2, 85.4) | **88.6** (85.0, 92.2) | 42.0 (36.8, 47.3) | 83.3 (79.0, 87.2) | 64.0 (59.0, 69.3) | 45.4 (40.3, 50.8) | 63.7 (58.4, 69.2) | 76.7 (71.9, 81.1) |
| PPV (%) | **80.0** (75.1, 84.5) | 48.6 (36.2, 60.9) | 41.6 (35.7, 46.9) | 74.1 (68.1, 79.9) | 67.7 (63.0, 72.6) | 23.1 (17.7, 28.9) | 54.0 (48.2, 60.1) | 60.4 (53.6, 67.2) |
| NPV (%) | **90.8** (87.2, 94.2) | 54.8 (50.4, 58.9) | 49.6 (43.7, 55.6) | 69.8 (65.4, 74.5) | 88.3 (83.8, 92.5) | 40.2 (35.1, 45.4) | 60.7 (55.4, 66.1) | 61.4 (56.3, 66.2) |
| Accuracy (%) | **85.2** (82.2, 88.2) | 54.0 (49.9, 58.0) | 45.3 (41.0, 49.2) | 71.4 (67.4, 75.0) | 75.8 (72.4, 79.2) | 33.6 (30.0, 37.6) | 57.8 (54.0, 62.1) | 61.1 (57.1, 65.0) |

The DL model we establish yielded a classification sensitivity of 90.2% (95% confidence
interval (CI): 86.1% - 93.8%) and specificity of 81.1% (95% CI: 76.2% - 85.4%) for single
image inputs, with areas under the receiver operating characteristic curve of 0.893 (95% CI:
0.866 - 0.917) (Fig. 6.6 (a) and Table 6.2). Deep learning achieved both a higher true positive
rate and lower false positive rate than the average human volunteer. Figure 6.6 (b) shows the
confusion matrix for predictions. Our model correctly classified 497 out of 583 frames,
resulting in a classification accuracy of 85.2% (95% CI: 82.2% - 88.2%). In comparison, the
average human accuracy on the same testing frames was 61.1% (95% CI: 57.1% - 65.0%).
The model outperformed humans on average as well as individually (Fig. 6.6 (c), two-tailed
$z$-test, $p<0.0001$).

We wished to further extend our deep learning model to identify the exact point at which
polarization occurs in time-sequence videos. To this end, we evaluated polarization onset
time point predictions from the classification results of both the model and human volunteers,

using a temporal smoothing method (Fig. 6.4 (c)). For time-smoothened data, the frames were first returned to time-order, after which the polarity onset point was determined by finding the point at which the prediction switched from an unpolarized majority to a polarized majority. All frames after this polarity onset point were then classified as polarized, and all frames before this point were classified as unpolarized, therefore 'smoothening' out any anomalous predictions using time point information. For time point analysis, the polarity onset point (as determined from the smoothening process) was used. For each testing embryo time-lapse recording, the time discrepancy for the model/volunteer was calculated as the actual time difference (to the nearest second) between the predicted polarity onset frame and the annotated polarity onset frame, using the knowledge of the frame-to-frame time difference for each recording. Where no predicted onset frame was given within the allocated recording, for this analysis the frame immediately after the final frame of the time-lapse recording was used as the predicted onset of polarization. These time discrepancies for each embryo were used in pairwise comparisons. Our model had significantly smaller timestamp prediction errors than the average human volunteer by pairwise comparison (two-tailed Wilcoxon matched-pairs signed-rank test, $p<0.01$, Fig. 6.7 (a)).

We next wished to investigate whether smoothened results from our ensemble classification model could outperform even human volunteers who are given access to temporal information during testing that the model does not use. To this end, we provided each volunteer with the complete AIF DIC videos in frame-by-frame time sequence for each embryo and asked for their estimate of the polarization onset time point. Compared with the smoothened model classification results performed on individual unordered images, the average human timestamp discrepancy was significantly larger than that of our model (two-tailed Wilcoxon signed-rank test, $p<0.05$, Fig. 6.7 (b)). The model identified exact polarization time points more precisely than the human volunteers, even when the volunteers utilized temporally ordered full video frames that the model did not have access to during training. It means that our deep learning model was able to identify the exact time point of polarization onset amongst temporally sequenced video frames better than all human volunteers, even with a severe disadvantage in data.
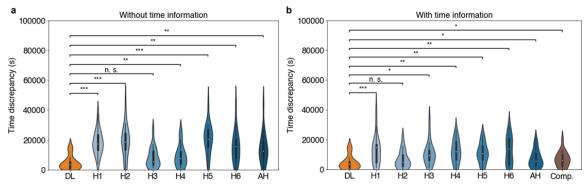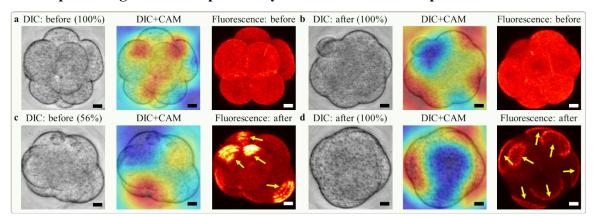
Figure 6.7: Comparative analysis of the polarity onset time point prediction by the ensemble deep learning model, six human volunteers, their average, and the compaction proxy. The violin plots represent time discrepancies between the annotated and the predicted polarity onset time index of the 19 testing embryos by the ensemble DL model, each of six human volunteers (H1-H6), their average (AH), and the compaction proxy (Comp.). (a) Six human volunteers were given the randomized testing frames without any time information. Their predicted labels were then chronologically ordered for each testing embryo and temporally smoothened to extract the polarity onset time point prediction. Their average result was processed in the same way. (b) Six humans were given the chronologically ordered frames for each testing embryo. They directly estimated the polarity onset time point. Their average result was the arithmetic mean of predicted time indexes for each testing embryo. Comparison between the ensemble DL model and each human is given in the figure. Statistical analysis uses the ensemble DL model result as the reference to test their difference significance and the *p*-values are specified in the figure for *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, NS, not significant, two-sided Wilcoxon matched-pairs signed-rank test.

Here we want to clarify that all image classification results were compared using a two-tailed *z*-test of two population proportions with significance classified for *p*-values as: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, and not significant (NS). Time prediction discrepancies were compared using two-sided Wilcoxon matched-pairs signed-rank test since our testing data size is small and not guaranteed as normal. Significance was given for *p*-values as the same with the above. Further details are given with each result. Statistical analyses were performed using the statistics module in SciPy package with Python. All the 95% confidence intervals were estimated by bootstrapping the testing dataset with 1000 replicates.

## 6.5 Deep learning model interpretability and relation with compaction



Figure 6.8: Visualization of the decision-making by the ensemble deep learning model. Heat maps obtained by the class activation mapping (CAM) technique highlight how the ensemble deep learning model attends the discriminative regions in the testing frame when giving the predicted class label. The red regions indicate positive focus of the model (in alignment with the predicted label) and the blue regions negative focus (in opposition to the predicted label). (a)-(d) correspond to four cases in confusion matrix, true negatives (TN), false positives (FP), false negative (FN), and true positives (TP), respectively. Yellow arrows point to the polarized blastomeres. In each subfigure, from left to right are the testing DIC image, its overlay with the focus heat map, and its corresponding fluorescence channel image. On top of the test DIC image is the predicted label of the ensemble DL model with its confidence (from 0 to 100%). On top of the fluorescence image is the annotated label by the expert. All the heat maps show that our DL model either attends to the individual blastomeres or the inter-blastomere angles. For example, TP heat map (d) focuses on the truly polarized blastomeres. At a certain time-point, some blastomeres have started polarization but the others have not, as shown in the FN case (c). This issue resulted in the DL model making a Type II error with low confidence in the case given. Scale bar = 20 $\mu$m.

To identify focus areas of our ensemble model, we generated attention heat maps using the class activation map (CAM) [191] technique. CAM is a heat map representation to highlight pixels that trigger a model to associate an image with a particular class (before or after onset). To be specific, we multiplied each feature map passing through the global average pooling (GAP) layer of ResNet backbone with their corresponding weight connecting the GAP layer and the fully-connected layer. Then we added the weighted feature maps in an element-wise manner. Each weight tells us how much importance needs to be given to individual feature maps. The final weighted sum gives us a heat map of a particular class (in our case, the before/after polarity onset class), which indicates what pixels our model favors or dislikes to make the final prediction. The heat map size is the same as that of the feature maps.

Therefore, to impose it on the input AIF DIC image, we scaled it to the size of the input image and finally got results shown in Fig. 6.8.

In Fig. 6.8, we have overlaid the CAM heat map with the input testing AIF DIC image. In each heat map, red pixels indicate regions of the embryo containing features that correlate positively with the predicted polarization class, while blue pixels indicate regions containing features that correlate negatively (i.e., correlate positively with the opposing class). To understand which regions of the input testing image influence our model most, we evaluated each possible prediction outcome: true negative (TN) (Fig. 6.8 (a)), false positive (FP) (Fig. 6.8 (b)), false negative (FN) (Fig. 6.8 (c)), and true positive (TP) (Fig. 6.8 (d)). When the model classified image frames as after polarization, it appeared to use inter-blastomere angle as a cue. Misclassifications tended to result from mismatched polarity between individual blastomeres and the overall embryo, producing weak prediction probabilities for both classes near 50% while the model was forced to choose one class (Fig. 6.8 (c)). Predictions in this probability range are more reasonably interpreted as not sure or cannot tell, but these were not options for the model.

The use of inter-blastomere angle as a cue by our model to determine embryo polarization (Fig. 6.8 (a)) was not surprising. Inter-blastomere angle is an indicator of embryo compaction [192, 193], a morphological change during development that typically precedes polarization (Fig. 6.9 (a)). To assess the extent to which our deep learning model uses just compaction for its polarization prediction, we annotated each embryo's AIF DIC frame sequence with the time point of compaction. We defined the time of compaction as the first frame at which the smallest inter-blastomere angle of the embryo is over 120 degrees, in agreement with previous research (Fig. 6.10) [193]. All frames after and including this point were considered compacted, and all frames prior to this point were considered uncompacted. To find the model's predicted time point of polarization, we re-aligned embryo frames in their original time sequence and applied temporal smoothing on the predicted label sequence for each testing embryo based on majority voting to output a single time point for polarization (Fig. 6.4 (c)).
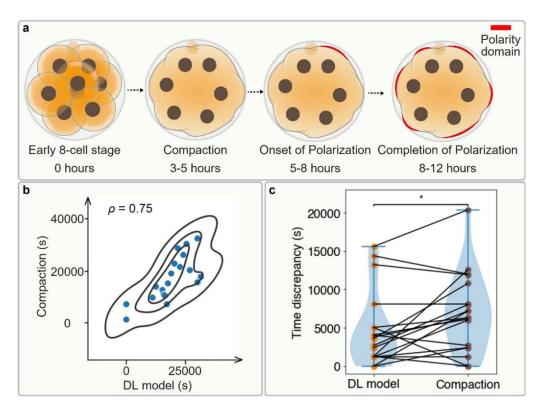
Figure 6.9: Comparative analysis of the ensemble deep learning model prediction and the compaction-based prediction for polarization. (a) Chronological order of compaction and polarization events during the 8-cell stage for a normal mouse embryo. (b) Correlation analysis between time points of DL model polarity prediction and compaction. The *x* and *y* coordinate are the predicted polarity onset time index of testing embryos (marked in blue solid balls) by the ensemble DL model and the annotated compaction time index, respectively. Their pairwise relationship shows a Pearson correlation coefficient ($\rho$) of 0.75. (c) Violin plot to visualize the time discrepancy between the annotated and the predicted polarity onset time index on 19 testing embryos by ensemble DL model and compaction proxy, overlaid with a slopegraph showing each testing embryo prediction time discrepancy in pair. From the kernel density estimate (blue shade) of violin plot and the connection line trends of slopegraph, we can tell that the prediction time discrepancy of DL model is significantly lower than the one of compaction proxy. The *p*-value is specified in the figure for *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, NS, not significant, two-sided Wilcoxon matched-pairs signed-rank test.

The Pearson correlation coefficient between compaction time point and the model's predicted time point of polarization onset was 0.75 across the 19 embryos used for testing (Fig. 6.9 (b)), suggesting that whilst compaction is indeed a utilized cue, it is not the only factor used by the model. We evaluated whether our model was superior to using compaction alone as a proxy for polarization, by calculating the time discrepancies between annotated

polarization time indexes (ground truth) and predicted time indexes by either our model or the compaction proxy. The model had significantly smaller time point prediction errors compared to the latter (two-tailed Wilcoxon matched-pairs signed-rank test, $p<0.05$, Fig. 6.9 (c)). That is, the model was superior to the use of compaction alone for predicting polarization and has likely managed to learn additional cues we do not yet understand.
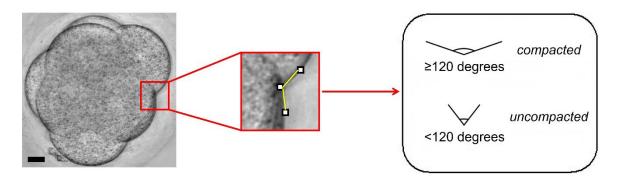


Figure 6.10: Criteria for annotating compaction. We measured compaction using the inter-blastomere angle. The first time point at which the minimal inter-blastomere angle was $\geqslant$ 120 degrees was defined as the compaction point. All frames including and succeeding this point were defined as compacted, whilst all frames prior to this point were defined as uncompacted. Scale bar = 10 $\mu$m.

To sum up, in this section we investigated possible reasons for the successes and failures of our model using the CAM technique and concluded that inter-blastomere angle, an indicator of compaction, was one of the model's cues for prediction. However, compaction alone was an inferior predictor of polarization compared to the model, suggesting that our model learned additional features informative of polarization that we currently do not understand. The intriguing implication is that more discriminative biology is apparent in simple unstained embryo images than we currently realize.

## 6.6 Summary

In this chapter, we show that an ensemble deep learning model can identify polarization in unstained embryo images from the DIC microscope with an accuracy surpassing that of humans by a wide margin. When classifying 583 test DIC 8-cell stage frames, our model yielded an accuracy of 85% [95% confidence interval (CI): 82.2% - 88.2%] compared to corresponding average human accuracy of 61% [95% CI: 57.1% - 65.0%].

It is important to note the difficulty of the polarization detection task using unstained embryo images, since to the naked human eye, unstained images do not have any clear features which allow identification of the cellular apical domain. This is reflected in our observed human accuracy of 61%, which represents a performance level barely higher than random chance. Expressed as odds, the odds of a human volunteer correctly differentiating polarization were 1.5—that is, humans were right 1.5 times for each time they were wrong. In contrast, our deep learning model was right 5.7 times for each time it was wrong.

Current embryo selection in *in vitro* fertilization (IVF) clinics relies on crude and qualitative expert inspection of live embryos under plain microscopy that equates to an educated guess. Deep learning is an unusually well-suited solution to providing a more accurate assessment of embryo health for IVF, since deep neural networks recognize subtle features that are difficult for humans to identify [194–196]. Prior research in this field [176, 177, 180] limited itself only to features that are obvious on bright field or DIC imaging such as cell count and size, or to directly predict implantation potential without investigating underlying biological processes [179]. Our model can potentially enable embryo quality assessment using an important developmental milestone and thereby overcome some limitations of these prior deep learning studies. To our knowledge, there is currently no other known way to adequately evaluate the developmentally critical polarization milestone for embryo health screening prior to selection for implantation. By detecting an underlying developmental feature of the embryo using unstained embryo images, our study provides a platform for a potential future solution to improve IVF technology.

Our deep learning-based approach presents a significantly more accurate and less laborious way to evaluate mammalian embryo polarization compared to manual analysis. In future studies, this approach can be used to predict other features that indicate the quality of developing embryos, which can be measured by blastocyst implantation success rate. Furthermore, it could be useful for stainless tracking of polarization in live human embryos, allowing IVF clinics to assess embryo polarity status and its precise timing using non-invasive methods, and move away from empirical embryo grading to a system grounded in

established developmental milestones of embryogenesis. In the future, in order to further enhance the generalization ability of our trained model, we can utilize more diverse data, e.g., from different institutes and clinics.

In conclusion, we have developed a powerful non-invasive deep learning method to detect embryo polarization from images without the use of fluorescence, while surpassing human performance. This method has a great potential to provide the first example of detecting an underlying developmental feature of the mammalian embryo from unstained images, which could be an important step towards improving IVF technology from the rate of increase in cell number or assessment of morphological features independently of developmental events.

# BIBLIOGRAPHY

[1] A. Darmont, High dynamic range imaging: sensors and architectures. Society of Photo-Optical Instrumentation Engineers (SPIE), 2013.

[2] G.J. Lee, C. Choi, D.-H. Kim, and Y.M. Song, "Bioinspired Artificial Eyes: Optic Components, Digital Cameras, and Visual Prostheses.," Advanced Functional Materials. vol. 28, no. 24, p. 1705202, 2017.

[3] D.B. Williams and C.B. Carter, "The Transmission Electron Microscope.," In: D.B. Williams and C.B. Carter, Eds. Transmission Electron Microscopy: A Textbook for Materials Science. pp. 3–17. Springer US, Boston, MA (1996).

[4] K.L. (Katherine L. Bouman, "Extreme imaging via physical model inversion : seeing around corners and imaging black holes," https://dspace.mit.edu/handle/1721.1/113998, (2017).

[5] Y. Zhang and H. Gross, "Systematic design of microscope objectives. Part I: System review and analysis.," Advanced Optical Technologies. vol. 8, no. 5, pp. 313–347, 2019.

[6] H.H. Radamson, H. Zhu, Z. Wu, et al., "State of the Art and Future Perspectives in Advanced CMOS Technology.," Nanomaterials. vol. 10, no. 8, p. 1555, 2020.

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

[8] T. Cacace, V. Bianco, and P. Ferraro, "Quantitative phase imaging trends in biomedical applications.," Optics and Lasers in Engineering. vol. 135, p. 106188, 2020.

[9] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models.," Acta Numerica. vol. 28, pp. 1–174, 2019.

[10] J. Dong, L. Valzania, A. Maillard, T. Pham, S. Gigan, and M. Unser, "Phase Retrieval: From Computational Imaging to Machine Learning: A tutorial.," IEEE Signal Processing Magazine. vol. 40, no. 1, pp. 45–57, 2023.

[11] B. Wen, S. Ravishankar, Z. Zhao, R. Giryes, and J.C. Ye, "Physics-Driven Machine Learning for Computational Imaging [From the Guest Editor].," IEEE Signal Processing Magazine. vol. 40, no. 1, pp. 28–30, 2023.

[12] G. Ongie, A. Jalal, C.A. Metzler, R.G. Baraniuk, A.G. Dimakis, and R. Willett, "Deep Learning Techniques for Inverse Problems in Imaging.," IEEE Journal on Selected Areas in Information Theory. vol. 1, no. 1, pp. 39–56, 2020.

[13] U.S. Kamilov, C.A. Bouman, G.T. Buzzard, and B. Wohlberg, "Plug-and-Play Methods for Integrating Physical and Learned Models in Computational Imaging: Theory, algorithms, and applications.," IEEE Signal Processing Magazine. vol. 40, no. 1, pp. 85–97, 2023.

[14] A. Mikš and J. Novák, "Method of initial design of a two-element double-sided telecentric optical system.," Optics Express. vol. 31, no. 2, pp. 1604–1614, 2023.

[15] H. Gross, H. Zügge, M. Peschka, and F. Blechinger, Handbook of Optical Systems, Volume 3: Aberration Theory and Correction of Optical Systems. Wiley-Vch, 2006.

[16] A. Parmar, R. Katariya, and V. Patel, "A Review on Random Forest: An Ensemble Classifier.," In: J. Hemanth, X. Fernando, P. Lafata, and Z. Baig, Eds. International

Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018. pp. 758–763. Springer International Publishing, Cham (2018).

[17] W.S. Noble, "What is a support vector machine?," Nature Biotechnology. vol. 24, no. 12, pp. 1565–1567, 2006.

[18] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review.," Computational Intelligence and Neuroscience. vol. 2018, p. e7068349, 2018.

[19] L. Lu, Y. Zheng, G. Carneiro, and L. Yang, Eds., Deep Learning and Convolutional Neural Networks for Medical Image Computing. Springer International Publishing, Cham, 2017.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database.," In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009).

[21] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks.," Communications of the ACM. vol. 60, no. 6, pp. 84–90, 2017.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition.," In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE (2016).

[23] L. Liu, W. Ouyang, X. Wang, et al., "Deep learning for generic object detection: A survey.," International Journal of Computer Vision. vol. 128, pp. 261–318, 2019.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks.," In: Advances in Neural Information Processing Systems 28 (2015).

[25] W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single Shot MultiBox Detector.," In: B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Computer Vision – ECCV 2016. pp. 21–37. Springer International Publishing, Cham (2016).

[26] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy.," Nature Photonics. vol. 7, no. 9, pp. 739–745, 2013.

[27] E. Pirard, "Multispectral imaging of ore minerals in optical microscopy.," Mineralogical Magazine. vol. 68, no. 2, pp. 323–333, 2004.

[28] Y. Roggo, A. Edmond, P. Chalus, and M. Ulmschneider, "Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms.," Analytica Chimica Acta. vol. 535, no. 1–2, pp. 79–87, 2005.

[29] Y.W. Chu, F. Chen, Y. Tang, et al., "Diagnosis of nasopharyngeal carcinoma from serum samples using hyperspectral imaging combined with a chemometric method.," Optics Express. vol. 26, no. 22, pp. 28661–28671, 2018.

[30] D.L. Barton, K. Bernhard-Höfer, and E.I. Cole Jr, "FLIP-chip and 'backside' techniques.," Microelectronics Reliability. vol. 39, no. 6–7, pp. 721–730, 1999.

[31] C.D. Tran, "Principles, instrumentation, and applications of infrared multispectral imaging, an overview.," Analytical Letters. vol. 38, no. 5, pp. 735–752, 2005.

[32] K. Agarwal, R. Chen, L.S. Koh, C.J.R. Sheppard, and X. Chen, "Crossing the resolution limit in near-infrared imaging of silicon chips: targeting 10-nm node technology.," Physical Review X. vol. 5, no. 2, p. 021014, 2015.

[33] J. Chung, G.W. Martinez, K.C. Lencioni, S.R. Sadda, and C. Yang, "Computational aberration compensation by coded-aperture-based correction of aberration obtained from optical Fourier coding and blur estimation.," Optica. vol. 6, no. 5, pp. 647–661, 2019.

[34] B.C. Platt and R. Shack, "History and principles of Shack-Hartmann wavefront sensing.," Journal of Refractive Surgery. vol. 17, no. 5, pp. 573–577, 2001.

[35] J.L. Beverage, R.V. Shack, and M.R. Descour, "Measurement of the three-dimensional microscope point spread function using a Shack-Hartmann wavefront sensor.," Journal of Microscopy. vol. 205, no. 1, pp. 61–75, 2002.

[36] M.R. Teague, "Deterministic phase retrieval: a Green's function solution,." Journal of the Optical Society of America A. vol. 73, no. 11, pp. 1434–1441, 1983.

[37] L. Waller, L. Tian, and G. Barbastathis, "Transport of intensity phase amplitude imaging with higher order intensity derivatives.," Optics Express. vol. 18, no. 12, pp. 12552–12561, 2010.

[38] L. Tian, J.C. Petruccelli, and G. Barbastathis, "Nonlinear diffusion regularization for transport of intensity phase imaging.," Optics Letters. vol. 37, no. 19, pp. 4131–4133, 2012.

[39] C. Zuo, Q. Chen, Y. Yu, and A. Asundi, "Transport-of-intensity phase imaging using Savitzky-Golay differentiation filter–theory and applications.," Optics Express. vol. 21, no. 5, pp. 5346–5362, 2013.

[40] C. Zuo, Q. Chen, W. Qu, and A. Asundi, "High-speed transport-of-intensity phase microscopy with an electrically tunable lens.," Optics Express. vol. 21, no. 20, pp. 24060–24075, 2013.

[41] R.A. Gonsalves and R. Chidlaw, "Wavefront sensing by phase retrieval.," In: Applications of Digital Image Processing III. pp. 32–39 (1979).

[42] J.R. Fienup, "Phase retrieval algorithms: a comparison.," Applied Optics. vol. 21, no. 15, pp. 2758–2769, 1982.

[43] 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. vol. 9, no. 7, pp. 1072–1085, 1992.

[44] L. Waller, M. Tsang, S. Ponda, S.Y. Yang, and G. Barbastathis, "Phase and amplitude imaging from noisy images by Kalman filtering.," Optics Express. vol. 19, no. 3, pp. 2805–2815, 2011.

[45] J. Zhong, L. Tian, P. Varma, and L. Waller, "Nonlinear optimization algorithm for partially coherent phase retrieval and source recovery.," IEEE Transactions on Computational Imaging. vol. 2, no. 3, pp. 310–322, 2016.

[46] C. Shen, J. Tan, C. Wei, and Z. Liu, "Coherent diffraction imaging by moving a lens.," Optics Express. vol. 24, no. 15, pp. 16520–16529, 2016.

[47] C. Shen, X. Bao, J. Tan, S. Liu, and Z. Liu, "Two noise-robust axial scanning multi-image phase retrieval algorithms based on Pauta criterion and smoothness constraint.," Optics Express. vol. 25, no. 14, pp. 16235–16249, 2017.

[48] A. Greenbaum, A. Feizi, N. Akbari, and A. Ozcan, "Wide-field computational color imaging using pixel super-resolved on-chip microscopy.," Optics Express. vol. 21, no. 10, pp. 12469–12483, 2013.

[49] A. Greenbaum, Y. Zhang, A. Feizi, et al., "Wide-field computational imaging of pathology slides using lens-free on-chip microscopy.," Science Translational Medicine. vol. 6, no. 267, p. 267ra175, 2014.

[50] J.M. Rodenburg and H.M.L. Faulkner, "A phase retrieval algorithm for shifting illumination.," Applied Physics Letters. vol. 85, no. 20, pp. 4795–4797, 2004.

[51] A. Pan and B. Yao, "Three-dimensional space optimization for near-field ptychography.," Optics Express. vol. 27, no. 4, pp. 5433–5446, 2019.

[52] A. Pan, M. Zhou, Y. Zhang, J. Min, M. Lei, and B. Yao, "Adaptive-window angular spectrum algorithm for near-field ptychography.," Optics Communications. vol. 430, pp. 73–82, 2019.

[53] X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy.," Optics Express. vol. 22, pp. 4960–4972, 2014.

[54] L.H. Yeh, J. Dong, J. Zhong, et al., "Experimental robustness of Fourier ptychography phase retrieval algorithms.," Optics Express. vol. 23, no. 26, pp. 33214–33240, 2015.

[55] A. Pan, Y. Zhang, T. Zhao, et al., "System calibration method for Fourier ptychographic microscopy.," Journal of Biomedical Optics. vol. 22, no. 9, p. 096005, 2017.

[56] C. Zuo, J. Sun, and Q. Chen, "Adaptive step-size strategy for noise-robust Fourier ptychographic microscopy.," Optics Express. vol. 24, no. 18, pp. 20724–20744, 2016.

[57] J. Chung, H. Lu, X. Ou, H. Zhou, and C. Yang, "Wide-field Fourier ptychographic microscopy using laser illumination source.," Biomedical Optics Express. vol. 7, no. 11, pp. 4787–4802, 2016.

[58] L. Tian and L. Waller, "3D intensity and phase imaging from light field measurements in an LED array microscope.," Optica. vol. 2, no. 2, pp. 104–111, 2015.

[59] R. Horstmeyer, J. Chung, X. Ou, G. Zheng, and C. Yang, "Diffraction tomography with Fourier ptychography.," Optica. vol. 3, no. 8, pp. 827–835, 2016.

[60] S. Dong, R. Horstmeyer, R. Shiradkar, et al., "Aperture-scanning Fourier ptychography for 3D refocusing and super-resolution macroscopic imaging.," Optics Express. vol. 22, no. 11, pp. 13586–13599, 2014.

[61] X. Ou, J. Chung, R. Horstmeyer, and C. Yang, "Aperture scanning Fourier ptychographic microscopy.," Biomedical Optics Express. vol. 7, no. 8, pp. 3140–3150, 2016.

[62] A. Levin, R. Fergus, F. Durand, and W.T. Freeman, "Image and depth from a conventional camera with a coded aperture.," ACM Transactions on Graphics (TOG). vol. 26, no. 3, pp. 70-es, 2007.

[63] R. Horstmeyer, X. Ou, J. Chung, G. Zheng, and C. Yang, "Overlapped Fourier coding for optical aberration removal.," Optics Express. vol. 22, no. 20, pp. 24062–24080, 2014.

[64] G. Zheng, X. Ou, R. Horstmeyer, and C. Yang, "Characterization of spatially varying aberrations for wide field-of-view microscopy.," Optics Express. vol. 21, no. 13, pp. 15131–15143, 2013.

[65] R. Horstmeyer, R. Heintzmann, G. Popescu, L. Waller, and C. Yang, "Standardizing the resolution claims for coherent microscopy.," Nature Photonics. vol. 10, no. 2, pp. 68–71, 2016.

[66] C.T. Rueden, J. Schindelin, M.C. Hiner, et al., "ImageJ2: ImageJ for the next generation of scientific image data.," BMC Bioinformatics. vol. 18, p. 529, 2017.

[67] K. Vigil, Y. Lu, A. Yurt, et al., "Integrated circuit super-resolution failure analysis with solid immersion lenses.," Electronic Device Failure Analysis. vol. 16, no. 2, pp. 26–32, 2014.

[68] Y. Lu, T. Bifano, S. Ünlü, and B. Goldberg, "Aberration compensation in aplanatic solid immersion lens microscopy.," Optics Express. vol. 21, no. 23, pp. 28189–28197, 2013.

[69] P. Marquet, B. Rappaz, P.J. Magistretti, et al., "Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy.," Optics Letters. vol. 30, pp. 468–470, 2005.

[70] C.S. Seelamantula, N. Pavillon, C. Depeursinge, and M. Unser, "Exact complex-wave reconstruction in digital holography.," Journal of the Optical Society of America A. vol. 28, no. 6, pp. 983–992, 2011.

[71] G. Popescu, T. Ikeda, R.R. Dasari, and M.S. Feld, "Diffraction phase microscopy for quantifying cell structure and dynamics.," Optics Letters. vol. 31, no. 6, pp. 775–777, 2006.

[72] Z. Wang, L. Millet, M. Mir, et al., "Spatial light interference microscopy (SLIM).," Optics Express. vol. 19, no. 2, pp. 1016–1026, 2011.

[73] A. Anand, V. Chhaniwal, and B. Javidi, "Tutorial: common path self-referencing digital holographic microscopy.," APL Photonics. vol. 3, no. 7, p. 071101, 2018.

[74] N.T. Shaked, Y. Zhu, M.T. Rinehart, and A. Wax, "Two-step-only phase-shifting interferometry with optimized detector bandwidth for microscopy of live cells.," Optics Express. vol. 17, no. 18, pp. 15585–15591, 2009.

[75] P. Gao, B. Yao, J. Min, et al., "Parallel two-step phase-shifting point-diffraction interferometry for microscopy based on a pair of cube beamsplitters.," Optics Express. vol. 19, no. 3, pp. 1930–1935, 2011.

[76] C. Joo, T. Akkin, B. Cense, B.H. Park, and J.F. Boer, "Spectral domain optical coherence phase microscopy for quantitative phase contrast imaging.," Optics Letters. vol. 30, no. 16, pp. 2131–2133, 2005.

[77] Z. Yaqoob, W. Choi, S. Oh, et al., "Improved phase sensitivity in spectral domain phase microscopy using line-field illumination and self phase referencing.," Optics Express. vol. 17, no. 13, pp. 10681–10687, 2009.

[78] R.W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures.," Optik. vol. 35, no. 2, pp. 237–246, 1972.

[79] A. Pan, Y. Zhang, K. Wen, et al., "Subwavelength resolution Fourier ptychography with hemispherical digital condensers.," Optics Express. vol. 26, no. 18, pp. 23119–23131, 2018.

[80] S.B. Mehta and C.J. Sheppard, "Quantitative phase-gradient imaging at high resolution with asymmetric illumination-based differential phase contrast.," Optics Letters. vol. 34, no. 13, pp. 1924–1926, 2009.

[81] L. Tian and L. Waller, "Quantitative differential phase contrast imaging in an LED array microscope.," Optics Express. vol. 23, no. 9, pp. 11394–11403, 2015.

[82] H. Lu, J. Chung, X. Ou, and C. Yang, "Quantitative phase imaging and complex field reconstruction by pupil modulation differential phase contrast.," Optics Express. vol. 24, no. 22, pp. 25345–25361, 2016.

[83] Y. Park, M. Diez-Silva, G. Popescu, et al., "Refractive index maps and membrane dynamics of human red blood cells parasitized by plasmodium falciparum.," Proceedings of the National Academy of Sciences. vol. 105, no. 37, pp. 13730–13735, 2008.

[84] W.J. Eldridge, S. Ceballos, H.S. Park, and A. Wax, "Comparing quantitative phase derived cellular mechanical parameters with atomic force microscopy measurements (Conference Presentation).," In: Biophysics, Biology and Biophotonics IV: the Crossroads. pp. 1088803. SPIE, San Francisco, California, United States (2019).

[85] R. Horstmeyer, X. Ou, G. Zheng, P. Willems, and C. Yang, "Digital pathology with Fourier ptychography.," Computerized Medical Imaging and Graphics. vol. 42, pp. 38–43, 2015.

[86] J. Miao, D. Sayre, and H. Chapman, "Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects.," Journal of the Optical Society of America A. vol. 15, no. 6, pp. 1662–1669, 1998.

[87] A. Anand, V.K. Chhaniwal, P. Almoro, G. Pedrini, and W. Osten, "Shape and deformation measurements of 3D objects using volume speckle field and phase retrieval.," Optics Letters. vol. 34, no. 10, pp. 1522–1524, 2009.

[88] W. Zhang, L. Cao, D.J. Brady, et al., "Twin-image-free holography: a compressive sensing approach.," Physical Review Letters. vol. 121, no. 9, p. 093902, 2018.

[89] C. Zheng, R. Zhou, C. Kuang, G. Zhao, Z. Yaqoob, and P.T. So, "Digital micromirror device-based common-path quantitative phase imaging.," Optics Letters. vol. 42, no. 7, pp. 1448–1451, 2017.

[90] Y. Baek, K. Lee, S. Shin, and Y. Park, "Kramers–Kronig holographic imaging for high-space-bandwidth product.," Optica. vol. 6, no. 1, pp. 45–51, 2019.

[91] E.C. Titchmarsh, Introduction to the theory of Fourier integrals. Clarendon Press, 1937.

[92] J.P. Havlicek, J.W. Havlicek, and A.C. Bovik, "The analytic image.," In: Proceedings of International Conference on Image Processing. pp. 446–449. IEEE (1997).

[93] M. Chen, Z.F. Phillips, and L. Waller, "Quantitative differential phase contrast (DPC) microscopy with computational aberration correction.," Optics Express. vol. 26, no. 25, pp. 32888–32899, 2018.

[94] J. Sun, C. Zuo, J. Zhang, Y. Fan, and Q. Chen, "High-speed Fourier ptychographic microscopy based on programmable annular illuminations.," Scientific Reports. vol. 8, no. 1, p. 7669, 2018.

[95] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment.," IEEE Transactions on Image Processing. vol. 20, no. 8, pp. 2378–2386, 2011.

[96] Y. Zhang, H. Wang, Y. Wu, M. Tamamitsu, and A. Ozcan, "Edge sparsity criterion for robust holographic autofocusing.," Optics Letters. vol. 42, no. 19, pp. 3824–3827, 2017.

[97] Y. Baek and Y. Park, "Intensity-based holographic imaging via space-domain Kramers–Kronig relations.," Nature Photonics. vol. 15, no. 5, pp. 354–360, 2021.

[98] G. Zheng, C. Shen, S. Jiang, P. Song, and C. Yang, "Concept, implementations and applications of Fourier ptychography.," Nature Reviews Physics. vol. 3, no. 3, pp. 207–223, 2021.

[99] L. Bian, J. Suo, G. Zheng, K. Guo, F. Chen, and Q. Dai, "Fourier ptychographic reconstruction using Wirtinger flow optimization.," Optics Express. vol. 23, no. 4, p. 4856, 2015.

[100] L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for Fourier Ptychography with an LED array microscope.," Biomedical Optics Express. vol. 5, no. 7, pp. 2376–2389, 2014.

[101] H. Zhou, C. Shen, M. Liang, and C. Yang, "Analysis of postreconstruction digital refocusing in Fourier ptychographic microscopy.," Optical Engineering. vol. 61, no. 7, p. 073102, 2022.

[102] M. Liang, C. Bernadt, S.B.J. Wong, C. Choi, R. Cote, and C. Yang, "All-in-focus fine needle aspiration biopsy imaging based on Fourier ptychographic microscopy.," Journal of Pathology Informatics. vol. 13, p. 100119, 2022.

[103] C. Shen, M. Liang, A. Pan, and C. Yang, "Non-iterative complex wave-field reconstruction based on Kramers–Kronig relations.," Photonics Research. vol. 9, no. 6, pp. 1003–1012, 2021.

[104] C. Shen, A.C.S. Chan, J. Chung, D.E. Williams, A. Hajimiri, and C. Yang, "Computational aberration correction of VIS-NIR multispectral imaging microscopy based on Fourier ptychography.," Optics Express. vol. 27, no. 18, p. 24923, 2019.

[105] T. Kohlberger, "Whole-slide image focus quality: Automatic assessment and impact on AI cancer detection.," Journal of Pathology Informatics. vol. 10, no. 1, p. 39, 2019.

[106] R. Wang, R. Wang, P. Song, et al., "Virtual brightfield and fluorescence staining for Fourier ptychography via unsupervised deep learning.," Optics Letters. vol. 45, no. 19, pp. 5405–5408, 2020.

[107] A.W. Lambert, D.R. Pattabiraman, and R.A. Weinberg, "Emerging biological principles of metastasis.," Cell. vol. 168, no. 4, pp. 670–691, 2017.

[108] R. Taftaf, X. Liu, S. Singh, et al., "ICAM1 initiates CTC cluster formation and trans-endothelial migration in lung metastasis of breast cancer.," Nature Communications. vol. 12, p. 4867, 2021.

[109] V. Plaks, C.D. Koopman, and Z. Werb, "Circulating tumor cells.," Science. vol. 341, no. 6151, pp. 1186–1188, 2013.

[110] S.C.P. Williams, "Circulating tumor cells.," Proceedings of the National Academy of Sciences. vol. 110, no. 13, pp. 4861–4861, 2013.

[111] P.D. Potdar and N.K. Lotey, "Role of circulating tumor cells in future diagnosis and therapy of cancer.," Journal of Cancer Metastasis and Treatment. vol. 1, pp. 44–56, 2015.

[112] M. Pimienta, M. Edderkaoui, R. Wang, and S. Pandol, "The potential for circulating tumor cells in pancreatic cancer management.," Frontiers in Physiology. vol. 8, p. 381, 2017.

[113] C. Yang, F. Chen, S. Wang, and B. Xiong, "Circulating tumor cells in gastrointestinal cancers: current status and future perspectives.," Frontiers in Oncology. vol. 9, p. 1427, 2019.

[114] V. Hofman, S. Heeke, C.-H. Marquette, M. Ilié, and P. Hofman, "Circulating tumor cell detection in lung cancer: but to what end?," Cancers. vol. 11, no. 2, p. 262, 2019.

[115] Y.-P. Yang, T.M. Giret, and R.J. Cote, "Circulating tumor cells from enumeration to analysis: current challenges and future opportunities.," Cancers. vol. 13, no. 11, p. 2723, 2021.

[116] N. Aceto, A. Bardia, D.T. Miyamoto, et al., "Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis.," Cell. vol. 158, no. 5, pp. 1110–1122, 2014.

[117] Y. Hong, F. Fang, and Q. Zhang, "Circulating tumor cell clusters: What we know and what we expect.," International Journal of Oncology. vol. 49, no. 6, pp. 2206–2216, 2016.

[118] E. Schuster, R. Taftaf, C. Reduzzi, M.K. Albert, I. Romero-Calvo, and H. Liu, "Better together: circulating tumor cell clustering in metastatic cancer.," Trends in Cancer. vol. 7, no. 11, pp. 1020–1032, 2021.

[119] Z. Ao, S.H. Shah, L.M. Machlin, et al., "Identification of Cancer-Associated Fibroblasts in Circulating Blood from Patients with Metastatic Breast Cancer Identification of cCAFs from Metastatic Cancer Patients.," Cancer Research. vol. 75, no. 22, pp. 4681–4687, 2015.

[120] V.S. LeBleu and R. Kalluri, "A peek into cancer-associated fibroblasts: origins, functions and translational impact.," Disease Models & Mechanisms. vol. 11, no. 4, p. dmm029447, 2018.

[121] E. Sahai, I. Astsaturov, E. Cukierman, et al., "A framework for advancing our understanding of cancer-associated fibroblasts.," Nature Reviews Cancer. vol. 20, no. 3, pp. 174–186, 2020.

[122] Q. Ping, R. Yan, X. Cheng, et al., "Cancer-associated fibroblasts: overview, progress, challenges, and directions.," Cancer Gene Therapy. vol. 28, no. 9, pp. 984–999, 2021.

[123] M. Boya, T. Ozkaya-Ahmadov, B.E. Swain, et al., "High throughput, label-free isolation of circulating tumor cell clusters in meshed microwells.," Nature Communications. vol. 13, no. 1, p. 3385, 2022.

[124] P. Patil, T. Kumeria, D. Losic, and M. Kurkuri, "Isolation of circulating tumour cells by physical means in a microfluidic device: a review.," RSC Advances. vol. 5, no. 109, pp. 89745–89762, 2015.

[125] L. Chen, A.M. Bode, and Z. Dong, "Circulating tumor cells: moving biological insights into detection.," Theranostics. vol. 7, no. 10, pp. 2606–2619, 2017.

[126] Y. Maertens, V. Humberg, F. Erlmeier, et al., "Comparison of isolation platforms for detection of circulating renal cell carcinoma cells.," Oncotarget. vol. 8, no. 50, pp. 87710–87717, 2017.

[127] G. Vona, A. Sabile, M. Louha, et al., "Isolation by size of epithelial tumor cells: a new method for the immunomorphological and molecular characterization of circulating tumor cells.," American Journal of Pathology. vol. 156, no. 1, pp. 57–63, 2000.

[128] S. Zheng, H. Lin, J.-Q. Liu, et al., "Membrane microfilter device for selective capture, electrolysis and genomic analysis of human circulating tumor cells.," Journal of Chromatography A. vol. 1162, no. 2, pp. 154–161, 2007.

[129] I. Desitter, B.S. Guerrouahen, N. Benali-Furet, et al., "A new device for rapid isolation by size and characterization of rare circulating tumor cells.," Anticancer Research. vol. 31, no. 2, pp. 427–441, 2011.

[130] S.H. Seal, "Silicone flotation: A simple quantitative method for the isolation of free-floating cancer cells from the blood.," Cancer. vol. 12, no. 3, pp. 590–595, 1959.

[131] R. Gertler, R. Rosenberg, K. Fuehrer, M. Dahm, H. Nekarda, and J.R. Siewert, "Detection of circulating tumor cells in blood using an optimized density gradient centrifugation.," In: Molecular Staging of Cancer. pp. 149–155. Springer, Berlin, Heidelberg (2003).

[132] F.Z. Shahneh, "Sensitive antibody-based CTCs detection from peripheral blood.," Human Antibodies. vol. 22, no. 1–2, pp. 51–54, 2013.

[133] C. Yang, K. Zou, Z. Yuan, T. Guo, and B. Xiong, "Prognostic value of circulating tumor cells detected with the CellSearch System in patients with gastric cancer: evidence from a meta-analysis.," OncoTargets and Therapy. vol. 11, pp. 1013–1023, 2018.

[134] C. Mansilla, E. Soria, and N. Ramírez, "The identification and isolation of CTCs: a biological Rubik's cube,." Critical Reviews in Oncology/Hematology. vol. 126, pp. 129–134, 2018.

[135] Y. Xu, T. Qin, J. Li, et al., "Detection of circulating tumor cells using negative enrichment immunofluorescence and an in situ hybridization system in pancreatic cancer.," International Journal of Molecular Sciences. vol. 18, no. 4, p. 622, 2017.

[136] W. Guo, X.-R. Yang, Y.-F. Sun, et al., "Clinical significance of EpCAM mRNA-positive circulating tumor cells in hepatocellular carcinoma by an optimized negative enrichment and qRT-PCR–based platform.," Clinical Cancer Research. vol. 20, no. 18, pp. 4794–4805, 2014.

[137] E. Heitzer, M. Auer, C. Gasch, et al., "Complex Tumor Genomes Inferred from Single Circulating Tumor Cells by Array-CGH and Next-Generation SequencingCTC Analysis by Array-CGH and Next-Generation Sequencing.," Cancer Research. vol. 73, no. 10, pp. 2965–2975, 2013.

[138] M. Sha, S. Jeong, B. Qiu, et al., "Isolation of cancer-associated fibroblasts and its promotion to the progression of intrahepatic cholangiocarcinoma.," Cancer Medicine. vol. 7, no. 9, pp. 4665–4677, 2018.

[139] R. Jiang, S. Agrawal, M. Aghaamoo, R. Parajuli, A. Agrawal, and A.P. Lee, "Rapid isolation of circulating cancer associated fibroblasts by acoustic microstreaming for assessing metastatic propensity of breast cancer patients.," Lab on a Chip. vol. 21, no. 5, pp. 875–887, 2020.

[140] A.J. Williams, J. Chung, X. Ou, et al., "Fourier ptychographic microscopy for filtration-based circulating tumor cell enumeration and analysis.," Journal of Biomedical Optics. vol. 19, no. 6, p. 066007, 2014.

[141] C. Svensson, S. Krusekopf, J. Lücke, and M.T. Figge, "Automated detection of circulating tumor cells with naive Bayesian classifiers.," Cytometry Part A. vol. 85, no. 6, pp. 501–511, 2014.

[142] C.-M. Svensson, R. Hübler, and M.T. Figge, "Automated classification of circulating tumor cells and the impact of interobsever variability on classifier training and performance.," Journal of Immunology Research. vol. 2015, p. 573165, 2015.

[143] T.B. Lannin, F.I. Thege, and B.J. Kirby, "Comparison and optimization of machine learning methods for automated classification of circulating tumor cells.," Cytometry Part A. vol. 89, no. 10, pp. 922–931, 2016.

[144] M. Stevens, A. Nanou, L.W.M.M. Terstappen, C. Driemel, N.H. Stoecklein, and F.A.W. Coumans, "StarDist Image Segmentation Improves Circulating Tumor Cell Detection.," Cancers. vol. 14, no. 12, p. 2916, 2022.

[145] Y. Mao, Z. Yin, and J. Schober, "A deep convolutional neural network trained on representative samples for circulating tumor cell detection.," In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1–6 (2016).

[146] L.L. Zeune, Y.E. Boink, G. Dalum, et al., "Deep learning of circulating tumour cells.," Nature Machine Intelligence. vol. 2, no. 2, pp. 124–133, 2020.

[147] K. Drews-Elger, J.A. Brinkman, P. Miller, et al., "Primary breast tumor-derived cellular models: characterization of tumorigenic, metastatic, and cancer-associated fibroblasts in dissociated tumor (DT) cultures.," Breast Cancer Research and Treatment. vol. 144, pp. 503–517, 2014.

[148] A.D. Edelstein, M.A. Tsuchida, N. Amodaj, H. Pinkard, R.D. Vale, and N. Stuurman, "Advanced methods of microscope control using μManager software.," Journal of Biological Methods. vol. 1, no. 2, p. e10, 2014.

[149] T. Gehring, "thorlabs_apt," https://github.com/qpit/thorlabs_apt.

[150] "harvesters," https://github.com/genicam/harvesters.

[151] W. Boecker, W. Rolf, W.-U. Muller, and C. Streffer, "Autofocus algorithms for fluorescence microscopy.," In: Applications of Digital Image Processing XIX. pp. 445–456. SPIE, Denver, CO, United States (1996).

[152] T. Peng, K. Thorn, T. Schroeder, et al., "A BaSiC tool for background and shading correction of optical microscopy images.," Nature Communications. vol. 8, p. 14836, 2017.

[153] S.J. Yang, D.M.A. Marc Berndl, M. Barch, A. Narayanaswamy, E. Christiansen, and S. Hoyer, "Assessing microscope image focus quality with deep learning.," BMC Bioinformatics. vol. 19, p. 77, 2018.

[154] P. Bankhead, M.B. Loughrey, J.A. Fernández, et al., "QuPath: Open source software for digital pathology image analysis.," Scientific Reports. vol. 7, no. 1, pp. 1–7, 2017.

[155] T.-Y. Lin, M. Maire, S. Belongie, et al., "Microsoft coco: Common objects in context.," In: Computer Vision–ECCV 2014. pp. 740–755. Springer, Cham, Zurich, Switzerland (2014).

[156] Z.-Q. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: A review.," IEEE Transactions on Neural Networks and Learning Systems. vol. 30, no. 11, pp. 3212–3232, 2019.

[157] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection.," In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 2980–2988 (2017).

[158] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks.," In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1492–1500. IEEE (2017).

[159] M. Everingham, L. Gool, C.K.I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge.," International Journal of Computer Vision. vol. 88, pp. 303–338, 2009.

[160] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2.

[161] M.H. Johnson and C.A. Ziomek, "The foundation of two distinct cell lineages within the mouse morula.," Cell. vol. 24, no. 1, pp. 71–80, 1981.

[162] T.P. Fleming, P.M. Cannon, and S.J. Pickering, "The cytoskeleton, endocytosis and cell polarity in the mouse preimplantation embryo.," Developmental Biology. vol. 113, no. 2, pp. 406–419, 1986.

[163] C. Gerri, A. McCarthy, G. Alanis-Lobato, et al., "Initiation of a conserved trophectoderm program in human, cow and mouse embryos.," Nature. vol. 587, pp. 443–447, 2020.

[164] M. Zhu, M. Shahbazi, A. Martin, et al., "Human embryo polarization requires PLC signaling to mediate trophectoderm specification.," eLife. vol. 10, p. e65068, 2021.

[165] S. Anani, S. Bhat, N. Honma-Yamanaka, D. Krawchuk, and Y. Yamanaka, "Initiation of Hippo signaling is linked to polarity rather than to cell position in the pre-implantation mouse embryo.," Development. vol. 141, no. 14, pp. 2813–2824, 2014.

[166] C.R. Samarage, M.D. White, Y.D. Álvarez, et al., "Cortical Tension Allocates the First Inner Cells of the Mammalian Embryo.," Developmental Cell. vol. 34, no. 4, pp. 435–447, 2015.

[167] C. Gerri, S. Menchero, S.K. Mahadevaiah, J.M.A. Turner, and K.K. Niakan, "Human Embryogenesis: A Comparative Perspective.," Annual Review of Cell and Developmental Biology. vol. 36, no. 1, pp. 411–440, 2020.

[168] M.H. Johnson and C.A. Ziomek, "Induction of polarity in mouse 8-cell blastomeres: specificity, geometry, and stability.," Journal of Cell Biology. vol. 91, no. 1, pp. 303–308, 1981.

[169] S. Louvet, J. Aghion, A. Santa-Maria, P. Mangeat, and B. Maro, "Ezrin Becomes Restricted to Outer Cells Following Asymmetrical Division in the Preimplantation Mouse Embryo.," Developmental Biology. vol. 177, no. 2, pp. 568–579, 1996.

[170] S. Vinot, T. Le, S. Ohno, T. Pawson, B. Maro, and S. Louvet-Vallée, "Asymmetric distribution of PAR proteins in the mouse embryo begins at the 8-cell stage during compaction.," Developmental Biology. vol. 282, no. 2, pp. 307–319, 2005.

[171] B. Plusa, S. Frankenberg, A. Chalmers, et al., "Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo.," Journal of Cell Science. vol. 118, no. 3, pp. 505–515, 2005.

[172] M. Zhu, J. Cornwall-Scoones, P. Wang, et al., "Developmental clock and mechanism of de novo polarization of the mouse embryo.," Science. vol. 370, no. 6522, p. eabd2703, 2020.

[173] V.B. Alarcon, "Cell Polarity Regulator PARD6B Is Essential for Trophectoderm Formation in the Preimplantation Mouse Embryo1.," Biology of Reproduction. vol. 83, no. 3, pp. 347–358, 2010.

[174] E. Korotkevich, R. Niwayama, A. Courtois, et al., "The Apical Domain Is Required and Sufficient for the First Lineage Segregation in the Mouse Embryo.," Developmental Cell. vol. 40, no. 3, pp. 235–247, 2017.

[175] A. Ralston and J. Rossant, "Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo.," Developmental Biology. vol. 313, no. 2, pp. 614–629, 2008.

[176] A. Khan, S. Gould, and M. Salzmann, "Deep Convolutional Neural Networks for Human Embryonic Cell Counting.," In: G. Hua and H. Jégou, Eds. Computer Vision – ECCV 2016 Workshops. pp. 339–348. Springer, Cham (2016).

[177] R.M. Rad, P. Saeedi, J. Au, and J. Havelock, "Blastomere Cell Counting and Centroid Localization in Microscopic Images of Human Embryo.," In: 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP). pp. 1–6. IEEE (2018).

[178] T.-J. Chen, W.-L. Zheng, C.-H. Liu, I. Huang, H.-H. Lai, and M. Liu, "Using Deep Learning with Large Dataset of Microscope Images to Develop an Automated Embryo Grading System.," Fertility & Reproduction. vol. 1, no. 1, pp. 51–56, 2019.

[179] C.L. Bormann, M.K. Kanakasabapathy, P. Thirumalaraju, et al., "Performance of a deep learning based neural network in the selection of human blastocysts for implantation.," eLife. vol. 9, p. e55301, 2020.

[180] B.D. Leahy, W.-D. Jang, H.Y. Yang, et al., "Automated Measurements of Key Morphological Features of Human Embryos for IVF.," In: A.L. Martel, P. Abolmaesumi, D. Stoyanov, et al., Eds. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 25–35. Springer, Cham (2020).

[181] D. Tran, S. Cooke, P.J. Illingworth, and D.K. Gardner, "Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer.," Human Reproduction. vol. 34, no. 6, pp. 1011–1018, 2019.

[182] N. Dard, S. Louvet, A. Santa-Maria, et al., "In Vivo Functional Analysis of Ezrin during Mouse Blastocyst Formation.," Developmental Biology. vol. 233, no. 1, pp. 161–173, 2001.

[183] M. Zernicka-Goetz, J. Pines, S.M. Hunter, et al., "Following cell fate in the living mouse embryo.," Development. vol. 124, no. 6, pp. 1133–1137, 1997.

[184] J. Zenker, M.D. White, M. Gasnier, et al., "Expanding Actin Rings Zipper the Mouse Embryo for Blastocyst Formation.," Cell. vol. 173, no. 3, pp. 776–791, 2018.

[185] A. Ioannidou, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "Deep Learning Advances in Computer Vision with 3D Data: A Survey.," ACM Computing Surveys. vol. 50, no. 2, pp. 1–38, 2017.

[186] S.P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3D Deep Learning on Medical Images: A Review.," Sensors. vol. 20, no. 18, p. 5097, 2020.

[187] K. Weiss, T.M. Khoshgoftaar, and D. Wang, "A survey of transfer learning.," Journal of Big Data. vol. 3, no. 1, p. 9, 2016.

[188] B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser, "Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images.," Microscopy Research and Technique. vol. 65, no. 1–2, pp. 33–42, 2004.

[189] G. Huang, Z. Liu, L. van der Maaten, and K.Q. Weinberger, "Densely Connected Convolutional Networks.," In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4700–4708. IEEE (2017).

[190] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks.," In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7132–7141. IEEE (2018).

[191] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization.," In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2921–2929. IEEE (2016).

[192] W.J.D. Reeve, "Cytoplasmic polarity develops at compaction in rat and mouse embryos.," Development. vol. 62, no. 1, pp. 351–367, 1981.

[193] M. Zhu, C.Y. Leung, M.N. Shahbazi, and M. Zernicka-Goetz, "Actomyosin polarisation through PLC-PKC triggers symmetry breaking of the mouse embryo.," Nature Communications. vol. 8, no. 1, p. 921, 2017.

[194] E.M. Christiansen, S.J. Yang, D.M. Ando, et al., "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images.," Cell. vol. 173, no. 3, pp. 792–803, 2018.

[195] C. Ounkomol, S. Seshamani, M.M. Maleckar, F. Collman, and G.R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy.," Nature Methods. vol. 15, pp. 917–920, 2018.

[196] R. Poplin, A.V. Varadarajan, K. Blumer, et al., "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning.," Nature Biomedical Engineering. vol. 2, pp. 158–164, 2018.