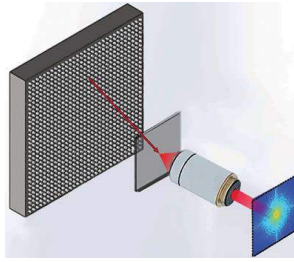




DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
CHENNAI – 600036

Learned Wirtinger Gradients for Fourier Ptychographic Phase Retrieval



A Thesis

Submitted by

ATREYEE SAHA

For the award of the degree

Of

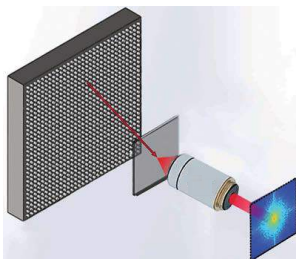
MASTER OF SCIENCE

July 2022



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
CHENNAI – 600036

Learned Wirtinger Gradients for Fourier Ptychographic Phase Retrieval



A Thesis

Submitted by

ATREYEE SAHA

For the award of the degree

Of

MASTER OF SCIENCE

July 2022

*One of the pleasures of looking at the world
through mathematical eyes is that you can
see certain patterns that would otherwise be
hidden.*

– Steven Strogatz

THESIS CERTIFICATE

This is to undertake that the Thesis titled **LEARNED WIRTINGER GRADIENTS FOR FOURIER PTYCHOGRAPHIC PHASE RETRIEVAL**, submitted by me to the Indian Institute of Technology Madras, for the award of **Master of Science**, is a bona fide record of the research work done by me under the supervision of **Dr. Kaushik Mitra**. The contents of this Thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Chennai 600036

Atreyee Saha

Date: July 2022

Dr. Kaushik Mitra

Research advisor

Assistant Professor

Department of Electrical Engineering

IIT Madras

ACKNOWLEDGEMENTS

I acknowledge my supervisors, Dr. Kaushik Mitra and Dr. Shanti Bhattacharya for providing me the opportunity to work with them and learn, for their thoughtful guidance throughout the course of the project, and for being considerate towards me during the difficult times. I express my gratitude towards Salman for his valuable insights and Sagar for his constant efforts. Finally, I wish to thank my labmates Gopi and Prasan for supporting me and my friends at IIT Madras with whom I share my journey.

ABSTRACT

Fourier Ptychographic Microscopy (FPM) is an imaging procedure that has been under development for the past few years. The technique realizes a higher Space-Bandwidth Product (SBP) utilizing images captured using a low NA objective and enables high-resolution phase imaging. Recent developments include mainly two approaches – iterative methods based on the knowledge of the forward imaging model and the data-driven approach, which commonly employs a deep learning framework to learn the mapping between intensity images and phase. Here, a hybrid ResNet architecture called Learned Wirtinger Gradient Network (LWGNet) is proposed, which combines the knowledge of the forward imaging model with the data-driven approach. It unrolls the traditional Wirtinger flow optimization algorithm into a novel neural network design that enhances the gradients through CNN layers. The proposed method also incorporates complex-valued operations into the conventional real-valued neural network building blocks. Using simulation and experiments, it is shown that incorporating the physical model into the data-driven framework successfully preserves finer details for low-cost and low dynamic range CMOS sensors and performs better than entirely data-driven methods. This improvement in performance for low-bit depth and low-cost sensors can significantly bring down the FPM imaging setup’s cost. The proposed method has shown consistently improved performance on collected real data. Finally, a method is proposed to extend the model to multiplexed LED illumination scheme.

CONTENTS

	Page
ACKNOWLEDGEMENTS	i
ABSTRACT	iii
LIST OF TABLES	vii
LIST OF FIGURES	ix
ABBREVIATIONS	xiii
NOTATION	xv
CHAPTER 1 INTRODUCTION	1
1.1 Microscopic Imaging	1
1.2 Fourier Ptychographic Microscopy	3
1.3 Motivation and Scope	4
1.4 Contributions	6
CHAPTER 2 PHYSICS OF IMAGE FORMATION	7
2.1 Diffraction-limited optical system	7
2.1.1 Abbe theory of imaging	8
2.1.2 Point Spread Function	9
2.1.3 Relation between NA and resolvable capacity	10
2.2 Forward Imaging model	11
CHAPTER 3 PRELIMINARY STUDY OF PHASE RETRIEVAL ALGORITHMS FOR FPM	15
3.1 The Problem of Phase retrieval	15
3.2 Related Works	16
3.3 Iterative Methods for Fourier Ptychographic Phase Retrieval	19
3.3.1 Alternate Projection based method for Fourier Ptychographic Phase Retrieval	19
3.3.2 Gradient descent based method for Fourier Ptychographic Phase Retrieval	21
3.4 Data-driven Methods for Fourier Ptychographic Phase Retrieval	25
3.4.1 Complex-valued Neural Network for FPM Phase Retrieval	26
3.4.2 Network architecture	26

CHAPTER 4	LEARNED WIRTINGER GRADIENTS NETWORK (LWGNET) FOR FPM PHASE RETRIEVAL	31
4.1	Proposed Network Architecture	32
4.1.1	Loss Function	34
4.2	Multiplexed illumination	35
4.2.1	Network architecture	35
4.2.2	Multiplexing scheme	36
CHAPTER 5	EXPERIMENTS AND RESULTS	37
5.1	Experimental Setup	37
5.2	Dataset	38
5.2.1	Simulated Dataset	38
5.2.2	Captured Real Dataset	39
5.2.3	Training and Evaluation	39
5.3	Results on Simulated dataset	40
5.3.1	Computational efficiency comparison	41
5.4	Results on Real Dataset	42
5.4.1	Sequential Illumination	43
5.4.2	Multiplexed Illumination	45
5.4.3	Sensor Quantization Analysis	45
5.5	Ablation Study	46
CHAPTER 6	CONCLUSION	51
6.1	Future Work	51
APPENDIX A	APPENDIX: EXPERIMENTAL SPECIFICATION	53
A.1	Calculation of Resolvable Capacity	53
A.2	Experimental Dataset	54
A.3	Submissions	56
REFERENCES		57
CURRICULUM VITAE		61
GENERAL TEST COMMITTEE		63

LIST OF TABLES

Table	Caption	Page
3.1	Comparison of iterative and neural-network based approaches.	18
4.1	Channel dimension of $\Psi_t\{\cdot\}$ for sequential scheme.	34
4.2	Channel dimensions of $\Psi_t\{\cdot\}$ for multiplexed scheme.	36
5.1	Compared to previous works, the proposed LWGNet achieves a better reconstruction quality at lower bit-depths.	42
5.2	Compared to previous works, the proposed LWGNet achieves a better reconstruction quality at lower bit-depths using fewer trainable parameters.	42

LIST OF FIGURES

Figure	Caption	Page
1.1	The compound microscope is equivalent to a magnifier, but contains multiple lenses. A light source is placed below the object. The overall system magnification is contributed by the objective and the eyepiece. The objective is placed at a working distance from the object and produces a real image on the imaging plane. The eyepiece projects and further magnifies the real image to produce a virtual image of the object. <i>source: edm (2022)(edmundoptics.jp)</i>	1
1.2	(a) Light path for bright-field and dark-field images. (b) Camera I (left) and Camera II (right) specifications. (c) Examples of input low resolution measurements under different sensor quantization	4
2.1	At the front focal plane x , a planar light beam is obliquely incident at an angle θ and spatial frequency f_x on the x -boundary. The beam is convergent on the back focal plane x' . The image is a delta function whose location is a function of the spatial frequency, f_x	7
2.2	Equivalent model of a microscopic imaging system.	8
2.3	(a) Experimental Setup of FPM. The object is placed between a planar LED array and objective of the inverted microscope. Low resolution images corresponding to ℓ^{th} LED are captured. (b) Representation of the equivalent forward model, comprises of down-sampling the object spectrum by the system pupil function centred at $(k_{x\ell}, k_{y\ell})$ for the ℓ^{th} LED.	11
3.1	The forward and inverse problems.	15
3.2	Flowchart for alternating projections algorithm for phase reconstruction (based on Gerchberg-Saxton algorithm Gerchberg (1972))	20
3.3	FPM phase reconstruction using Wirtinger update method. Artifacts at 50 and 100 iterations are highlighted in red. (a) Ground truth (b)Wirtinger Flow (c) Wirtinger Flow with NAG, and (d) Stochastic update.	25
3.4	FP-UNet architecture	27
3.5	FP-ResNet architecture	27
4.1	The iterative and the direct approaches for solving inverse problems. . .	31
4.2	The model architecture of LWGNet comprising of three stages of the LWG update block. The inputs are the object and the pupil initialization and measurement images for every stage. LWG block includes a Gradient computation block followed by three stages of Conv+CTanh layers followed by a Fully Connected(FC) layer. The adder block updates the input object field using the modified gradients.	33
4.3	LED positions on a 15×15 grid (top) and Sample LR images of Osteosarcoma slide (bottom).	36

5.1	FPM Experimental setup. Camera I is the high-dynamic range sCMOS camera that captures 16-bit images, while Camera I is the low cost low dynamic range machine vision camera to capture low bit-depth images.	37
5.2	Phase reconstruction on simulated data using proposed LWGNet method are perceptually superior at lower-bit depth depths compared to existing methods. Traditional algorithms WF and AP do not perform well under lower bit-depths. Among deep methods, cDIP performs poorly at lower bit depths.	41
5.3	Reconstruction on 16-bit LR images using Lung carcinoma sample (top) and Osteosarcoma (bottom). (a) Sample FoV LWGNet reconstruction, (b) LWGNet insets (c) F-cGAN reconstruction (d) cDIP reconstruction and (e) LR bright-field image.	43
5.4	Reconstruction on 12-bit LR images using Lung carcinoma sample (top and bottom).(a) Sample FoV LWGNet reconstruction, (b) LWGNet insets (c) F-cGAN reconstruction (d) cDIP reconstruction and (e) LR bright-field image.	44
5.5	Reconstruction on 8-bit LR images for Lung carcinoma sample (top) and Osteosarcoma (bottom). (a) Sample FoV LWGNet reconstruction, (b) LWGNet insets (c) F-cGAN reconstruction (d) cDIP reconstruction and (e) LR bright-field image.	44
5.6	Comparison of phase reconstruction results for Osteosarcoma(top) and Lung carcinoma (bottom) samples for (a) Sequential and (b) Multiplexed illumination schemes using 225 and 45 images respectively.	45
5.7	Comparison of phase reconstruction using the proposed LWGNet on Osteosarcoma (top) and Lung carcinoma (bottom) at various bit depth settings. The reconstruction quality is consistent even under high quantization noise.	46
5.8	Left: Representation of complex-valued filters(top) and interleaved real-valued filters(bottom). Right: Using complex-valued operations leads to superior quality of reconstructions.	47
5.9	Phase reconstruction on 16-bit simulated data using (b) FP-ResNet architecture and (c) FP-UNet architecture. The FP-ResNet shows marginal improvement over FP-UNet which shows contrast artifacts. . .	48
5.10	Left: Representation of processing of gradients(top) and processing $\hat{\phi}^i$ post update(bottom). Right: The proposed way of learning the network on gradient leads to more effective use of gradient information than learning a network on estimated object field.	49
5.11	Left: PSNR vs Number of unrolled stages. Right: corresponding reconstructions for 3 and 5 stage LWGNet under different bit depths. For 16 and 12 bit data, after just 3 stages the performance saturates while 5 stages are needed for 8 bit data.	50
A.1	(a) USAF chart and (b) Measurement of cell image.	53
A.2	Histological samples used for training. Top row shows the brightfield images captured using sCMOS camera while the bottom row shows the corresponding AP phase reconstructions.	55

A.3	Histological samples used for testing. Top row shows the brightfield images captured using sCMOS camera while the bottom row shows the corresponding AP phase reconstructions.	55
-----	--	----

ABBREVIATIONS

- AP** Alternating Projections.
- CNN** Convolutional Neural Network.
- FP** Fourier Ptychography.
- FPM** Fourier Ptychographic Microscopy.
- HR** High Resolution.
- LR** Low Resolution.
- LWG** Learned Wirtinger Gradient.
- NA** Numerical Aperture.
- OTF** Optical Transfer Function.
- PSF** Point Spread Function.
- VGG** Visual Geometry Group.
- WF** Wirtinger Flow.

NOTATION

\mathbb{O}	Object-field
I	Intensity measurements
P	Pupil function

CHAPTER 1

INTRODUCTION

1.1 MICROSCOPIC IMAGING

The optical microscope is an indispensable tool for pathological examinations. Modern digital pathology interfaces the optical microscope with digital imaging devices for automated analysis. The simplest form of microscope is a compound microscope shown in Figure 1.1.

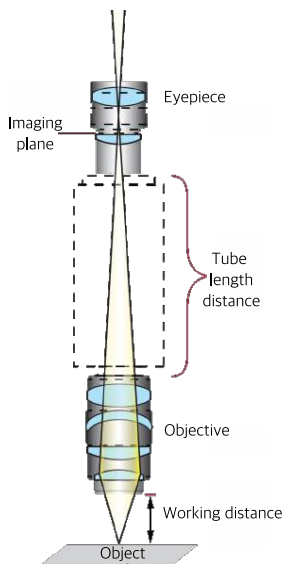


Figure 1.1: The compound microscope is equivalent to a magnifier, but contains multiple lenses. A light source is placed below the object. The overall system magnification is contributed by the objective and the eyepiece. The objective is placed at a working distance from the object and produces a real image on the imaging plane. The eyepiece projects and further magnifies the real image to produce a virtual image of the object.
source: edm (2022)(edmundoptics.jp)

Light waves undergo changes in amplitude and phase when passed through a specimen. The simple light microscope captures the amplitude variations since optical detectors convert photonic energy to electrical signals and can only measure the intensity (squared magnitude) of the light wave. i.e. the intensity. Phase contains information about material properties but is not measured directly. Specific sample characteristics are

difficult to image due to low contrast and are typically imaged in dark-field mode. For this, combining high-quality sensors and advanced digital image processing techniques enables optimized imaging. Such a microscopic system also offers a trade-off between resolution and Field-of-View (FoV), and movement of stage is required to image the entire slide. For characterization of 3D micro-structures, high resolution is achieved using an electronic beam of shorter wavelength than visible light.

Various observation modes have been developed for certain light characteristics arising out of specific sample properties that are required to be imaged. For example, for imaging transparent samples, *Phase contrast microscopy* uses phase shifting the background illumination beam with respect to the scattered beam from the sample to obtain high contrast. The *Differential Interference Contrast* mode uses constructive and destructive interference between the two polarized wavefronts vibrating perpendicularly to each other. *Fluorescence microscopy* images samples stained using a fluorescent compound.

Light microscopes interfaced with digital imaging systems are used to capture images of fine features; it requires detectors with high Signal-to-Noise Ratio (SNR) characteristics. However, as shown by Abbe (1873), the maximum attainable resolution by a diffraction-limited system is of the order of the wavelength of the light. The diffraction limit is in the order of fractions of microns. If operated at shorter wavelengths in the X-ray spectral region, optical components are found to suffer from aberrations.

One of the main challenges in computational medical imaging is to make imaging technology accessible for point of care diagnostics in resource-constrained communities. One way to make these techniques accessible includes designing cheaper and portable hardware. Algorithmic methods are an economic alternative to expensive hardware modifications. Breslauer *et al.* (2009); Smith *et al.* (2011) have used smartphone cameras for microscopy, while Aidukas *et al.* (2019) have used 3D printed microscopes for a wide field of view High-Resolution(HR) imaging. However, using cheaper and inefficient hardware limits the imaging capabilities of these systems through various degradations,

which can only be dealt with by designing effective computational algorithms. In the following section, a computational method of Fourier Ptychography is introduced, which addresses the problem of achieving HR phase imaging using an inexpensive hardware modification to the conventional light microscope.

1.2 FOURIER PTYCHOGRAPHIC MICROSCOPY

The number of resolvable pixels is measured by dividing the FoV by the imaging system's half-pitch resolution (in two dimensions). The quantity is the imaging system's Space-Bandwidth Product (SBP). FPM is a computational technique that increases the SBP of an imaging system by acquiring multiple images of the sample. FPM solves the problem of capturing a large Field of View (FoV) and achieving HR imaging with a minimum sample translation. It can also computationally compensate for system aberrations and overcomes the physical limitations of the optical system.

The FPM setup comprises of a low Numerical Aperture(NA) objective and an array illumination source, such that the sample may be illuminated from multiple angles. Each angle is associated with a region in the underlying object spectrum, and the intensity information is captured by the corresponding Low-Resolution (LR) image.

Imaging modes

Each low resolution image is captured with the sample illuminated from an angle, θ , corresponding to the position of the LED that determines the amount of light that passes through the aperture of the objective lens to the image sensor. There are two imaging modes,

- Bright-field imaging: It is the standard method of creating an image. Optically, thicker regions appear dark against parts with transmittance close to 1 appear lighter. The image formed is a 2D projection of the sample.
- Dark-field imaging: The object, when illuminated by peripheral LEDs, under

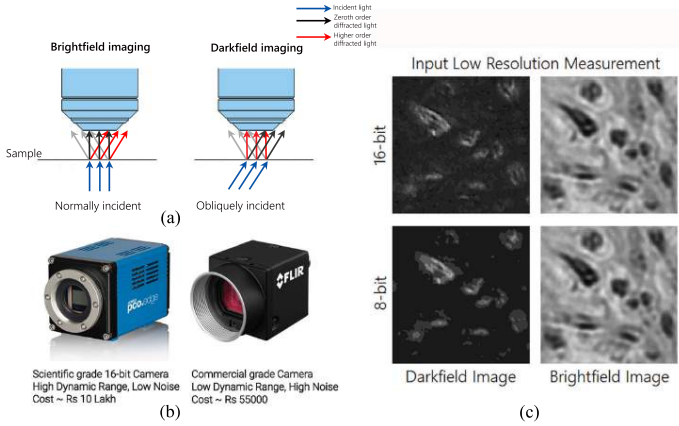


Figure 1.2: (a) Light path for bright-field and dark-field images. (b) Camera I (left) and Camera II (right) specifications. (c) Examples of input low resolution measurements under different sensor quantization

a higher angle of illumination, a significant fraction of the object light beam escapes the aperture. The light that passes through the lens is scattered light that corresponds to higher orders of the diffracted beam.

Figure 1.2 shows examples of bright-field and dark-field images captured from our experimental setup. We observe that there is significant loss of information due to quantization from 16-bit to 8-bit.

1.3 MOTIVATION AND SCOPE

FP finds applications in various imaging fields. Our work focuses on the following potential applications,

1. FPM can capture a large area at high resolution and has the potential to find application in histopathological imaging. Portable FPM systems, if developed, can provide inexpensive whole-slide imaging in remote areas with limited resources.

2. FPM offers higher throughput and phase reconstruction and can improve machine-learning-based algorithms' accuracy for diagnosis in medical fields such as radiology and pathology.
3. Due to minimum or no slide translation, FPM offers high-speed phase imaging of in-vitro cultures. An inexpensive method to improve the image acquisition speed is the multiplexing scheme, as discussed in later chapters. The captured images have high contrast.

Despite extensive study, there has not been enough work done to make the FPM system accessible for point-of-care diagnostics. The challenges are threefold. First, for implementation of FPM, like most low-light imaging modalities, suffer from significant noise and dynamic range problems, especially for dark-field images. Current FPM systems rely on expensive optics and scientific-grade sCMOS sensors to increase the SNR and dynamic range of low-resolution dark-field captures, which significantly increases the system cost Tian *et al.* (2014). Second, existing reconstruction algorithms designed for FPM systems have shown results only for these expensive systems and are unlikely to perform optimally when the quality of the sensor degrades. Third, the existing reconstruction techniques are either slow or model-independent data-driven techniques that entirely ignore the forward imaging process. One related work on making FPM systems accessible is Aidukas *et al.* (2019). However, the authors rely on extensive pre-processing and calibration before reconstructing with traditional algorithms.

Keeping the above challenges in mind, the 'LWGNet' for sequential FPM reconstruction is proposed. LWGNet is a novel physics-driven unrolled neural network that combines the expressiveness of data-driven techniques with the interpretability of iterative Wirtinger flow based techniques Bian *et al.* (2015). Unlike existing unrolled networks, LWGNet performs most of its operations on image gradient through *complex-valued* functions. Specifically, it learns a non-linear mapping from complex stochastic gradients to intermediate object fields through complex-valued neural networks. Such a learned

mapping helps preserve the HR details in the peripheral dark-field images, especially for low dynamic range sensors. LWGNet outperforms both traditional and deep-learning methods in terms of reconstruction quality, especially for low-cost machine vision sensors with poor dynamic range. This is shown by performing extensive evaluations on simulated and real histopathological data captured under different bit depths.

1.4 CONTRIBUTIONS

In summary, the following contributions are made:

- The LWGNet is proposed, which is a physics-inspired complex-valued feed-forward network for FPM reconstructions that exploits the physics of the FPM model and data-driven methods.
- The proposed approach uses a learned complex-valued non-linear mapping from gradients to object field that helps restore the finer details under a low dynamic range scenario, thereby reducing the gap in reconstruction quality between expensive HDR sCMOS cameras and low-cost 8-bit CMOS sensors. This enables reducing the cost of the experimental setup to a large extent.
- A real dataset is collected of 8, 12 and 16 bit LR measurements using a CMOS sensor and the corresponding aligned ground-truth for finetuning our method. It is the only FPM dataset captured with multiple bit depth settings using a low-cost CMOS sensor to the best of our knowledge.
- The proposed network outperforms existing traditional and learning-based algorithms in terms of reconstruction quality for both simulated and real data, as verified through extensive experiments on challenging histopathological samples.

CHAPTER 2

PHYSICS OF IMAGE FORMATION

In this chapter, the concepts related to the forward imaging model of FPM are discussed. The chapter starts with the properties of a diffraction-limited optical system. The behaviour of the imaging lens for an input planar wavefront and the microscopic model of image formation is discussed. Then the concept of pupil function is defined. Finally, the chapter is concluded with the mathematical model of FPM in Section 2.2.

2.1 DIFFRACTION-LIMITED OPTICAL SYSTEM

The concept of Fourier transform and the Fourier space is essential to develop a mathematical model microscopic image formation. Let a simple sine wave be considered,

$$U_0(x, y) = a \cos(k_x x + k_y y + \alpha)$$

where a is the amplitude, α is the phase or the propagation angle.

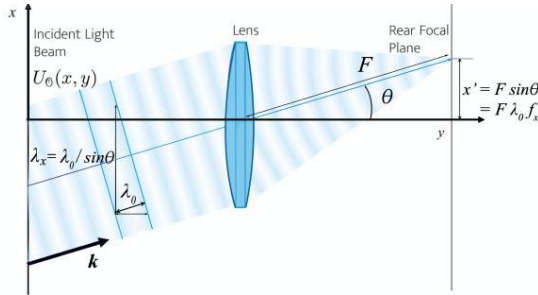


Figure 2.1: At the front focal plane x , a planar light beam is obliquely incident at an angle θ and spatial frequency f_x on the x -boundary. The beam is convergent on the back focal plane x' . The image is a delta function whose location is a function of the spatial frequency, f_x .

Let the microscope be represented by an equivalent focal length F of its objective. Let U_o form an angle θ to the optical axis and focus on the rear focal plane at a height $x' = F \sin \theta = F \lambda_0 f_x$ on the back focal plane; similarly, the location of an image on the y -boundary can be determined. If the intensity distribution at the rear focal plane is captured, it is found to be a function of spatial frequency $\mathbf{k} = (k_x, k_y)$, measured from the centre of the objective.

2.1.1 Abbe theory of imaging

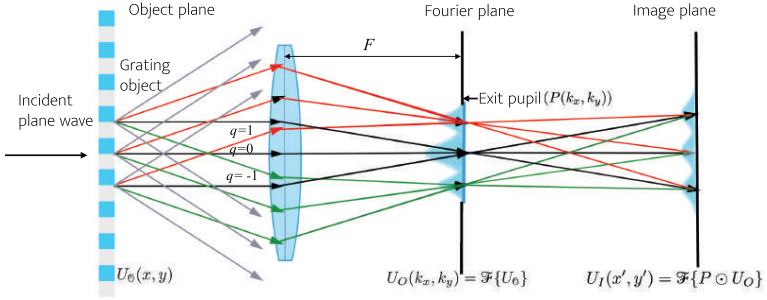


Figure 2.2: Equivalent model of a microscopic imaging system.

The previous section discussed that the imaging system represented by an equivalent lens performs a Fourier transform on the rear focal plane. The microscope, in effect, performs two successive Fourier transforms. On the object plane of Figure 2.2, we have a diffraction grating. The light wavefront can be decomposed into Fourier components as it undergoes diffraction at the grating. U_O Abbe's theory suggests that the image formed is the interference of the diffraction phenomenon. Each spatial frequency generates pairs of plane waves (diffraction orders) that propagate symmetrically with the optical axis at

the object plane. Then, U_O is Fourier transform of U_Θ . U'_I is the Fourier transform of U_O .

$$U'_I(x', y') = \iint_{-\infty}^{\infty} U_O(k_x, k_y) e^{-j(k_x x' + k_y y')} dk_x dk_y \quad (2.1)$$

$$U_O(k_x, k_y) = \iint_{-\infty}^{\infty} U_\Theta(x, y) e^{-j(k_x x + k_y y)} dx dy \quad (2.2)$$

Therefore, due to the successive Fourier transforms,

$$U'_I(x', y') = U_\Theta(-x, -y) \quad (2.3)$$

Equation 2.3 shows why an inverted image is observed on the image plane. It is to be noted that Equation 2.3 holds only approximately. As shown in Figure 2.2, the exit pupil (represented by the function $P(k_x, k_y)$) limits the maximum diffraction angle. As a result, higher frequencies are cut-off, and the imaging system acts as a Low Pass filter (LPF). Therefore,

$$U_I(x', y') = \mathcal{F}\{P \odot U_\Theta\} \quad (2.4)$$

where, P is the optical transfer function.

2.1.2 Point Spread Function

Let a point source be kept at the origin, and the optical axis is made to pass through it. For slight deviations of the source from the optical axis, the image essentially does not undergo any change. The system is then known to be Linear Shift Invariant (LSI) system. The impulse response of an LSI system is the Point Spread Function (PSF). The Fourier transform of the PSF is the Optical Transfer function (OTF). The equivalent mathematical model is

$$U_I(x, y) = \iint_{-\infty}^{\infty} h(x - \tilde{x}, y - \tilde{y}) U_\Theta(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y} \quad (2.5)$$

$$= h(x, y) \otimes U_\Theta(x, y) \quad (2.6)$$

The intensity pattern on the image plane is the squared modulus of U_I .

At this point, a binary coherent OTF can be assumed,

$$P(k_x, k_y) = \begin{cases} 1, & \text{if } k_x^2 + k_y^2 \leq k_m^2 \\ 0, & \text{otherwise} \end{cases} \quad (2.7)$$

The Fourier Transform is of P is

$$h(\rho) = \frac{J_1(k_m \rho)}{k_m \rho} \quad (2.8)$$

where, $\rho = \sqrt{k_x^2 + k_y^2}$ and $k_m = \frac{2\pi r_m}{\lambda F}$. $J_1(\cdot)$ is the Bessel function of the first kind and r_m is the radius of the exit pupil.

2.1.3 Relation between NA and resolvable capacity

According to the Rayleigh criterion, the images of two point sources are resolvable if the maximas of their corresponding diffraction patterns on the image plane are at least separated by impulse function's first minima(root). The size of the PSF is the size of the maximum lobe, i.e. up to the first root of the diffraction pattern. The root in Equation 2.8 is at 1.22π . The radius of the first minima is approximately given by

$$\frac{r_m}{\lambda F} \rho = 0.61 \quad (2.9)$$

$$\rho = \frac{0.61 \lambda}{NA} \quad (2.10)$$

The cut-off frequency f_c is related to the NA

$$f_c = \frac{2NA}{\lambda} \quad (2.11)$$

The resolution is limited by diffraction due to the wave nature of light, and real optical systems can only approach this ideal.

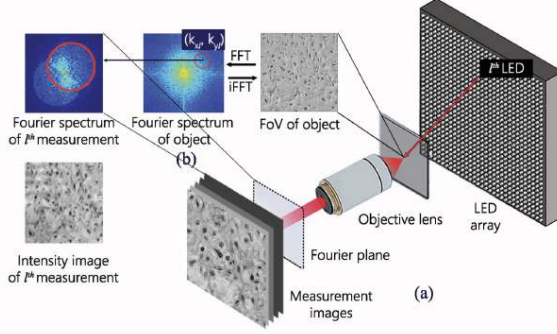


Figure 2.3: (a) Experimental Setup of FPM. The object is placed between a planar LED array and objective of the inverted microscope. Low resolution images corresponding to ℓ^{th} LED are captured. (b) Representation of the equivalent forward model, comprises of down-sampling the object spectrum by the system pupil function centred at $(k_{x\ell}, k_{y\ell})$ for the ℓ^{th} LED.

2.2 FORWARD IMAGING MODEL

For convenience, the object-field U_Θ is denoted by the object Θ . The mathematical model for the capture of multiple intensity images of the sample under different incident angles,

$$\phi_\ell(x, y) = h(x, y) \otimes (\Theta(x, y) e^{i(k_{x\ell}x + k_{y\ell}y)}). \quad (2.12)$$

In equation 2.12, Θ represents the complex amplitude of the object, ϕ_ℓ is an intermediate quantity that represents the output complex signal of the microscope system, $h(x, y)$ represents the coherent PSF in the spatial domain, and \otimes represents convolution operation. The term $e^{i(k_{x\ell}x + k_{y\ell}y)}$ represents the incident plane wave with a wave vector $(k_{x\ell}, k_{y\ell})$ for the ℓ^{th} LED. Equation. 2.12 can be written in Fourier domain as

$$\Phi_\ell(k_x, k_y) = P \odot O(k_x - k_{x\ell}, k_y - k_{y\ell}) \quad (2.13)$$

In Equation 2.13, O represents the object spectrum in the Fourier domain, Φ_ℓ represents the output spectrum of the microscope platform, and P is the equivalent representation of the microscopic system. It is to be noted that, the multiplication of $e^{i(k_{x\ell}x + k_{y\ell}y)}$ in the

spatial domain is equivalent to shifting the object spectrum by the amount of $(k_{x\ell}, k_{y\ell})$ in the Fourier domain. Therefore, a shifted spectrum is observed on the right hand side of Equation.2.13. Also, since the sample is assumed to be a 2D thin layer, the captured low-resolution images under different incident angles be uniquely mapped to different pass-bands in the Fourier domain.

Algorithm 1 Forward_computation(\mathbb{O}, K, P)

- 1: **Input:** Object-field \mathbb{O} , LED frequency measurement $K = [(k_{x\ell}, k_{y\ell})]_{\ell \in L}$, Pupil function P
 - 2: **Output:** Intensity measurement $[I]_{\ell \in L}$
 - 3: **Local:** Let L be the number of LEDs and O be the Fourier transform of \mathbb{O}
 - 4: $O \leftarrow \mathcal{F}\{\mathbb{O}\}$
 - 5: **for** $\ell = 1, 2, \dots, L$
 - 6: $\Phi_\ell \leftarrow P \odot O(k_x - k_{x\ell}, k_y - k_{y\ell})$
 - 7: $\phi_\ell \leftarrow \mathcal{F}^{-1}\{\Phi_\ell\}$
 - 8: $I_\ell \leftarrow |\phi_\ell|^2$
 - 9: **RETURN** $[I]_{\ell \in L}$
-

Algorithm 1 simulates the forward imaging process of FPM as defined by Zheng *et al.* (2013). The starting point is a high-resolution complex image as \mathbb{O} . From system specifications, the incident wave vector for the LED matrix is calculated. The object is assumed to be placed on the optical axis at $(0, 0)$. Lastly, the output images are produced by the low-pass filtering operation of Equation. 2.13. The the set of LR intensity images is denoted by $[I_\ell]_{\ell \in L}$. From Algorithm 1, it is observed that the most expensive steps are the FFT operation and its inverse. Therefore, the time-complexity of the forward computation algorithm depends primarily on the number of image pixels or the size of the FoV. For a large FoV, it is divided into overlapping patches for simulation of the LR images.

Algorithm 1 is a simplified approximation of the microscopic imaging system. In order to simulate the imaging system more accurately, non-linearities such as optical aberrations and various noise sources in the camera pipeline must be considered. We have considered Optical aberrations are simulated using a non-binary Pupil function; the knowledge

of Zernike coefficients (Lakshminarayanan and Fleck (2011)) for a typical microscope objective is required. For the simulation of the camera pipeline, the model proposed by Hasinoff *et al.* (2010) has been used as a reference in the following chapters.

CHAPTER 3

PRELIMINARY STUDY OF PHASE RETRIEVAL ALGORITHMS FOR FPM

3.1 THE PROBLEM OF PHASE RETRIEVAL

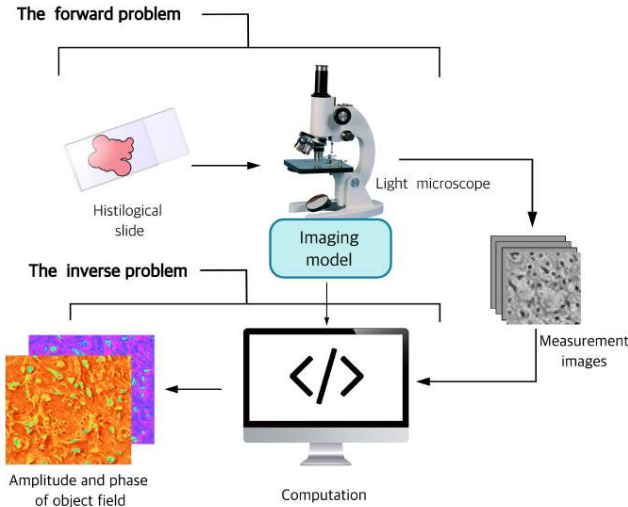


Figure 3.1: The forward and inverse problems.

The forward computation model has been introduced in the previous chapter. As shown in Figure 2.3, a known sample is imaged by the microscope and the intensity patterns are captured by a sensor. As a result, the phase information is lost. In this chapter, the methods for algorithmic phase retrieval are discussed. Figure 3.1 shows that given the model parameters, the intensity measurements can be estimated using the knowledge of the imaging model. The inverse problem of algorithmic phase retrieval, is the recovery of the phase given the measurement of the magnitude of the optical far-field and a mathematical representation of the imaging model.

While the phase retrieval problem appears in many different setups, the FPM problem is taken up in this work. LR intensity images $[I]_{\ell \in L}$ are captured using a low NA objective of the FPM experimental setup (Figure 2.3). The following objective function is to be minimized,

$$\min_{\mathbb{C}} \sum_{\ell=1}^L \|I_{\ell} - |A_{\ell}\{\mathbb{C}\}|^2\|_2^2 \quad (3.1)$$

where $A\{\cdot\}$ represents the forward model as discussed in Section 2.2. Apart from the challenges imposed due to the ill-posedness and non-convexity of phase-retrieval problems, FPM reconstruction suffers from several other challenges:

1. From the forward imaging model, it is known that intensity images are the band-pass filtered version of the original signal where the cut-off depends on the illumination angle. This necessitates high-quality sensors, especially for large angle illumination, where signal strength is low. In the absence of high-quality sensors, the problem of FPM reconstruction becomes even more challenging.
2. FPM reconstruction algorithms are sensitive to inexact system calibration and aberrations.
3. FPM reconstruction is essentially a complex-valued non-convex inverse problem, and existing learning-based inverse algorithms cannot be extended to the FPM system trivially.

In order to solve Equation 3.3, iterative methods are employed as discussed in the later sections.

3.2 RELATED WORKS

The earliest works on ptychographic phase retrieval are based on the alternating projection algorithm and date back to the pioneering work of Gerchberg and Saxton (Gerchberg (1972)). They attempted to minimize the least-squares error and iteratively imposed the

constraints of the image plane and the diffraction plane to obtain convergence. Later on, the application of variants of the Wirtinger Flow algorithm, such as truncated Wirtinger flow and reshaped Wirtinger flow by Zhang et al. Zhang *et al.* (2017) has also been exploited. The initial papers published on FP phase retrieval algorithms derive ideas from the literature, as mentioned earlier.

A description of the FPM setup is given by Ou et al. Ou *et al.* (2013), who successfully performed whole slide high-resolution phase and amplitude imaging without translation of stage. It is a first-order technique based on the alternate projections algorithm. Since then, numerous works on FPM have come up. Song et al. Song *et al.* (2019) have proposed modifications, including algorithms to overcome system aberration and sample motion correction Bian *et al.* (2016). The work by Tian et al. Tian *et al.* (2014) extended the AP algorithm that allows multiplexing of the LED measurements. Wirtinger flow-based methods are shown to perform well under low SNR Bian *et al.* (2015). The Wirtinger flow algorithm is also extended to the multiplexed scheme by Bostan et al. Bostan *et al.* (2018). There have been several hardware modifications to the experimental setup described in Figure 2.3, which are beyond the context of this article but have been researched by Zheng et al. Zheng Guoan (2021).

Recently, deep learning techniques have also been explored to solve the FP problem. Jiang et al. Jiang *et al.* (2018) show a novel approach by treating the object as a learnable network parameter. However, as it requires optimization for each patch, it requires a long time to reconstruct the entire FoV. Kappeler *et al.* (2017) have performed high-resolution amplitude recovery using a CNN architecture. Nguyen *et al.* (2018) have performed time-lapse high-resolution phase imaging via adversarial training of a U-Net architecture Mirza and Osindero (2014). The authors have successfully performed phase imaging only using a limited number of Dark-field images. Using adversarial training, the authors performed HR phase imaging without the aid of the information contained in the darkfield images corresponding to the peripheral LEDs. Their training

scheme involves training (transfer learning in the case of a new biological type of cells) of U-Net on LR measurements at 0 min and predicts the phase in the subsequent frames of the videos of dynamic cell samples. FP phase reconstruction under various overlap conditions between adjacent low-resolution images in the frequency domain has been studied by Boominathan *et al.* (2018). However, the performance of their model under a low SNR scenario has not been demonstrated. Kellman *et al.* (2019) have designed a Physics-based Neural Network (PbNN) that learns the multiplexed design by optimizing for the weights of the LEDs in the multiplexed design formed by digitally combining single-LED measurements. The PbNN is memory intensive since it unrolls 100 iterations of gradient descent with a step size of 0.5.

In summary, iterative algorithms have been the standard for FP reconstruction, but with certain limitations such as sensitivity to noise and calibration errors and higher time-complexity. In contrast, recent approaches involving neural network models can be optimized to overcome these challenges. The obvious advantage of direct inference using neural network is the lower inference time. A comparison of the iterative and direct inference methods is provided in Table 3.1.

Direct methods	Iterative methods
Uses a learnt map to reconstruct in one forward propagation comprising of finite number of steps.	Iteration steps depend on a given error tolerance.
The computational burden of one forward propagation step depends upon the complexity of the model architecture.	Iterative methods are designed for lower computational burden per iteration.
Generally requires a large amount of data for training.	No prior training data is required.
Network parameters are learnt from data.	Parameter tuning is a non-trivial task.

Table 3.1: Comparison of iterative and neural-network based approaches.

3.3 ITERATIVE METHODS FOR FOURIER PTYCHOGRAPHIC PHASE RETRIEVAL

The case of phase recovery from intensity measurements only is taken up in this work. Mathematically, the phase retrieval problem while using structured illumination for intensity measurement is as follows (Zheng *et al.* (2013)):

$$\text{Find } \mathbb{O}, \quad \text{subject to} \quad I_\ell = |\mathcal{F}^{-1}\{P_\ell \odot \mathcal{F}\{\mathbb{O}(\mathbf{k} - \mathbf{k}_\ell)\}\}|^2, \text{ for } \ell = 1, 2, \dots, L \quad (3.2)$$

where \mathcal{F} is the 2D spatial Fourier transform and \mathcal{F}^{-1} . $P_\ell \in \mathbb{C}^N$ is the pupil function for $\mathbf{k}_\ell = (k_{x\ell}, k_{y\ell})$, the unique spatial frequency corresponding to the ℓ^{th} illumination source. It is assumed that \mathbb{O} is an optically thin sample, i.e. its transmittance is close to unity. Classical retrieval algorithms seek to solve the following minimization problem,

$$\min_{\mathbb{O}} \sum_{\ell=1}^L \|I_\ell - |A_\ell\{\mathbb{O}\}|^2\|_2^2 \quad (3.3)$$

where $A\{\cdot\} \triangleq \mathcal{F}^{-1} \circ P_\ell \odot \mathcal{F}\{\cdot\}$. The objective is a non-convex, real-valued function of a complex-valued object field. Conventional approaches like gradient descent will converge to a stationary point. As discussed by Osherovich (2012), convex relaxation methods such as restriction of the search region can reduce it to a convex objective, and Wirtinger calculus can solve for a global minima. In the following subsections, iterative algorithms are discussed in further detail.

3.3.1 Alternate Projection based method for Fourier Ptychographic Phase Retrieval

A widely implemented algorithm for phase retrieval was introduced by Gerchberg and Saxton (G-S) (Gerchberg (1972)) and is based on the principle of Alternate Projections (AP). The algorithm was initially developed to infer the phase distribution of the electron beam given its intensity distributions. Since then, it has found various applications in optical imaging. Fienup (1982) showed that the G-S algorithm is equivalent to the steepest descent method.

The G-S algorithm was extended by Zheng *et al.* (2013) to perform FPM phase retrieval.

Algorithm 2 Alternating_projection($[I]_{\ell \in L}, K, P, \mu$)

```

1: Input: Low-resolution measurements  $[I]_{\ell \in L}$ , LED frequency measurement  $K = [(k_{x\ell}, k_{y\ell})]_{\ell \in L}$ , Pupil function  $P$ , learning rate  $\mu$ 
2: Output: Reconstructed object field  $\mathbb{O}$ . Initialize  $\mathbb{O}^0 = I_0$ 
3: Local: tolerance  $\epsilon$ 
4: while  $\|[I]_{\ell \in L} - [\hat{I}]_{\ell \in L}\|_2^2 \leq \epsilon$ 
5:   for  $\ell \in L$ 
6:      $\Phi_\ell \leftarrow A_\ell\{O(k_x - k_{x\ell}, k_y - k_{y\ell})\}$ 
7:      $\phi_\ell \leftarrow \mathcal{F}^{-1}\{\Phi_\ell\}$ 
8:      $\hat{I}_\ell \leftarrow |\phi_\ell|^2$ 
9:      $\phi'_\ell \leftarrow \sqrt{\hat{I}_\ell} \frac{\phi_\ell}{|\phi_\ell|}$ 
10:     $\Phi'_\ell \leftarrow \mathcal{F}\{\phi'_\ell\}$ 
11:     $O^{i+1}(k_x - k_{x\ell}, k_y - k_{y\ell}) \leftarrow O^i(k_x - k_{x\ell}, k_y - k_{y\ell}) - \mu \cdot P \odot (\Phi_\ell - \Phi'_\ell)$ 
12: RETURN  $\mathbb{O}$ 

```

3.3.2 Gradient descent based method for Fourier Ptychographic Phase Retrieval

Gradient descent on the classical phase retrieval problem converges sub-optimally. However, in the case of FPM, given enough measurements and sufficient overlap between measurements in the frequency domain, it imposes additional constraints and is possible to design a feasible algorithm. It seeks to directly solve the minimization problem in Equation 3.3. The naive gradient descent approach (i^{th} iteration, over objective function f),

1. Find a descent direction $(\nabla f)_i$.
2. Along that direction find the learning rate μ_i that sufficiently decreases the objective function value.
3. Update as $\mathbb{O}_{i+1} = \mathbb{O}_i + \mu_i(\nabla f)_i$

Wirtinger Gradients

The objective in Equation 3.3 is a function of complex-valued object field. A complex-valued function $f(z)$ of a complex-valued variable $z = x + jy$,

$$f(z) = u(x, y) + jv(x, y), \quad (3.4)$$

The following pair of derivatives for the function $f(z)$ are defined with respect to the pair of complex variables z, z^* (Tanaka (2013)),

$$\frac{\partial f}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right) \quad (3.5)$$

$$\frac{\partial f}{\partial z^*} = \frac{1}{2} \left(\frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right). \quad (3.6)$$

For holomorphic $f(z)$, the *Cauchy-Reimann* equations are satisfied,

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y} \quad (3.7)$$

$$\frac{\partial u(x, y)}{\partial y} = -\frac{\partial v(x, y)}{\partial x}. \quad (3.8)$$

For holomorphic f , $\frac{\partial f}{\partial z^*}$ reduces to zero, using the Cauchy-Riemann equations,

$$\frac{\partial f}{\partial z} = \frac{1}{2} \left(\frac{\partial u}{\partial x} + j \frac{\partial v}{\partial y} \right) \quad (3.9)$$

$$\frac{\partial f}{\partial z^*} = 0. \quad (3.10)$$

This is equivalent to the real derivative of the complex-valued function. The Wirtinger gradients are defined for performing gradient descent using complex variables. If $f(z)$ is the objective function, it can be optimized using either of $\frac{\partial}{\partial z}$ or $\frac{\partial}{\partial z^*}$ operator. If the latter is preferred, the Wirtinger gradient operator will be defined as,

$$\nabla_z f = \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \quad (3.11)$$

$$= 2 \frac{\partial}{\partial z^*} \quad (3.12)$$

In the case of objective function in Equation 3.3, it is noted to be non-holomorphic and real-valued. The gradients are defined in the following Section using the operator in Equation 3.12.

Wirtinger Gradients for Fourier Ptychographic Phase Retrieval

The naive Wirtinger flow algorithm as proposed by Bian *et al.* (2015), begins with Equation 3.3.

$$f(\mathbb{O}, I_\ell) = \min_{\mathbb{O}} \sum_{\ell=1}^L \|I_\ell - |A_\ell\{\mathbb{O}\}|^2\|_2^2 \quad (3.13)$$

$$= \sum_{\ell=1}^L \|I_\ell - A_\ell\{\mathbb{O}\} \odot \overline{A_\ell\{\mathbb{O}^i\}}\|_2^2 \quad (3.14)$$

where $A\{\cdot\} \triangleq \mathcal{F}^{-1} \circ P_\ell \odot \mathcal{F}\{\cdot\}$. The i^{th} gradient update rule,

$$\mathbb{O}^{i+1} = \mathbb{O}^i + \eta \nabla_{\mathbb{O}} f^i \quad (3.15)$$

where η is the step-size and,

$$\nabla_{\mathbb{O}} f^i = \frac{1}{L} \sum_{\ell=1}^L A_\ell^H \{(|A_\ell\{\mathbb{O}^i\}|^2 - I_\ell) \odot A_\ell\{\mathbb{O}^i\}\} \quad (3.16)$$

Here, $A_\ell^H\{\cdot\} \triangleq \mathcal{F}^{-1} \circ P_\ell^H \odot \mathcal{F}\{\cdot\}$.

$$\text{ZEROPAD}_{N,\ell} \{\phi(k_x, k_y)\} = \begin{cases} \phi(k_x, k_y), & |k_x - k_{x\ell}|, |k_y - k_{y\ell}| \leq N/2 \\ 0, & \text{otherwise} \end{cases}$$

The pseudo-code for the naive approach is shown in Algorithm 3. The naive Wirtinger flow algorithm suffers from slow convergence, and the choice of the step size is crucial. Bian *et al.* (2015) have suggested methods to optimize the step size. Our initial experiments also involve modifications of the naive approach. Algorithm 4 is the Nesterov Accelerated version. In Algorithm 5, the first-order Wirtinger flow algorithm is modified by sequentially updating an auxiliary variable corresponding to every LED while computing the gradients. It is observed that the approach reduces the number of iteration stages and avoids phase wrapping(refer to Appendix ??). An exponentially decaying step size is used for the experiments.

Algorithm 3 Gradient descent($[I]_{\ell \in L}, K, P, \mu$)

- 1: **Input:** low-resolution measurements $[I]_{\ell \in L}$, LED frequency measurement $K = [(k_{x\ell}, k_{y\ell})]_{\ell \in L}$, Pupil function P , learning rate μ
 - 2: **Output:** Reconstructed object field \mathbb{O} . Initialize $\mathbb{O}^0 = I_0$.
 - 3: **Local:** Let L be the number of input measurements. Initialize $[G]_{\ell \in L}$ as zero. tolerance ϵ .
 - 4: **while** $\|[I]_{\ell \in L} - [\hat{I}]_{\ell \in L}\|_2^2 \leq \epsilon$
 - 5: **for** $\ell = 1, 2, \dots, L$
 - 6: $\Phi_\ell \leftarrow A_\ell \{O(k_{x\ell}, k_{y\ell})\}$
 - 7: $\phi_\ell \leftarrow \mathcal{F}^{-1}\{\Phi_\ell\}$
 - 8: $\hat{\mathbf{I}}_\ell \leftarrow |\mathcal{F}^{-1}\{\Phi_\ell\}|^2$
 - 9: $G_\ell \leftarrow A_\ell^{-1}\{(\hat{\mathbf{I}}_\ell - I_\ell) \odot \phi_\ell\}$
 - 10: $\mathbb{O}_{i+1} \leftarrow \mathbb{O}_i - \mu \cdot \frac{1}{L} \sum_{\ell=1}^L G_\ell$
 - 11: **RETURN** \mathbb{O}
-

Algorithm 4 Accelerated gradient descent($\Psi, \mathbf{I}, K, P, \mu$)

- 1: **Input:** low-resolution measurements $[I]_{\ell \in L}$, LED frequency measurement $K = [(k_{x\ell}, k_{y\ell})]_{\ell \in L}$, Pupil function P , learning rate μ , acceleration η .
 - 2: **Output:** Reconstructed object field \mathbb{O} . Initialize $\mathbb{O}^0 = I_0$.
 - 3: **Local:** Let L be the number of input measurements. Initialize $\Phi_1 = \mathbb{O}$, $[G]_{\ell \in L}$ as zero. tolerance ϵ . Initialize $\Phi_1 = \mathbb{O}$.
 - 4: **while** $\|[I]_{\ell \in L} - [\hat{I}]_{\ell \in L}\|_2^2 \leq \epsilon$
 - 5: **for** $\ell = 1, 2, \dots, L$
 - 6: $\Phi_\ell \leftarrow A_\ell \{\mathbb{O}(k_{x\ell}, k_{y\ell})\}$
 - 7: $\phi_\ell \leftarrow \mathcal{F}^{-1}\{\Phi_\ell\}$
 - 8: $\hat{\mathbf{I}}_\ell \leftarrow |\mathcal{F}^{-1}\{\Phi_\ell\}|^2$
 - 9: $G[\ell] \leftarrow A_\ell^{-1}\{(\hat{\mathbf{I}}_\ell - I_\ell) \odot \phi_\ell\}$
 - 10: $v_{i+1} \leftarrow (1 - \eta) \cdot v_i - \eta \cdot \frac{1}{L} \sum_{\ell=1}^L \mathbf{G}[\ell]$
 - 11: $\mathbb{O}_{i+1} \leftarrow \mathbb{O}_i - \mu \cdot \frac{1}{L} \sum_{\ell=1}^L \mathbf{G}[\ell]$
 - 12: **RETURN** \mathbb{O}
-

Algorithm 5 Stochastic gradient descent($\Psi, \mathbf{I}, K, P, \mu, \eta$)

- 1: **Input:** low-resolution measurements $[I]_{\ell \in L}$, LED frequency measurement $K = [(k_{x\ell}, k_{y\ell})]_{\ell \in L}$, Pupil function P , learning rates μ and η
 - 2: **Output:** Gradient $\mathbf{g} = \{g_\ell\}$
 - 3: **Local:** Let L be the number of input measurements. Initialize $\Phi_1 = \mathbb{O}$, $[G]_{\ell \in L}$ as zero. tolerance ϵ . Initialize $\Phi_1 = \mathbb{O}$.
 - 4: **while** $\|[I]_{\ell \in L} - [\hat{I}]_{\ell \in L}\|_2^2 \leq \epsilon$
 - 5: **for** $\ell = 1, 2, \dots, L$
 - 6: $\Phi_\ell \leftarrow A_\ell \{\mathbb{O}(k_{x\ell}, k_{y\ell})\}$
 - 7: $\phi_\ell \leftarrow \mathcal{F}^{-1}\{\Phi_\ell\}$
 - 8: $\hat{\mathbf{I}}_\ell \leftarrow |\mathcal{F}^{-1}\{\Phi_\ell\}|^2$
 - 9: $g_\ell \leftarrow A_\ell^{-1}\{(\hat{\mathbf{I}}_\ell - I_\ell) \odot \phi_\ell\}$
 - 10: $\Phi_{\ell+1} \leftarrow \Phi_\ell - \eta \cdot g_\ell$
 - 11: $\mathbf{G}[\ell] \leftarrow g_\ell$
 - 12: $\mathbb{O}_{i+1} \leftarrow \mathbb{O}_i - \mu \cdot \frac{1}{L} \sum_{\ell=1}^L \mathbf{G}[\ell]$
 - 13: **RETURN** \mathbb{O}
-

A comparison of update methods of the Wirtinger Flow algorithm (Algorithm 3, 4 and 5) is shown in Figure 3.3. While phase wrapping artifacts are frequently encountered for phase retrieval from intensity and has been studied by Ou *et al.* (2014), the experiment shows that phase wrapping is avoided by tuning the learning rate and modifying the gradient update method. The naive Wirtinger flow and its accelerated version display phase wrapping even at lower learning rate of $\mu \approx 0.005$ and suffers from slow convergence. The Stochastic update method in Algorithm 5 performs update for every LED in each iteration. This prevents the phase wrapping artifacts and achieves faster convergence.

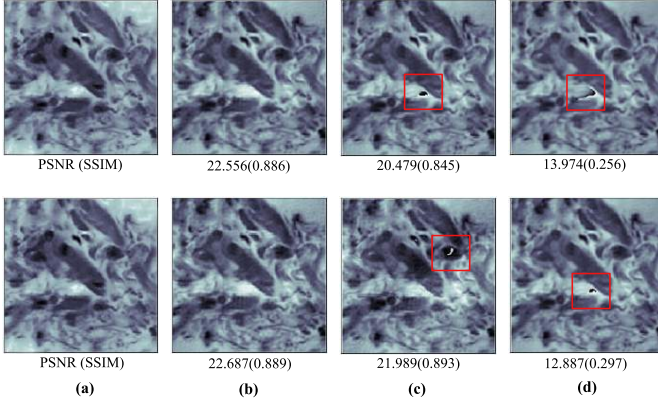


Figure 3.3: FPM phase reconstruction using Wirtinger update method. Artifacts at 50 and 100 iterations are highlighted in red. (a) Ground truth (b) Wirtinger Flow (c) Wirtinger Flow with NAG, and (d) Stochastic update.

3.4 DATA-DRIVEN METHODS FOR FOURIER PTYCHOGRAPHIC PHASE RETRIEVAL

Neural networks are powerful in approximating unknown functions due to their high degree of freedom, owing to many parameters. In this section, a deep learning framework for complex-valued operations is described.

3.4.1 Complex-valued Neural Network for FPM Phase Retrieval

Complex-valued neuronal units are shown to have a richer representative capacity and are biologically motivated (Trabelsi *et al.* (2018)). Recent deep nets for FPM phase retrieval reconstruct amplitude and phase independently (Nguyen *et al.* (2018), Boominathan *et al.* (2018)). Although the focus is to retrieve the phase information, it is more intuitive to reconstruct the object field and extract the phase information. Furthermore, our motive to integrate the iterative approach, which operates in the complex domain, makes it essential to define the Neural network parameters in the complex domain. A complex-valued framework also makes it convenient to implement routines in the frequency domain. The following sections describe the implementation of Complex-valued Neural Networks (CVNNs). Conventional derivation of optimization algorithms such as the back-propagation algorithm employs real-valued derivatives. It can be reformulated in the complex domain by defining the Wirtinger derivatives. Such reformulation is necessary since non-holomorphic functions often occur in CVNNs, for which the definitions using real-valued derivatives do not hold. There can be real-valued objective functions and non-holomorphic activation functions. However, the derivatives of such functions can be defined using Wirtinger calculus, enabling learning algorithms such as back-propagation in the complex domain. It generalizes the notion of complex-valued derivatives, and the holomorphic function becomes a particular case only.

The conventional real-valued network architectures U-Net and Res-Net are reformulated in the complex domain in the following sections. The primary idea is derived from the work by Trabelsi *et al.* (2018). Operating in the complex domain gives us the advantage of easy implementation of FFT-based routines.

3.4.2 Network architecture

The standard neural network architectures, the UNet and ResNet architecture are compared using the complex-valued building blocks (refer Figures 3.4 and 3.5).

1. **Complex-convolution:** An equivalent 2D complex-valued convolution is

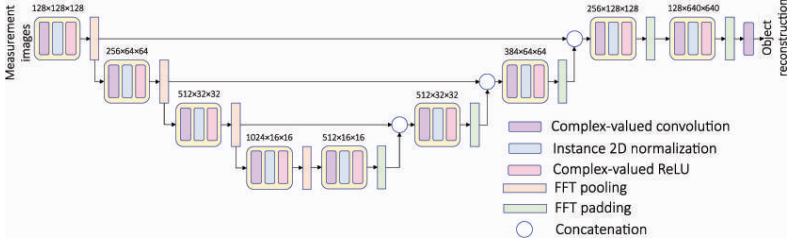


Figure 3.4: FP-UNet architecture

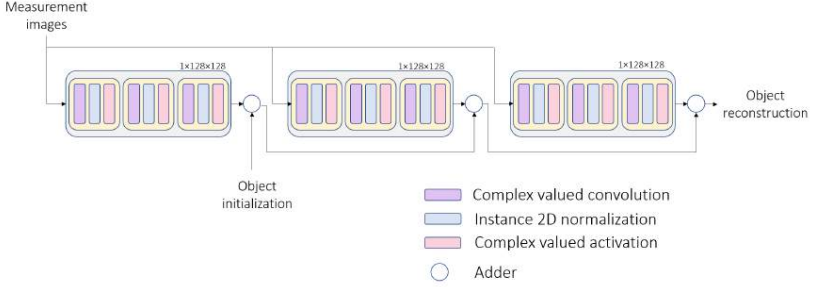


Figure 3.5: FP-ResNet architecture

performed using real-valued CNN blocks as follows: let there be M input channels and N output channels for a given CNN layer. Then, complex-valued filter matrix weight for m^{th} input channel and n^{th} output channel is $W_{mn} = A_{mn} + jB_{mn}$ that convolves with a 2D complex input $\mathbf{z}_m = \mathbf{x}_m + j\mathbf{y}_m$. Here, A_{mn} and B_{mn} are real-valued kernels, and \mathbf{x}_m and \mathbf{y}_m are also real.

The matrix notation for complex-convolution is,

$$\begin{bmatrix} \mathcal{R}(W_{mn} * \mathbf{z}) \\ \mathcal{I}(W_{mn} * \mathbf{z}) \end{bmatrix} = \begin{bmatrix} A_{mn} & -B_{mn} \\ B_{mn} & A_{mn} \end{bmatrix} * \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad (3.17)$$

The marginalized form the complex-convolution,

$$W_{mn} * \mathbf{z}_m = A_{mn} * \mathbf{x}_m + jB_{mn} * \mathbf{y}_m. \quad (3.18)$$

The filter weights are initialized as a uniform distribution.

2. **Activation function:** An amplitude-phase type non-linear activation function acts on the normalised outputs of the convolution layer. The complex-valued form of the Tanh activation is defined as,

$$\mathbb{C}\text{Tanh}(\mathbf{z}_k) = \text{Tanh}(|\mathbf{z}_k|)e^{j\theta_{z_k}}. \quad (3.19)$$

Complex-valued activation functions can also be implemented in their marginalised forms in terms of real-valued operations acting separately on the real and imaginary parts,

$$\mathbb{C}\text{Tanh}(\mathbf{z}_i) = \text{Tanh}(\mathcal{R}(\mathbf{z}_i)) + j\text{Tanh}(\mathcal{I}(\mathbf{z}_i)). \quad (3.20)$$

3. **FFT Pooling:** Rippel *et al.* (2015) proposed spectral pooling which performs dimensionality reduction by truncating the representation in the frequency domain. This approach preserves considerably more information per parameter than other pooling strategies and enables flexibility in the choice of pooling output dimensionality.

Algorithm 6 Forward FFT pooling(\mathbf{z}_k, W_o, H_o)

- 1: **Input:** $\mathbf{z}_k \in \mathbb{C}^{H_i \times W_i}$, output size $H_o \times W_o$
 - 2: **Output:** Pooled map $\hat{\mathbf{z}}_k \in \mathbb{C}^{H_o \times W_o}$
 - 3: $\mathbf{y} \leftarrow \mathcal{F}\{\mathbf{z}_k\}$
 - 4: $\mathbf{y} \leftarrow \text{CROP SPECTRUM}(\mathbf{y}, W_o, H_o)$
 - 5: $\hat{\mathbf{z}}_k \leftarrow \mathcal{F}^{-1}(\mathbf{y})$
 - 6: **RETURN** $\hat{\mathbf{z}}_k$
-

Algorithm 7 Backpropagation FFT pooling($\mathbf{g}_{\mathbf{z}_k}$)

- 1: **Input:** Gradient with respect to output, $\mathbf{g}_{\mathbf{z}_k} = \frac{\partial f}{\partial \mathbf{z}_k} \in \mathbb{C}^{H_o \times W_o}$
 - 2: **Output:** Gradient with respect to input, $\mathbf{g}_{\mathbf{z}_k} = \frac{\partial f}{\partial \mathbf{z}_k} \in \mathbb{C}^{H_i \times W_i}$
 - 3: $\mathbf{y} \leftarrow \mathcal{F}\{\mathbf{g}_{\mathbf{z}_k}\}$
 - 4: $\mathbf{y} \leftarrow \text{ZEROPAD}(\mathbf{y}, W_i, H_i)$
 - 5: $\hat{\mathbf{z}} \leftarrow \mathcal{F}^{-1}(\mathbf{y})$
 - 6: **RETURN** $\hat{\mathbf{z}}$
-

4. **FFT Upsampling:** The frequency domain upsampling is similar to zero-padding.

The forward and backward propagation routines can be written by referring to Algorithm 6 and 7.

CHAPTER 4

LEARNED WIRTINGER GRADIENTS NETWORK (LWGNET) FOR FPM PHASE RETRIEVAL

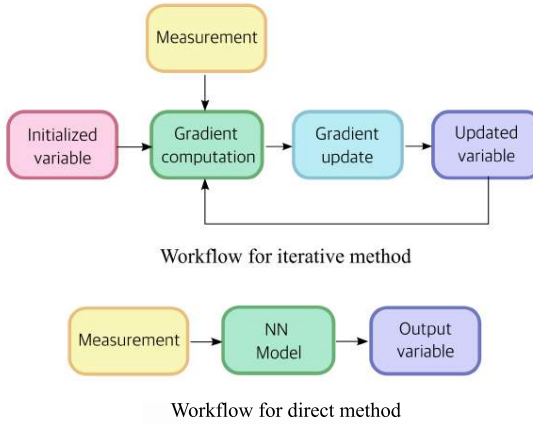


Figure 4.1: The iterative and the direct approaches for solving inverse problems.

The previous chapter discusses the iterative and direct methods for FPM phase reconstruction. In the first case (Figure 4.1), a physical model uses the experimental measurement to compute the inverse solution incrementally. In the second case, given the measurements, the inverse mapping is approximated using a neural network using supervised learning. The inference is performed on unseen data. In this chapter, a hybrid model of the data-driven and the model-based approaches, Learned Wirtinger Gradients Network (LWGNet), is proposed. The network is designed to address the challenge of phase recovery from poor quality intensity images, as mentioned in Section 1.3. First,

the methodology for the sequential illumination scheme is described, and then a simple modification is used in the existing reconstruction pipeline to extend it to the multiplexed illumination scheme.

4.1 PROPOSED NETWORK ARCHITECTURE

Starting with the Wirtinger flow algorithm, described in Section 3.3.2. Minimizing Equation 3.3 with respect to \mathbb{O} using gradient descent leads to the following i^{th} step update,

$$\mathbb{O}^{i+1} = \mathbb{O}^i - \eta \frac{1}{L} \sum_{\ell=1}^L \nabla_{\mathbb{O}} f_{\ell}^i, \quad (4.1)$$

where η is the step-size and

$$\nabla_{\mathbb{O}} f_{\ell}^i = A_{\ell}^H \{ (|A_{\ell} \{\mathbb{O}^i\}|^2 - I_{\ell}) \odot A_{\ell} \{\mathbb{O}^i\} \}. \quad (4.2)$$

A naive way to unroll the above gradient descent Wirtinger flow algorithm could be to design a neural network with each stage performing the gradient update step of Eq.(4.1). However, it has been found that such a naive unrolling converges slowly and fails to provide meaningful results under limited GPU memory constraints. To overcome this, a K-stage unrolled network is designed with the i^{th} stage performing the following operation

$$\mathbb{O}^{i+1} = \mathbb{O}^i + \Psi^i \{ [\mathbf{g}_{\ell}]_{\ell \in L}^i \}. \quad (4.3)$$

Here,

$$\mathbf{g}_{\ell}^i = A_{\ell}^H \{ (|A_{\ell} \{\phi_{\ell-1}^i\}|^2 - I_{\ell}) \odot A_{\ell} \{\phi_{\ell-1}^i\} \}, \quad (4.4)$$

$$\phi_{\ell}^i = \phi_{\ell-1}^i - \eta \nabla_{\phi} f_{\ell-1}^i, \quad (4.5)$$

$\phi_0^i = O^i$, $\phi_1^0 = O^0 = I_0$, $[\mathbf{g}_{\ell}]_{\ell \in L}^i$ is a stack of \mathbf{g}_{ℓ}^i and $\Psi\{\cdot\}$ is a learned complex neural network. \mathbf{g}_{ℓ}^i in the above equations can be interpreted as stochastic gradients

corresponding to each illumination source. $\Psi\{.\}$ then learns to non-linearly combine these stochastic gradients (refer to Algorithm 3) and update the object field as shown in Equation (4.6). Experimentally, it has been found that such an unrolled network converged faster with fewer K than naively unrolled Wirtinger flow. $\Psi\{.\}$ consists of 3×3 complex convolutions followed by $\mathbb{C}\text{Tanh}$ non-linearity, 2D instance norm, and a fully connected layer. The learned non-linear $\Psi\{.\}$ helps combine the stochastic gradients more effectively, especially for peripheral darkfield images, which are typically degraded due to lower bit depth and noise. Each stage is equivalent to a gradient descent iteration

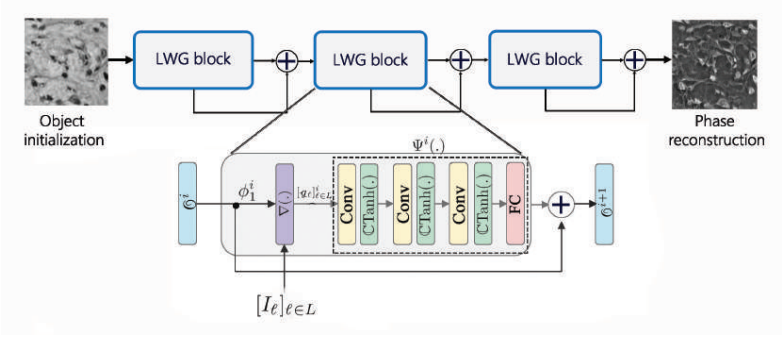


Figure 4.2: The model architecture of LWGNet comprising of three stages of the LWG update block. The inputs are the object and the pupil initialization and measurement images for every stage. LWG block includes a Gradient computation block followed by three stages of Conv+ $\mathbb{C}\text{Tanh}$ layers followed by a Fully Connected(FC) layer. The adder block updates the input object field using the modified gradients.

in the unrolled network. Fig 4.2 shows three such blocks similar to a residual network block and comprises three sub-blocks that implement the steps described above. The initial gradient estimates are the residues input to a complex convolution layer, followed by a non-linear activation function and linear layer. The residual addition is analogous to the gradient update step. A K -stage unrolled network is designed with the i^{th} stage

performing the following operation

$$\mathbb{O}^{i+1} = \mathbb{O}^i + \Psi^i\{[\mathcal{G}_\ell]^i_{\ell \in L}\}. \quad (4.6)$$

$\Psi\{\cdot\}$ uses complex as described in Section 3.4.2. In Figure 4.2, the Conv layer is followed by an 2D instance normalization layer. The amplitude-phase type $\mathbb{C}\text{Tanh}$ activation function acts on the normalised outputs as described in Section 3.4.2. Finally, after three stages of convolution and non-linear activation blocks, a Fully Connected(FC) block that acts on the channel dimension is used.

Layer	Input	Output	Groups
$\nabla(\cdot)$	225	225	-
Conv	225	128	1
Conv	128	64	1
Conv	64	32	1
FC	32	1	-

Table 4.1: Channel dimension of $\Psi_i\{\cdot\}$ for sequential scheme.

4.1.1 Loss Function

Let \mathbb{O} be the ground truth object and $\hat{\mathbb{O}}$ be the reconstruction. The following weighted combination of loss functions is used to optimize our network,

$$\mathcal{L} = \lambda_1 \mathcal{L}_{MSE} + \lambda_2 \mathcal{L}_{FMAE} + \lambda_3 \mathcal{L}_{VGG} \quad (4.7)$$

where,

$$\mathcal{L}_{MSE} = \|\mathcal{L}(\hat{\mathbb{O}}) - \mathcal{L}(\mathbb{O})\|_2^2 + \||\hat{\mathbb{O}}| - |\mathbb{O}|\|_2^2 \quad (4.8)$$

$$\mathcal{L}_{FMAE} = \||\mathcal{F}(\hat{\mathbb{O}})| - |\mathcal{F}(\mathbb{O})|\|_1 \quad (4.9)$$

$$\mathcal{L}_{VGG/i,j} = \beta_1 \|\psi_{i,j}(|\hat{\mathbb{O}}|) - \psi_{i,j}(|\mathbb{O}|)\|_2^2 + \beta_2 \|\psi_{i,j}(\mathcal{L}\hat{\mathbb{O}}) - \psi_{i,j}(\mathcal{L}\mathbb{O})\|_2^2 \quad (4.10)$$

$\|\cdot\|_p$ represents the p -norm. Eq.(4.8) defines the pixel-wise Mean Squared Error (MSE) loss over amplitude and phase components. The VGG loss function is a perceptual loss function, as defined by Ledig *et al.* (2017). It minimizes the MSE from the output of the feature maps of the pre-trained VGG-19 network. $\psi_{i,j}(\cdot)$ is the feature map obtained by the j -th convolution (after activation) before the i -th max-pooling layer in the network, which is considered given. Here, the ReLU output is used, $\psi_{2,2}$ and $\psi_{4,3}$ the VGG-19 network. The output feature maps corresponding to the reconstruction and ground truth amplitudes are compared. Eq.(4.9) minimizes the L_1 -norm of the magnitude Fourier spectrum between reconstruction and ground truth.

4.2 MULTIPLEXED ILLUMINATION

The multiplexed illumination scheme reduces the data acquisition overhead and is more desirable. Illuminating multiple LEDs at once reduces the number of images. Our multiplexing scheme uses a set of five LEDs for each image. The number of captured images is reduced from 225 images to 45 images, each LED being used only once. Thus, the possible reduction in total acquisition time for the entire FoV is 25 times.

4.2.1 Network architecture

The proposed neural network architecture is kept the same, with the convolution filters split into groups. A description of optimizable parameters $\Psi_i\{\cdot\}$ in each LWG block is given in Table 4.2. The measurement images replicated by a factor of 5 to obtain a set of 225 images serve as input to the network.

Layer	Input	Output	Groups
$\nabla(\cdot)$	225	225	-
Conv	225	135	45
Conv	135	90	45
Conv	90	45	1
FC	45	1	-

Table 4.2: Channel dimensions of $\Psi_i\{\cdot\}$ for multiplexed scheme.

4.2.2 Multiplexing scheme

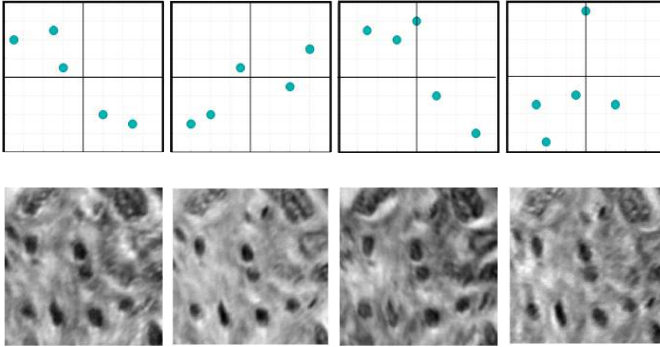


Figure 4.3: LED positions on a 15×15 grid (top) and Sample LR images of Osteosarcoma slide (bottom).

A multiplexing scheme is designed to balance the LED positions in every set of multiplexed images across the Fourier spectrum. The LED positions are sorted according to their corresponding NA value and split into five groups. During capture, an LED is chosen randomly and exhaustively from each group. Figure 4.3 plots the LED positions which shows an even distribution of central and peripheral LEDs.

CHAPTER 5

EXPERIMENTS AND RESULTS

5.1 EXPERIMENTAL SETUP

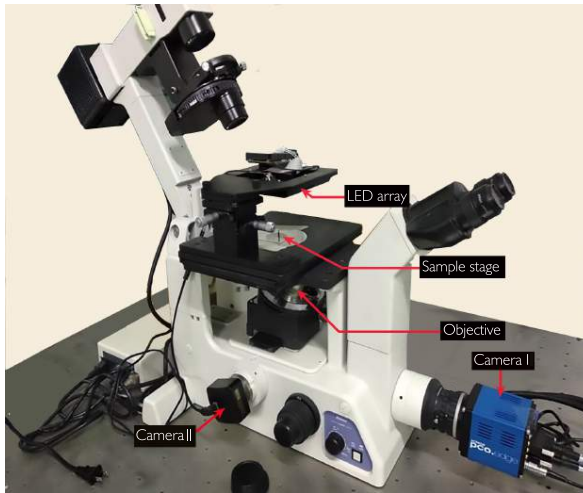


Figure 5.1: FPM Experimental setup. Camera I is the high-dynamic range sCMOS camera that captures 16-bit images, while Camera I is the low cost low dynamic range machine vision camera to capture low bit-depth images.

The basic experimental setup of Fourier Ptychographic Microscopy (FPM) consists of a low NA objective lens and structured coherent illumination source. The most common illumination source is a square grid LED array. A stack of images corresponding to every LED is collected. Following this, an algorithm operates in the spatial and frequency domain to generate high resolution object field from these low resolution images. Figure 5.1 shows the experimental setup used for capturing the real dataset. Camera I is a low cost machine vision CMOS camera that allows 3 bit depth settings - 8, 12, and 16.

Camera II is high cost sCMOS camera that only allows 16-bit imaging.

The experimental setup to capture real data for this work consists of a Nikon Eclipse TE300 inverted microscope and a 10X/0.25 NA objective lens for imaging. An AdaFruit programmable RGB LED array (32×32 , planar) is used to illuminate the sample from various angles using Red LED with a wavelength of 630nm. By using central 225 LEDs, sequential low-resolution images are captured using a High Dynamic Range (HDR) scientific grade sCMOS camera (PCO Edge 5.5) and a low cost machine vision CMOS camera (FLIR FLEA3). The sCMOS camera allows us to capture only 16-bit images while the CMOS allows to capture 8, 12 and 16-bit images. The sCMOS camera has a resolution of 2560×2160 pixels and a pixel pitch $6.5\mu m$ while the CMOS sensor has a resolution of 1552×2080 pixels and pixel pitch of $2.5\mu m$. Both the sensors are monochromatic. The sCMOS camera is used on the front port of the inverted microscope while the CMOS camera is used on the side port of the same microscope.

The LEDs on the array have a grid spacing of 4mm and the array is placed at a distance of 80 mm from the sample plane of the microscope. The illumination from the LED array is controlled using an Arduino MEGA 2560 microcontroller, and simultaneously sends a clock signal to external exposure start port of the sCMOS camera for controlling the camera shutter speed with respect to LED illumination.

5.2 DATASET

Quantitative experiments are performed on a simulated dataset. Then perform qualitative evaluation on a larger real dataset captured at different bit-depth settings of the camera.

5.2.1 Simulated Dataset

Images in Iowa histology dataset Iow (2021) are used for simulating objects fields with uncorrelated amplitude and phase. The FoV of the entire histology slide is divided into 320×320 pixels; the amplitude images are normalised, such that the values lie in

the range $[0, 1]$. The pixel values of the phase image are linearly mapped to the range $[-\pi, \pi]$. These objects are further divided into training, validation and test splits. Then the FPM forward model discussed in 1, is used to generate low resolution intensity images of 64×64 pixels from these object field samples.

5.2.2 Captured Real Dataset

For the experiments, histological sample slides are used, specifically Lung carcinoma and Osteosarcoma (HE stained). The dataset is described in detail Appendix A.2. For obtaining groundtruth, AP reconstruction Tian *et al.* (2014) was performed on the sequential FPM measurements captured using the sCMOS camera. Finally, before training, the 4 regions were divided into 640 non-overlapping patches. During training on 16-bit captured images the entire FoV of Camera I of size 1552×2080 pixels was considered. The entire FoV was divided into 160 samples, each of size 128×128 pixels. For fine-tuning the model on lower bit-depth (12-bit and 8-bit) images, the same slides were reused. For inference on the 5 different regions of size 688×688 pixels FoV were cropped from the 3 slides shown in Figure A.3. Each of these were again divided into 64 patches of size 128×128 with overlap of 48 pixels.

5.2.3 Training and Evaluation

Before testing the network on real data, it is network is finetuned on a small set of simulated training data. To capture a training set, first low resolution images are captured in a sequential manner using the sCMOS sensor and the CMOS sensor. To generate the groundtruth data for training, FPM phase reconstructions using the algorithm in Tian *et al.* (2015) is used on the captured low resolution sCMOS images. To compensate for the misalignment between sCMOS and CMOS images, image registration is also performed. In total 8 slides of cervical, cerebral cortical, lung carcinoma and osteosarcoma cells were used for real data. 4 slides were kept for training and 4 for testing.

The proposed neural network is implemented using PyTorch Paszke *et al.* (2019), and train the same on 2 GTX 1080Ti GPUs of 12GiB capacity, for 100 epochs. The loss

function parameters as described in Eq.(4.1.1) are $(\lambda_1, \lambda_2, \lambda_3) = (0.1, 0.05, 1)$ and $(\beta_1, \beta_2) = (0.5, 1)$. Adam optimizer is used with an initial learning rate (LR) of 10^{-4} and a learning rate scheduler that reduces the LR by a factor of 0.1 when the overall loss does not improve. On the simulated data, 3 stages of the update block are used as described in Fig 4.2 for 16-bit and 12-bit depth images, and increase the number of stages to 5 for 8-bit images.

5.3 RESULTS ON SIMULATED DATASET

Table 5.1 shows image quality with two metrics, namely PSNR and SSIM against existing FPM phase retrieval algorithms. The proposed algorithm is shown to perform at par for the 16-bit simulation results and outperforms the baseline methods at lower bit depths of 12-bit and 8-bit in terms of both PSNR and SSIM. Additionally, the proposed model involves fewer trainable parameters and requires about 100 epochs of training. The proposed method is shown to be less memory intensive compared to the deep-learning based models, cDIP and F-cGAN. However, the Gradient computation block is computationally intensive taking longer inference time, compared to matrix computations. The memory and inference time reported in Table 5.1 are for a 3 stage proposed network. Fig 5.2 shows a few phase reconstructions from simulated data at three bit depths. For 16-bit measurement images, both traditional and neural network based methods perform competitively well. Fig 5.2(b) shows AP-based reconstruction method Tian *et al.* (2014) suffers the most from degradation of peripheral LED measurements at lower bit-depths. Gradient-based methods Bian *et al.* (2015) display poor contrast and blurring at the edges as shown in Fig 5.2(c). Consequently, this is corrected for in the proposed method as shown in 5.2(f). cDIP Boominathan *et al.* (2018) shows checkerboard artifacts that become more prominent with under lower bit depth. Similarly, F-cGAN Nguyen *et al.* (2018) model successfully preserves finer details at higher bit-depths, but shows blur artifacts with increase in quantization.

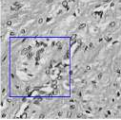
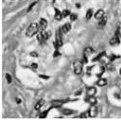
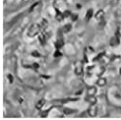
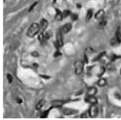
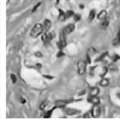
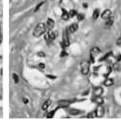
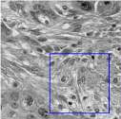
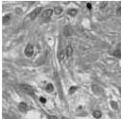
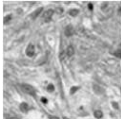
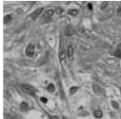
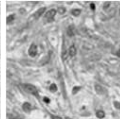
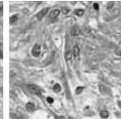
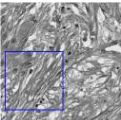
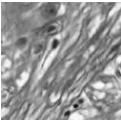
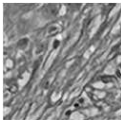
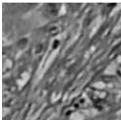
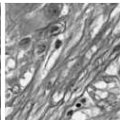
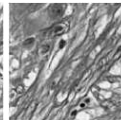
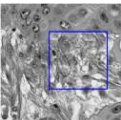
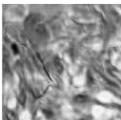
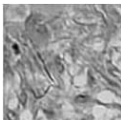
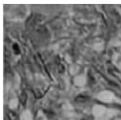
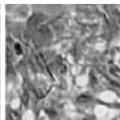
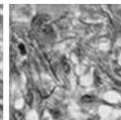
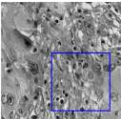
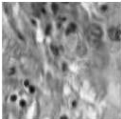
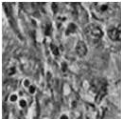
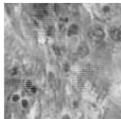
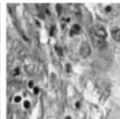
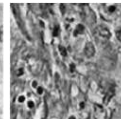
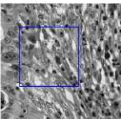
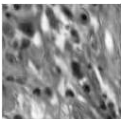
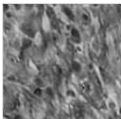
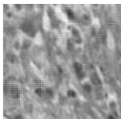
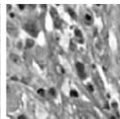
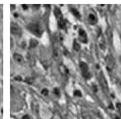
16-bit reconstruction						
	PSNR(SSIM)	29.726(0.898)	24.190(0.774)	23.574(0.820)	26.771(0.847)	29.208(0.810)
12-bit reconstruction						
	PSNR(SSIM)	27.347(0.929)	26.013(0.824)	25.229(0.887)	28.742(0.924)	28.626(0.873)
8-bit reconstruction						
	PSNR(SSIM)	21.706(0.703)	23.759 (0.816)	19.647(0.639)	26.634(0.815)	27.029(0.887)
8-bit reconstruction						
	PSNR(SSIM)	20.724(0.818)	23.832(0.829)	17.608 (0.730)	25.789(0.769)	27.133(0.840)
8-bit reconstruction						
	PSNR(SSIM)	20.638(0.709)	21.364 (0.743)	18.194 (0.712)	23.588(0.739)	26.468 (0.837)
8-bit reconstruction						
	PSNR(SSIM)	20.749(0.791)	20.759 (0.816)	18.803 (0.771)	24.827(0.785)	25.284(0.826)
	(a) Ground truth	(b) AP	(c) WF	(d) cDIP	(e) F-cGAN	(f) LWGNet

Figure 5.2: Phase reconstruction on simulated data using proposed LWGNet method are perceptually superior at lower-bit depth depths compared to existing methods. Traditional algorithms WF and AP do not perform well under lower bit-depths. Among deep methods, cDIP performs poorly at lower bit depths.

5.3.1 Computational efficiency comparison

To analyse the computational efficiency of the proposed method, metrics viz. peak memory usage, inference time and model parameters are considered as shown in Table 5.2. The

Method	16		12		8	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
AP	24.346	0.835	21.764	0.728	20.500	0.594
WF	23.172	0.769	22.706	0.744	21.621	0.730
cDIP	25.654	0.815	22.563	0.718	20.951	0.712
F-cGAN	29.021	0.907	<u>26.325</u>	<u>0.797</u>	<u>25.715</u>	<u>0.765</u>
LWGNet	<u>28.745</u>	<u>0.829</u>	27.726	0.807	26.035	0.802

Table 5.1: Compared to previous works, the proposed LWGNet achieves a better reconstruction quality at lower bit-depths.

experiments have been carried out on a 64bit Intel(R) Xeon(R) CPU and a NVIDIA GeForce RTX 3080 Ti. The inference time is the GPU run-time. The proposed model is compared against the deep architectures as mentioned in Section 5.3. It is observed that the proposed method involves lower number of model parameters and memory, while the inference time is considerably higher due to the gradient computation module.

Method	Memory (in GiB)	Inference time (in ms)	Parameters (in millions)
cDIP	3.24	5	54.9
F-cGAN	3.52	29	7.88
LWGNet	2.26	341	0.39

Table 5.2: Compared to previous works, the proposed LWGNet achieves a better reconstruction quality at lower bit-depths using fewer trainable parameters.

5.4 RESULTS ON REAL DATASET

In this section, the qualitative results on captured dataset have been discussed. The computed gradients of peripheral LED images contain high-frequency details. It is hypothesized that the learned complex-valued neural network parameters are optimized to map these useful gradients to desired object field more effectively than neural networks that use just the intensity images as input. Moreover, due to the extremely small parameter count of the proposed method, fine-tuning on a small dataset of real data doesn't lead to

any overfitting.

5.4.1 Sequential Illumination

The proposed model is tested first on sequentially captured dataset. Figures 5.3, 5.4 and 5.5 show the visual results for 16, 12 and 8 bit depths respectively. The proposed model is tested on real experimental data and compare the results against two deep-learning techniques, cDIP and F-cGAN. Prior to evaluation, all the methods are finetuned on real dataset described in Section 5.2.2. Phase reconstruction obtained from cDIP show low contrast and are prone to checkerboard artifacts. F-cGAN reconstructs the cellular structures but performs relatively poor in reconstruction of finer background details. On the contrary, the proposed algorithm successfully preserves edges and background details compared to the other approaches consistently for all bit depths. However, it is also observed that the images appear to lack density information in some of the cases such as 5.4 (bottom).

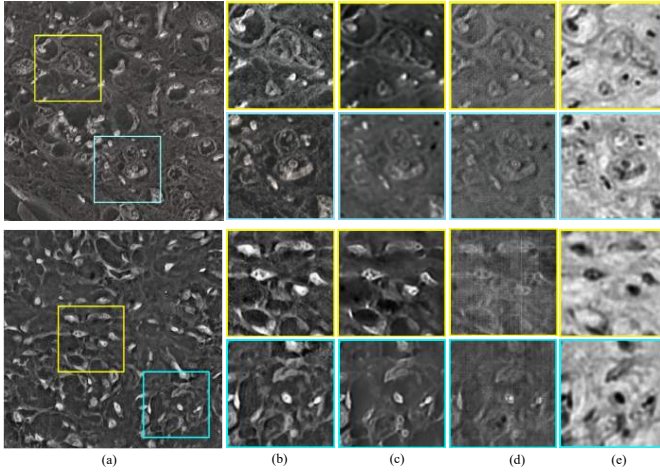


Figure 5.3: Reconstruction on 16-bit LR images using Lung carcinoma sample (top) and Osteosarcoma (bottom). (a) Sample FoV LWGNet reconstruction, (b) LWGNet insets (c) F-cGAN reconstruction (d) cDIP reconstruction and (e) LR bright-field image.

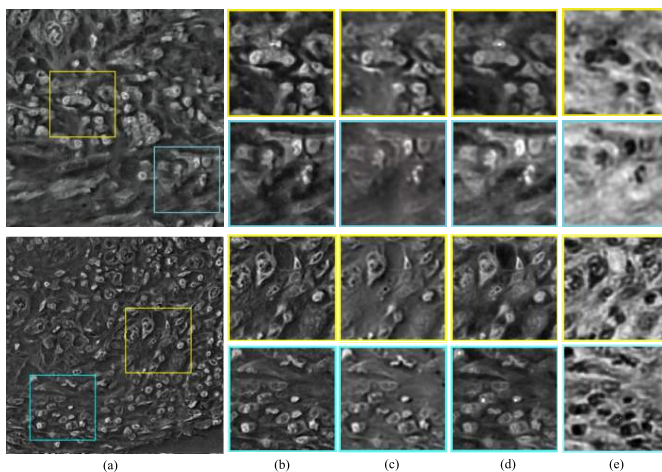


Figure 5.4: Reconstruction on 12-bit LR images using Lung carcinoma sample (top and bottom). (a) Sample FoV LWGNet reconstruction, (b) LWGNet insets (c) F-cGAN reconstruction (d) cDIP reconstruction and (e) LR bright-field image.

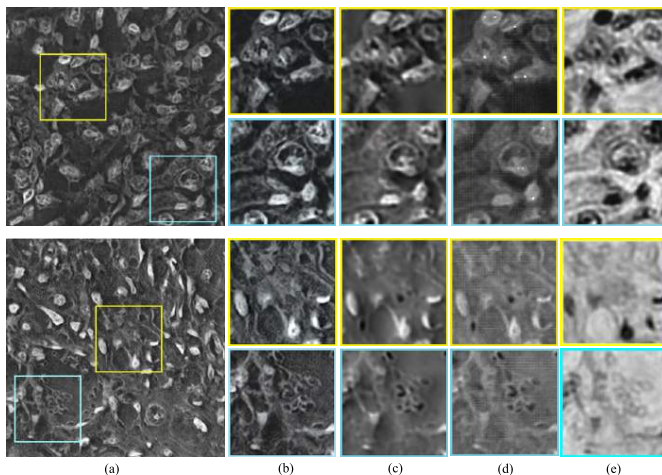


Figure 5.5: Reconstruction on 8-bit LR images for Lung carcinoma sample (top) and Osteosarcoma (bottom). (a) Sample FoV LWGNet reconstruction, (b) LWGNet insets (c) F-cGAN reconstruction (d) cDIP reconstruction and (e) LR bright-field image.

5.4.2 Multiplexed Illumination

Experiments show that using the proposed multiplexing scheme in Section 4.2.2, the LWGNet successfully performs phase reconstruction as shown in Figure 5.6. It is observed that foreground details are more prominent compared to the sequential illumination scheme. From the phase reconstruction of the Lung carcinoma sample in Figure 5.6 (bottom), it is observed that the perceived density variation comes out more prominent in the multiplexed illumination case.

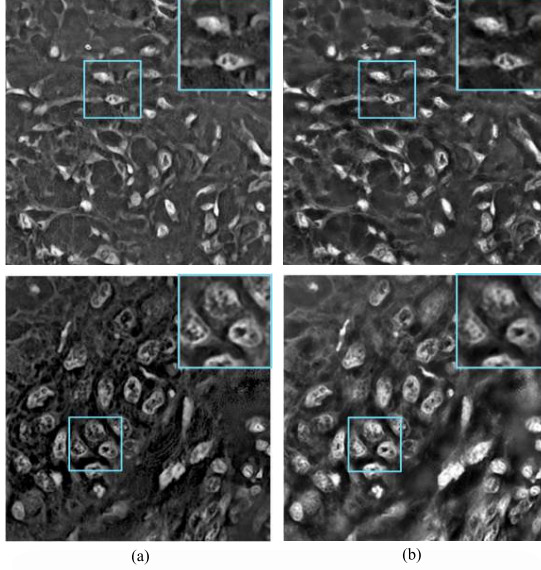


Figure 5.6: Comparison of phase reconstruction results for Osteosarcoma(top) and Lung carcinoma (bottom) samples for (a) Sequential and (b) Multiplexed illumination schemes using 225 and 45 images respectively.

5.4.3 Sensor Quantization Analysis

In this section, it is experimentally verified that the proposed method shows little perceptual variation over reconstruction with increasing quantization noise or decrease in the bit depth of the input data. The proposed approach is finetuned using the dataset

described in Section 5.2.2 prior to evaluation. Figure 5.7 shows reconstruction for Osteosarcoma (top) and Lung carcinoma (bottom) samples at different bit-depth settings of the camera. The reconstructions obtained are perceptually indistinguishable in the first case as shown in the inset. In the second case, minor are observed changes in reconstructions of sharp edge features with the bit-depth setting.

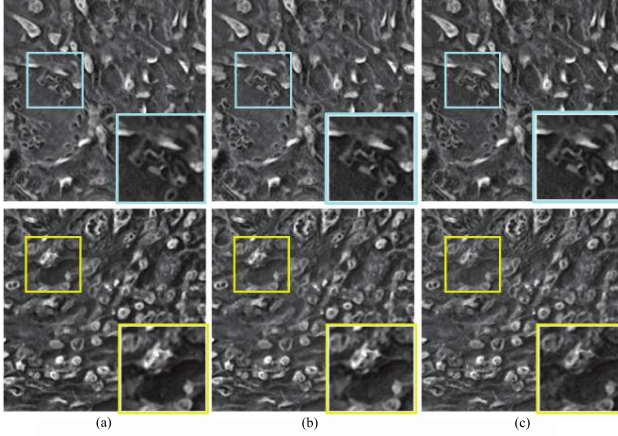


Figure 5.7: Comparison of phase reconstruction using the proposed LWGNet on Osteosarcoma (top) and Lung carcinoma (bottom) at various bit depth settings. The reconstruction quality is consistent even under high quantization noise.

5.5 ABLATION STUDY

In this section, the impact of different components of the proposed network is analysed using simulated data.

Effect of Complex-valued Operations

Here, the necessity of the complex-valued operations described in Chapter ?? is verified. To do that, the complex-valued operation of the proposed network is replaced with real-valued operations acting individually on the real and imaginary channels of the complex gradient. The rest of the architecture and loss functions are kept the same. Top

row of Figure 5.8 presents the visual comparison between reconstructions obtained from complex-valued and real-valued networks. It is shown that the use of complex valued operations increases the average PSNR by 5dB.

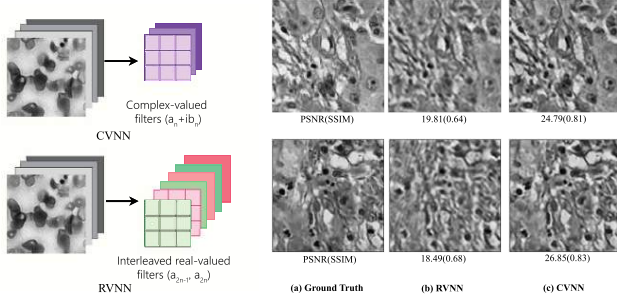


Figure 5.8: Left: Representation of complex-valued filters(top) and interleaved real-valued filters(bottom). Right: Using complex-valued operations leads to superior quality of reconstructions.

Choice of Network Architecture

Figure 5.9 shows phase reconstruction from simulated 16-bit LR images using complex-valued neural network architectures described in Section 3.4.2. The Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) are reported as image quality metrics for all the experiments. Each network is trained for 1000 epochs. The ResNet architecture is less memory intensive and shows an improvement of ≈ 1 dB of PSNR over its UNet counterpart which suffers from contrast artifacts. The average SSIM however does not display any significant difference remains ≈ 0.8 .

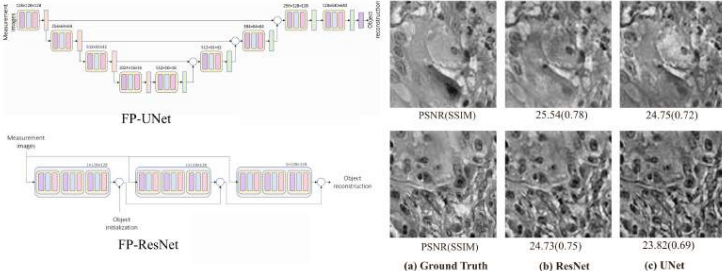


Figure 5.9: Phase reconstruction on 16-bit simulated data using (b) FP-ResNet architecture and (c) FP-UNet architecture. The FP-ResNet shows marginal improvement over FP-UNet which shows contrast artifacts.

Effect of Processing Gradients

To verify that processing gradients through a learned network actually helps, a variant of the proposed architecture is trained with the following update stage where the learned neural network acts on the estimated object field,

$$\mathbb{O}_f = \hat{\Psi}\{\mathbb{O}^K\} \quad (5.1)$$

where,

$$\mathbb{O}^{i+1} = (\mathbb{O}^i + \frac{1}{L} \sum_{\ell=1}^L g_{\ell}^i), i \in \{0, 1, \dots, K-1\} \quad (5.2)$$

and $\hat{\Psi}\{\cdot\}$ is N numbers of $\Psi^i\{\cdot\}$ (described in Section 4.1) stacked together. $K = 5$ iterations of Wirtinger flow and stack $N = 3$ numbers of $\Psi\{\cdot\}$ is used for this purpose to compare against 3 stages of proposed approach. The rest of the architecture and loss functions are kept the same as proposed method. Second row in Figure 5.10 shows the visual results highlighting the effect of processing the gradients. The proposed way of learning a mapping from the stochastic gradients to object fields leads to more effective use of the gradient information compared to an enhancing an object field using a neural

network after wirtinger updates.

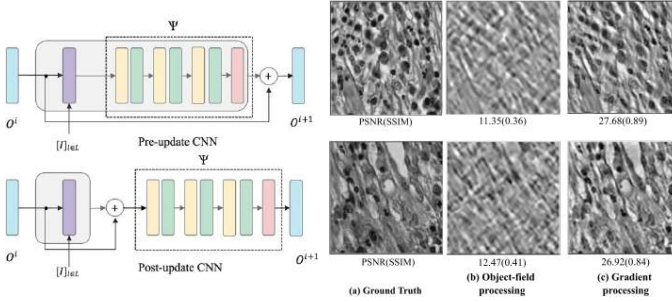


Figure 5.10: Left: Representation of processing of gradients(top) and processing O^i post update(bottom). Right: The proposed way of learning the network on gradient leads to more effective use of gradient information than learning a network on estimated object field.

Effect of Number of Unrolled Stages

In this section, the effect of the number of unrolled iterative stages (K in Section 4.1) is analysed. Figure 5.11 represents the variation of PSNR with the number of update blocks in the proposed architecture varied between 2 to 10 blocks. It is found that the PSNR saturates after 3 stages for 16-bit and 12-bit data, while it takes at least 5 stages to achieve a similar performance for 8 bit data.

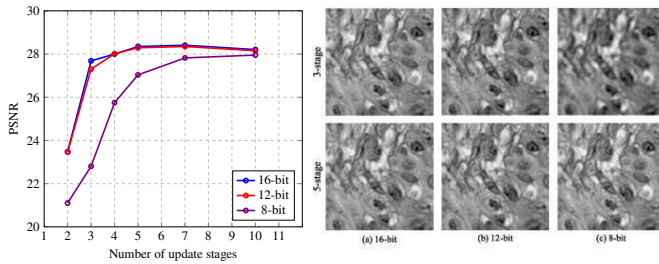


Figure 5.11: Left: PSNR vs Number of unrolled stages. Right: corresponding reconstructions for 3 and 5 stage LWGNet under different bit depths. For 16 and 12 bit data, after just 3 stages the performance saturates while 5 stages are needed for 8 bit data.

CHAPTER 6

CONCLUSION

A novel physics-based neural network for FPM reconstructions is proposed. The proposed network derives inspiration from conventional phase retrieval algorithms and combines it with data-driven neural networks. Unlike the naive Wirtinger flow or alternate projections, the network learns a non-linear mapping from stochastic gradients to object field intermediates. Complex-valued neural networks learn this non-linear mapping. It is then extended to a multiplexed illumination scheme. The proposed method performs better than existing FP phase reconstruction techniques, especially for challenging scenarios like low dynamic range. We attribute the success of our network on such challenging conditions to the novel non-linear physics-based mapping. The ability to perform high-resolution wide-field of view microscopy using low-cost sensors through physics-inspired data-driven techniques can significantly decrease system costs and make point-of-care diagnostics more accessible.

6.1 FUTURE WORK

The long term goal is to design and develop a compact prototype to be employed in remote areas without access to sophisticated microscopic imaging equipment. The design involves the use of low-cost component parts and further developments in the reconstruction pipeline. The problem of usage of commercial grade sensors is addressed in this work. It should be noted that, like most physics-based unrolled neural networks, our proposed architecture suffers from slower inference due to the full resolution operations involved in the physics based stages. In future it would be interesting to look into integrating memory efficient operations Kellman *et al.* (2020) with the proposed approach. Also, the performance of the proposed algorithm in presence of calibration errors needs

to be evaluated and a pre-processing stage is to be included to accommodate errors due to LED misalignment, and effectively eliminate background noise in dark-field images for improvement of SNR. We have also not studied the effect of optical aberrations (optical aberrations); it will be an interesting problem to look at. The challenge is to include additional computational modules to the proposed model without greatly increasing the complexity of the computation graph of the model in order to effectively facilitate backpropagation of gradients during training. The final task is developing a compact design for the accommodation of hardware components and deploying the proposed algorithm.

APPENDIX A

APPENDIX: EXPERIMENTAL SPECIFICATION

A.1 CALCULATION OF RESOLVABLE CAPACITY

The distance between the USAF resolution bars (a) is given for every Group and Element. From the chart, $Resolution = 2^{(Group\ Number + \frac{Element\ Number}{6})}$, for a particular element from a particular group that is resolved.

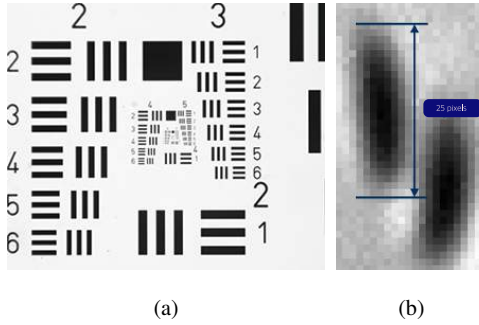


Figure A.1: (a) USAF chart and (b) Measurement of cell image.

For incoherent illumination, the resolvable capacity is given by $\frac{\lambda}{2NA}$, while that for coherent illumination is $\frac{\lambda}{NA}$. For partially coherent illumination, it will lie in the range of $[\frac{\lambda}{2NA}, \frac{\lambda}{NA}]$. The resolution can also be estimated analytically as follows,

$$\text{Effective pixel size of sensor} = \frac{\text{Pixel size of the sensor}}{\text{Total magnification}},$$

The Total magnification = Objective magnification \times System magnification.

For sCMOS sensor (PCO Edge), the pixel size is $6.5\mu m$ and that for CMOS sensor

(FLIR) is $2.5\mu m$.

Given, the System Magnification = $2\times$, the Total magnification = 8(using $4/0.1NA$ objective) or Total Magnification = 20 (using $10\times/0.25NA$ objective).

Cell Size = No. of pixels along length a cell \times Effective Pixel Size of sensor = $25 \times (6.5/20)\mu m = 8.5\mu m$, using sCMOS sensor and the $10\times/0.25NA$ objective.

Spatial sampling ratio = $\frac{\text{Effective Pixel Size}}{\text{resolvable capacity}} < 1$. Hence, Nyquist criterion is satisfied.

A.2 EXPERIMENTAL DATASET

We used histopathological slides (HE stained) for our experiments. We used 2 slides each of lung carcinoma, osteosarcoma, and cervical cells, and 1 slide of cerebral cortex. 1 slide each of lung carcinoma and cerebral cortex and 2 slides of cervical cells were used for training and the rest were kept for testing. From each slide, multiple regions were captured by translating the slides laterally. Specifically, from the training slides, 4 regions were captured while from the test slides 5 different regions were captured. For each region, 225 low-resolution sequential images were obtained corresponding to the 225 LEDs using 4 different imaging settings: 3 using the 3 bit-depth (8,12 and 16-bit) settings of camera II and 1 using 16-bit setting of camera I. The low resolution images from the two camera were registered to account for the misalignment. In Fig A.3, exemplar FoV representing each of 3 slides is shown.

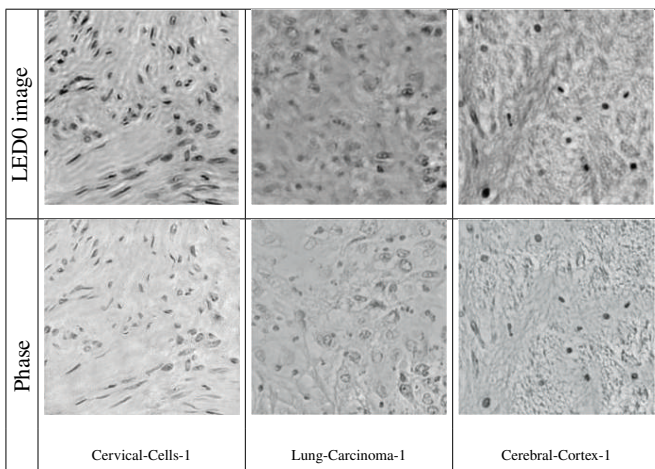


Figure A.2: Histological samples used for training. Top row shows the brightfield images captured using sCMOS camera while the bottom row shows the corresponding AP phase reconstructions.

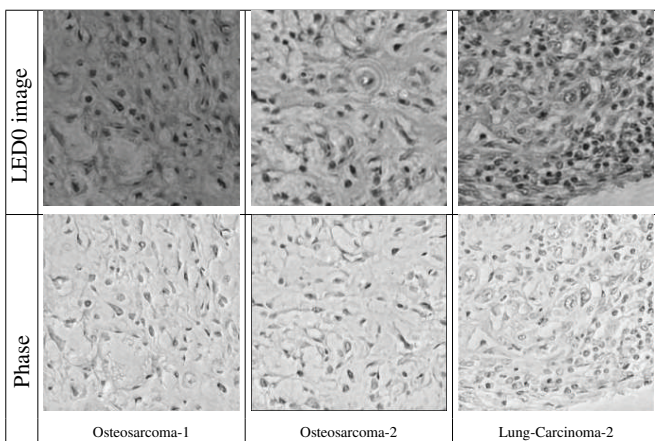


Figure A.3: Histological samples used for testing. Top row shows the brightfield images captured using sCMOS camera while the bottom row shows the corresponding AP phase reconstructions.

A.3 SUBMISSIONS

I. In review: European Conference in Computer Vision (ECCV 2022) **A. Saha, S. Khan, S. Schrawat, S. Prabhu, S. Bhattacharya, K. Mitra** (2022), LWGNet - Learned Wirtinger Gradients for Fourier Ptychographic Phase Retrieval.

II. Presentation: The work is to be presented at ICCP, 2022.

REFERENCES

1. (2021). Iowa virtual slidebox. URL <http://www.path.uiowa.edu/virtualslidebox/>.
2. (2022). Understanding microscopes and objectives. URL <https://www.edmundoptics.jp/knowledge-center/application-notes/microscopy/understanding-microscopes-and-objectives/>.
3. **Abbe, E.** (1873). Contributions to the theory of the microscope and microscopic perception. *Archive for microscopic anatomy*, **9**(1), 413–468.
4. **Aidukas, T., R. Eckert, A. R. Harvey, L. Waller, and P. C. Konda** (2019). Low-cost, sub-micron resolution, wide-field computational microscopy using opensource hardware. *Scientific reports*, **9**(1), 1–12.
5. **Bian, L., J. Suo, G. Zheng, K. Guo, F. Chen, and Q. Dai** (2015). Fourier ptychographic reconstruction using wirtinger flow optimization. *Opt. Express*, **23**(4), 4856–4866. URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-23-4-4856>.
6. **Bian, L., G. Zheng, K. Guo, J. Suo, C. Yang, F. Chen, and Q. Dai** (2016). Motion-corrected fourier ptychography. *Biomedical optics express*, **7**(11), 4543–4553.
7. **Boominathan, L., M. Maniparambil, H. Gupta, R. Baburajan, and K. Mitra** (2018). Phase retrieval for fourier ptychography under varying amount of measurements. *arXiv preprint arXiv:1805.03593*.
8. **Bostan, E., M. Soltanolkotabi, D. Ren, and L. Waller**, Accelerated wirtinger flow for multiplexed fourier ptychographic microscopy. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018.
9. **Breslauer, D. N., R. N. Maamari, N. A. Switz, W. A. Lam, and D. A. Fletcher** (2009). Mobile phone based clinical microscopy for global health applications. *PLoS one*, **4**(7), e6320.
10. **Fienup, J. R.** (1982). Phase retrieval algorithms: a comparison. *Appl. Opt.*, **21**(15), 2758–2769. URL <http://opg.optica.org/ao/abstract.cfm?URI=ao-21-15-2758>.
11. **Gerchberg, R. W.** (1972). A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, **35**, 237–246.
12. **Hasinoff, S. W., F. Durand, and W. T. Freeman**, Noise-optimal capture for high dynamic range photography. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010.
13. **Jiang, S., K. Guo, J. Liao, and G. Zheng** (2018). Solving fourier ptychographic

- imaging problems via neural network modeling and tensorflow. *Biomed. Opt. Express*, **9**(7), 3306–3319. URL <http://www.osapublishing.org/boe/abstract.cfm?URI=boe-9-7-3306>.
14. **Kappeler, A., S. Ghosh, J. Holloway, O. Cossairt, and A. Katsaggelos**, Ptychnet: Cnn based fourier ptychography. In *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.
 15. **Kellman, M., E. Bostan, M. Chen, and L. Waller**, Data-driven design for fourier ptychographic microscopy. In *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2019.
 16. **Kellman, M., K. Zhang, E. Markley, J. Tamir, E. Bostan, M. Lustig, and L. Waller** (2020). Memory-efficient learning for large-scale computational imaging. *IEEE Transactions on Computational Imaging*, **6**, 1403–1414.
 17. **Lakshminarayanan, V. and A. Fleck** (2011). Zernike polynomials: a guide. *Journal of Modern Optics*, **58**(7), 545–561.
 18. **Ledig, C., L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi** (2017). Photo-realistic single image super-resolution using a generative adversarial network.
 19. **Mirza, M. and S. Osindero** (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
 20. **Nguyen, T., Y. Xue, Y. Li, L. Tian, and G. Nehmetallah** (2018). Deep learning approach for fourier ptychography microscopy. *Opt. Express*, **26**(20), 26470–26484. URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-26-20-26470>.
 21. **Osherovich, E.** (2012). Numerical methods for phase retrieval. *arXiv preprint arXiv:1203.4756*.
 22. **Ou, X., R. Horstmeyer, C. Yang, and G. Zheng** (2013). Quantitative phase imaging via fourier ptychographic microscopy. *Opt. Lett.*, **38**(22), 4845–4848. URL <http://ol.osa.org/abstract.cfm?URI=ol-38-22-4845>.
 23. **Ou, X., G. Zheng, and C. Yang** (2014). Embedded pupil function recovery for fourier ptychographic microscopy. *Opt. Express*, **22**(5), 4960–4972. URL <http://opg.optica.org/oe/abstract.cfm?URI=oe-22-5-4960>.
 24. **Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala**, Pytorch: An imperative style, high-performance deep learning library. In **H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett** (eds.), *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, 8024–8035. URL <http://papers.neurips.cc/paper/>

9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

25. **Rippel, O., J. Snoek, and R. P. Adams** (2015). Spectral representations for convolutional neural networks. *Advances in neural information processing systems*, **28**.
26. **Smith, Z. J., K. Chu, A. R. Espenson, M. Rahimzadeh, A. Gryshuk, M. Molinaro, D. M. Dwyre, S. Lane, D. Matthews, and S. Wachsmann-Hogiu** (2011). Cell-phone-based platform for biomedical device development and education applications. *PloS one*, **6**(3), e17150.
27. **Song, P., S. Jiang, H. Zhang, X. Huang, Y. Zhang, and G. Zheng** (2019). Full-field fourier ptychography (ffp): Spatially varying pupil modeling and its application for rapid field-dependent aberration metrology. *APL Photonics*, **4**(5), 050802.
28. **Tanaka, G.** (2013). Complex-valued neural networks: Advances and applications [book review]. *ieee Computational intelligence magazine*, **8**(2), 77–79.
29. **Tian, L., X. Li, K. Ramchandran, and L. Waller** (2014). Multiplexed coded illumination for fourier ptychography with an led array microscope. *Biomed. Opt. Express*, **5**(7), 2376–2389. URL <http://www.osapublishing.org/boe/abstract.cfm?URI=boe-5-7-2376>.
30. **Tian, L., Z. Liu, L.-H. Yeh, M. Chen, J. Zhong, and L. Waller** (2015). Computational illumination for high-speed in vitro fourier ptychographic microscopy. *Optica*, **2**(10), 904–911. URL <http://www.osapublishing.org/optica/abstract.cfm?URI=optica-2-10-904>, <https://www.laurawaller.com/opensource/>.
31. **Trabelsi, C., O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal** (2018). Deep complex networks.
32. **Zhang, H., Y. Liang, and Y. Chi** (2017). A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms. *Journal of Machine Learning Research*, **18**(141), 1–35. URL <http://jmlr.org/papers/v18/16-572.html>.
33. **Zheng, G., R. Horstmeyer, and C. Yang** (2013). Wide-field, high-resolution fourier ptychographic microscopy. *Nature photonics*, **7**(9), 739–745.
34. **Zheng Guoan, J. S. S. P. Y. C., Shen Cheng** (2021). Concept, implementations and applications of fourier ptychography. *Nature Reviews Physics*, **3**, 207–223. URL <https://doi.org/10.1038/s42254-021-00280-y>.

CURRICULUM VITAE

NAME

Atreyee Saha

EDUCATION QUALIFICATIONS

Degree	Major	University/Institute	Year	CGPA/%
<i>MS in Engineering</i>	Electrical Engineering	IIT Madras	2021	8.19 CGPA
<i>Bachelor of Engineering</i>	Electrical Engineering	Jadavpur University	2019	8.68 CGPA
<i>Higher Secondary</i>	Science	Nava Nalanda High School	2015	86.0%

GENERAL TEST COMMITTEE

Chairperson

Dr. A. N. Rajagopalan
Professor,
Department of Electrical Engineering,
IIT Madras,
Chennai, India 600036.

Guide(s)

Dr. K. Mitra Assistant Professor,
Department of Electrical Engineering,
IIT Madras,
Chennai, India 600036.

Member(s)

Dr. S. Bhattacharya
Professor,
Department of Electrical Engineering,
IIT Madras,
Chennai, India 600036.

Dr. Anurag Mittal
Professor,
Department of Computer Science and Engineering,
IIT Madras,
Chennai, India 600036.