LWGNet - Learned Wirtinger Gradients for Fourier Ptychographic Phase Retrieval

Atreyee Saha, Salman S. Khan, Sagar Sehrawat, Sanjana S. Prabhu, Shanti Bhattacharya, and Kaushik Mitra

Indian Institute of Technology Madras, Chennai 600036, India

Abstract. Fourier Ptychographic Microscopy (FPM) is an imaging procedure that overcomes the traditional limit on Space-Bandwidth Product (SBP) of conventional microscopes through computational means. It utilizes multiple images captured using a low numerical aperture (NA) objective and enables high-resolution phase imaging through frequency domain stitching. Existing FPM reconstruction methods can be broadly categorized into two approaches: iterative optimization based methods, which are based on the physics of the forward imaging model, and datadriven methods which commonly employ a feed-forward deep learning framework. We propose a hybrid model-driven residual network that combines the knowledge of the forward imaging system with a deep datadriven network. Our proposed architecture, LWGNet, unrolls traditional Wirtinger flow optimization algorithm into a novel neural network design that enhances the gradient images through complex convolutional blocks. Unlike other conventional unrolling techniques, LWGNet uses fewer stages while performing at par or even better than existing traditional and deep learning techniques, particularly, for low-cost and low dynamic range CMOS sensors. This improvement in performance for low-bit depth and low-cost sensors has the potential to bring down the cost of FPM imaging setup significantly. Finally, we show consistently improved performance on our collected real data¹.

Keywords: Fourier ptychography, Physics-based network, Computational imaging

1 Introduction

One of the main challenges in medical computational 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. [6,18] have used smartphone cameras for microscopy, while [1] have used 3D printed microscopes for wide field of view high resolution imaging. However, the use of cheaper and inefficient hardware introduces limits on the imaging capabilities of these systems through various degradations which can only be dealt by designing effective computational algorithms.

¹ We have made the code available at: https://github.com/at3e/LWGNet.git

In this work, we will be focusing on a particular microscopy technique called Fourier Ptychographic Microscopy (FPM) [23]. It is a computational microscopy technique that allows us to perform high resolution and wide field of view (FOV) imaging. It circumvents the limit on SBP of conventional microscope by relying on multiple low-resolution captures of the sample under programmed illumination and reconstruction via phase retrieval algorithms. Development in FPM in the recent years has included improving the temporal resolution through multiplexed illumination [20], designing better reconstruction algorithms [4,2,5], and designing better illumination codes [9].

Despite the above-mentioned progress, there hasn't been enough work done to make these FPM systems accessible for point of care diagnostics. Some of the works on making FPM accessible are [1,10,12]. However, the authors rely on extensive hardware modifications without much changes to the algorithmic aspects. The following fundamental challenges have been the main reason behind that. First, the existing FPM systems, like most low-light imaging modalities, suffer from significant noise and dynamic range problems especially for darkfield images. These FPM systems rely on expensive optics and scientific grade sCMOS sensors to increase the SNR and dynamic range of the low-resolution darkfield captures which increases the system cost significantly [20]. 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 sensor degrades. Third, the existing reconstruction techniques are either slow or are model-independent data-driven techniques that ignore the forward imaging process completely.

Keeping the above challenges in mind, we propose 'LWGNet' for sequential FPM reconstruction. 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 [2]. Previous unrolling works perform their operation in the image or object field space, while LWGNet perform these operations in the gradient of the object fields. LWGNetlearns a non-linear mapping from complex stochastic gradients to intermediate object fields through complex-valued neural networks. Such a learned mapping helps preserve the high frequency details in the peripheral darkfield images, especially for low dynamic range sensor. LWGNet outperforms both traditional and deep-learning methods in terms of reconstruction quality, especially for low-cost machine vision sensors with poor dynamic range. We show this by performing extensive evaluations on simulated and real histopathological data captured under different bit depths. In summary, we make the following contributions:

- We propose LWGNet 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.

- We collect a real dataset of 8, 12 and 16 bit low resolution measurements using a CMOS sensor along with the corresponding aligned groundtruth for finetuning our method. To the best of our knowledge, this is the only FPM dataset captured with multiple bit depth settings using a low cost CMOS sensor. This dataset will be made public upon acceptance.
- 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.



Fig. 1: Overview of FPM phase reconstruction. (a) The object is placed between a planar LED array and objective of the inverted microscope. Low resolution images corresponding to ℓ^{th} LED is captured. (b) Representation of the equivalent forward model, comprises of down-sampling the object spectrum by the system pupil function centred at $(k_{x\ell}, k_{y\ell})$. (c) Examples of input low resolution measurements under different sensor quantization and corresponding reconstructions using F-cGAN and the proposed LWGNet.

2 Background

2.1 The nature of phase retrieval problem

We are interested in the case of phase recovery from intensity measurements only. The experimental setup for FPM involves illuminating the object from multiple angles and is described in detail in [23]. Mathematically, the phase retrieval problem while using multiple structured illuminations for intensity measurement is as follows:

Find 0, subject to
$$I_{\ell} = |\mathcal{F}^{-1}\{P_{\ell} \odot \mathcal{F}\{\mathbb{O}(\mathbf{k} - \mathbf{k}_{\ell})\}\}|^2$$
, for $\ell = 1, 2, \dots, L$, (1)

where \mathcal{F} is the 2D spatial Fourier transform and \mathcal{F}^{-1} is the corresponding inverse transform. $P_{\ell} \in \mathbb{C}^N$ 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 \mathcal{O} is an optically thin sample, i.e. its transmittance is close to unity. Classical retrieval algorithms seek to solve the following minimization problem,

$$\min_{0} \sum_{\ell=1}^{L} \|I_{\ell} - |A_{\ell}\{0\}|^2 \|_2^2,$$
(2)

where $A\{.\} \triangleq \mathcal{F}^{-1} \circ P_{\ell} \odot \mathcal{F}\{.\}.$

The objective is a non-convex, real-valued function of complex-valued object field. Conventional approaches like gradient descent will converge to a stationary point.

2.2 Related Works

There have been numerous works on FPM reconstruction. Ou et al.[15] successfully performed whole slide high-resolution phase and amplitude imaging using a first-order technique based on alternate projections (AP). Song et al.[19] proposed algorithms to overcome system aberration in FPM reconstruction. In [3], the authors proposed sample motion correction for dynamic scenes. The work by Tian et al. [20] extended the AP algorithm for multiplexed illumination in FPM setup. Besides AP, Wirtinger flow based methods are shown to perform well under low SNR [2]. The Wirtinger flow algorithm is also extended to the multiplexed scheme by Bostan et al. [5].

Recently, deep learning techniques have also been explored to solve the FPM reconstruction problem. Jiang et al. [7] show a novel approach by treating the object as a learnable network parameter. However, as it requires optimization for each patch, it has a large inference time. Kappeler et al. [8] have performed high-resolution amplitude recovery using a CNN architecture. Nguyen et al. [14] have performed time-lapse high-resolution phase imaging via adversarial training of a U-Net architecture [13]. The authors performed phase imaging using a limited set of low-resolution images. They train (transfer learn in the case of new biological type of cells) U-Net on a subset of low resolution video measurements. While testing, they predict phase of dynamic cell samples on the subsequent frames. FPM phase reconstruction under various overlap conditions between adjacent low-resolution images in the frequency domain has been studied by Boominathan et al. [4]. However, they show results only on simulated samples. Kellman et al. [9] have designed a Physics-based Neural Network (PbNN) that learns patterns for multiplexed illumination by optimizing the weights of the LEDs.

An important line of work is the employment of high-resolution phase imaging using low-cost components. To our best knowledge, the only attempt made so far is by Aidukas et al. [1], who employed traditional reconstruction algorithm accompanied with an elaborate calibration and pre-processing of low-resolution images taken from a commercial-grade camera. The authors performed amplitude reconstruction, but their system has fundamental limitations due to sensor size and optical aberrations.

3 Proposed Method: LWGNet



Fig. 2: Model architecture of LWGNet comprises of three stages of the LWG update block. The inputs are object, pupil initialization and measurement images for every stage. LWG block comprises of a Gradient computation block followed by three stages of Conv+Tanh layers followed by a Fully Connected(FC) layer. The adder block updates the input object field using the modified gradients.

The naive Wirtinger flow algorithm seeks to minimize the following objective [2],

$$f(\mathfrak{G}) = \sum_{\ell=1}^{L} \|I_{\ell} - |A_{\ell}\{\mathfrak{G}\}\|^2 \|_2^2$$
(3)

$$= \sum_{\ell=1}^{L} \|I_{\ell} - A_{\ell}\{0\} \odot \overline{A_{\ell}\{0^{i}\}}\|_{2}^{2},$$
(4)

where $A\{.\} \triangleq \mathcal{F}^{-1} \circ P_{\ell} \odot \mathcal{F}\{.\}$, \mathcal{F} is the 2D spatial Fourier transform, \mathcal{F}^{-1} is the corresponding inverse transform, P_{ℓ} is the pupil function corresponding to the ℓ^{th} illumination source and I_{ℓ} is the capture ℓ^{th} measurement for the same illumination source. Minimizing Eq.(4) with respect to \mathcal{O} using gradient descent leads to the following i^{th} step update,

$$\mathfrak{S}^{i+1} = \mathfrak{S}^{i} - \eta \frac{1}{L} \sum_{\ell=1}^{L} \nabla_{\mathfrak{S}} f_{\ell}^{i}, \tag{5}$$

where η is the step-size and,

$$\nabla_{\mathbb{G}} f_{\ell}^{i} = A_{\ell}^{H} \{ (|A_{\ell} \{ \mathbb{G}^{i} \}|^{2} - I_{\ell}) \odot A_{\ell} \{ \mathbb{G}^{i} \} \}.$$
(6)

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.(5) followed by convolutional block. However, we found that such a naive unrolling doesn't converge to any meaningful solution. To overcome this, we design a K-stage unrolled network with the i^{th} stage performing the following operation

$$\mathbb{O}^{i+1} = \mathbb{O}^i + \Psi^i([\boldsymbol{g}_\ell]^i_{\ell \in L}),\tag{7}$$

where

$$g_{\ell}^{i} = A_{\ell}^{H} \{ (|A_{\ell} \{ \phi_{\ell-1}^{i} \}|^{2} - I_{\ell}) \odot A_{\ell} \{ \phi_{\ell-1}^{i} \} \},$$
(8)

$$\phi_{\ell}^{i} = \phi_{\ell-1}^{i} - \eta g_{\ell-1}^{i}, \tag{9}$$

and $\Psi(.)$ is a learned complex neural network. , ϕ_{ℓ}^{i} is an intermediate object field that is estimated at the *i*-th stage using the gradients obtained with respect to the ℓ -th LED measurement. We initialize $\phi_{1}^{i}as \mathbb{G}^{i}$. 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 and update the object field as shown in Eq.(7). Experimentally, we found that such an unrolled network converged faster with fewer K than naively unrolled wirtinger flow. $\Psi(.)$ consists of three 3×3 complex convolutions followed by complex Tanh non-linearity, instance norm, and a fully connected layer. The learned non-linear $\Psi(.)$ helps combine the stochastic gradients in a more effective way especially for peripheral darkfield images which are typically degraded due to lower bit depth and noise.

We simulate complex arithmetic using real-valued entities [22] in $\Psi(.)$. We perform equivalent 2D complex-valued convolution 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 $\mathbf{w}_{mn} = \mathbf{a}_{mn} + j\mathbf{b}_{mn}$ that convolves with a 2D complex input $\mathbf{z}_m = \mathbf{x}_m + j\mathbf{y}_m$. Here, \mathbf{a}_{mn} and \mathbf{b}_{mn} are real-valued kernels, and \mathbf{x}_m and \mathbf{y}_m are also real.

$$\mathbf{w}_{mn} * \mathbf{z}_m = \mathbf{a}_{mn} * \mathbf{x}_m + j\mathbf{b}_{mn} * \mathbf{y}_m.$$
(10)

We initialize filter weights as a uniform distribution. The CNN layer is followed by an 2D instance normalization layer. An amplitude-phase type non-linear activation function for and complex-valued z = x + jy is defined as follows

$$\mathbb{C}\mathrm{Tanh}(\mathbf{z}) = \mathrm{Tanh}(|\mathbf{z}|)e^{i\theta_{\mathbf{z}}}$$
(11)

acts on the normalised outputs.

Finally, after 3 convolutional and non-linearity blocks, we use a fully connected block that acts on the channel dimension.

3.1 Loss Functions

Let O be the ground truth object and \hat{O} be the reconstruction. We use the following weighted combination of loss functions to optimize our network,

$$\mathscr{L} = \lambda_1 \mathscr{L}_{MSE} + \lambda_2 \mathscr{L}_{FMAE} + \lambda_3 \mathscr{L}_{VGG} \tag{12}$$

where,

$$\mathscr{L}_{MSE} = \|\angle(\hat{O}) - \angle(O)\|_2^2 + \||\hat{O}| - |O|\|_2^2$$
(13)

$$\mathscr{L}_{FMAE} = \||\mathscr{F}(\hat{O})| - |\mathscr{F}(O)|\|_1 \tag{14}$$

$$\mathscr{L}_{VGG/i,j} = \beta_1 \|\psi_{i,j}(|\hat{O}|) - \psi_{i,j}(|O|)\|_2^2 + \beta_2 \|\psi_{i,j}(\angle \hat{O}) - \psi_{i,j}(\angle O)\|_2^2$$
(15)

 $\|\cdot\|_p$ represents the *p*-norm. Eq.(13) 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 [11]. It minimizes the MSE from the output of the feature maps of the pre-trained VGG-19 network. $\psi_{i,j}(.)$ is the feature map obtained by the *j*-th convolution (after activation) before the *i*-th max-pooling layer in the network, which we consider given. Here, we use ReLU output $\psi_{2,2}$ and $\psi_{4,3}$ the VGG-19 network. The output feature maps corresponding to the reconstruction and ground truth amplitudes are compared. Eq.(14) minimizes the L_1 -norm of the magnitude Fourier spectrum between reconstruction and ground truth.

4 Experiments and Results

4.1 Simulated Dataset

Images in Iowa histology dataset [17] 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 [23], is used to generates low resolution intensity images of 64×64 pixels from these object field samples.

4.2 Our Captured Real Dataset

The experimental setup to capture real data 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, we captured sequential low-resolution images 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.

Before testing our network on real data, we finetune our network on a small set of real 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 [21] 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 and 4 for testing. We show the capture setup and few real captured data in the supplementary.

We implement our neural network using PyTorch [16], and train the same on 2 GTX 1080Ti GPUs of 12GiB capacity, for 100 epochs. The loss function parameters as described in Eq.(3.1) are $(\lambda_1, \lambda_2, \lambda_3) = (0.1, 0.05, 1)$ and $(\beta_1, \beta_2) =$ (0.5, 1). We use Adam optimizer 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, we use 3 stages of the update block as described in Fig 2 for 16-bit and 12-bit depth images, and increase the number of stages to 5 for 8-bit images.

4.3 Comparison on Simulated Data

We compare our simulated and experimental results against existing iterative procedure and deep-learning techniques. Under iterative methods, we have compared with the alternate projections (AP) algorithm proposed by Tian et al. [20]. Another class of iterative method for phase retrieval is the Wirtinger Flow Optimization proposed by [2]. Under data-driven techniques, we consider the conditional deep image prior (cDIP) architecture proposed by Boominathan et al. [4]. We also compare against Fourier-loss cGAN (F-cGAN) architecture provided by Nguyen et al. [14] and modify the input layer dimension of the U-Net architecture to take into account the difference in number of illumination LEDs.

Table 1 shows image quality with two metrics, namely Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity (SSIM) against existing FPM phase retrieval algorithms. The proposed LWGNet is shown to perform at par for the 16-



Fig. 3: 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 bitdepths. Among deep methods, cDIP performs poorly at lower bit depths.

Method	16		12		8		Memory I	nference time	Parameters
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	(in GiB)	(in ms)	(in millions)
AP	24.346	0.835	21.764	0.728	20.500	0.594	NA	NA	NA
WF	23.172	0.769	22.706	0.744	21.621	0.730	NA	NA	NA
cDIP	25.654	0.815	22.563	0.718	20.951	0.712	3.24	5	54.9
F-cGAN	29.021	0.907	$\underline{26.325}$	0.797	$\underline{25.715}$	0.765	3.52	29	7.88
LWGNet	28.745	$\underline{0.829}$	27.726	0.807	26.035	0.802	2.26	341	0.39

Table 1: Compared to previous works, the proposed LWGNet achieves a better reconstruction quality at lower bit-depths using fewer trainable parameters.

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. Although the F-cGAN [14] baseline performs better on 16-bit simulated data, the proposed method performs perceptually better on real data, as shown later in Section 4.4. Fig 3 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 3(b) shows AP-based reconstruction method [20] suffers the most from degradation of peripheral LED measurements at lower bitdepths. Gradient-based methods [2] display poor contrast and blurring at the edges as shown in Fig 3(c). Consequently, this is corrected for in the proposed method as shown in 3(f). cDIP [4] shows checkerboard artifacts that become more prominent with under lower bit depth. Similarly, F-cGAN [14] model successfully preserves finer details at higher bit-depths, but shows blur artifacts with increase in quantization. Additionally, LWGNet involves fewer trainable parameters and requires about 100 epochs of training. LWGNet 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 of fully convolutional networks. The memory and inference time reported in Table 1 are for a 3 stage proposed network.

4.4 Comparison on Real Data

We test our model on real experimental data and compare the results against two deep-learning techniques, cDIP and F-cGAN. Prior to evaluation, we finetune all the methods on real dataset described in Section 4.2. Figure 4 show the visual results for 16, 12 and 8 bit depths respectively. For our experiments, we use histological sample slides, specifically Lung carcinoma and Osteosarcoma (HE stained). 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. However, the proposed algorithm successfully preserves edges and background details compared to the other approaches consistently for all bit depths. The computed gradients of peripheral LED images contain high-frequency details. We hypothesize 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 our proposed method, finetuning on a small dataset of real data doesn't lead to any overfitting. 4.5Sensor Quantization Analysis

In this section, we experimentally verify that the proposed method shows little perceptual variation over reconstruction with increasing quantization noise or decrease in the bit depth of the input data. We finetune our proposed approach using the dataset described in Section 4.2 prior to evaluation. Fig 5 shows reconstruction for Osteosarcoma (top) and Lung carcinoma (bottom) samples at



Fig. 4: Phase reconstruction for *real* 16-bit Lung carcinoma (sCMOS and CMOS), 12-bit Osteosarcoma, and 8-bit Osteosarcoma (CMOS). (a)FoV reconstruction using LWGNet and inset shown in (b). (c) reconstruction obtained from F-cGAN, (d) reconstruction obtained from cDIP, (e) reconstruction obtained from AP, and (f) input brightfield image.

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, we observe minor changes at reconstructions of sharp edge features with the bit-depth setting.

12 Atreyee Saha et al.

Fig. 5: 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.

4.6 Ablation Study

In this subsection, we analyse the impact of different components of our proposed network on simulated data.

Effect of Complex-valued Operations Here we verify the necessity of the complex-valued operations described in Section 3. To do that, we replace the complex-valued operation of our proposed network with real-valued operations acting individually on the real and imaginary channels of the complex gradient. We keep the rest of the architecture and loss functions the same. Top row of Fig 6 presents the visual comparison between reconstructions obtained from complex-valued and real-valued networks. Our experiments show that the use of complex valued operations increases the average PSNR by 5dB.

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

$$\mathfrak{O}_f = \hat{\Psi}(\mathfrak{O}^K) \tag{16}$$

where,

$$\mathfrak{G}^{i+1} = (\mathfrak{G}^i + \frac{1}{L} \sum_{\ell=1}^{L} g^i_{\ell}), i \in \{0, 1, ..., K-1\}$$
(17)

Fig. 6: Ablation experiments: First row shows the using complex-valued operations leads to improved reconstruction . Second row shows that the proposed method of learning the network on gradient leads to effective utilization of gradient information than learning a network on estimated object field. Third row shows that using stochastic gradient improves our results compared to regular gradients.

and $\Psi(.)$ is N numbers of $\Psi^i(.)$ (described in Section 3) stacked together. We used K = 5 iterations of Wirtinger flow and stack N = 3 numbers of $\Psi(.)$ 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 Fig 6 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 a enhancing an object field using a neural network after wirtinger updates.

Effect of Stochastic Gradients In this experiment, we verify the efficacy of the stochastic gradients estimated in Eq.(8). To do that, we train a variant of the proposed network where regular gradients are used instead of the stochastic gradients and each stage is given by,

$$\mathbb{O}^{i+1} = \mathbb{O}^{i} + \Psi^{i}([\nabla_{\mathbb{O}} f^{i}_{\ell}]_{\ell \in L}).$$
(18)

where $\nabla_6 f_\ell^i$ is given by Eq.(6). The rest of the network and loss functions are kept the same as the proposed LWGNet. The third row of Fig 6 shows visual results for this experiment. The use of stochastic gradient clearly outperforms the use of regular gradients.

Fig. 7: 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.

Effect of Number of Unrolled Stages In this section, we analyse the effect of the number of unrolled iterative stages (K in Section 3). Fig ?? represents the variation of PSNR with the number of update blocks in the proposed architecture varied between 2 to 10 blocks. We 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.

5 Discussion and Conclusion

We propose a novel physics based neural network for FPM reconstructions. Our network derives inspiration from conventional wirtinger flow phase retrieval and combines it with data-driven neural networks. Unlike naive wirtinger flow, our network learns a non-linear mapping from stochastic gradients to object field intermediates. We use complex-valued neural networks to learn this non-linear mapping. We perform extensive experiments on simulated and real data to validate the proposed architecture's efficacy. Our method performs at par or even better than existing FPM reconstruction techniques especially for difficult scenarios like low dynamic range. We attribute the success of our network on such challenging conditions to the novel non-linear physics-based mapping. Moreover, we also collect a dataset of real samples using a low cost sensor under different sensor bit depths which will be made public upon acceptance of the paper. The ability to perform high-resolution wide-field of view microscopy using low-cost sensor through physics inspired data-driven techniques can significantly bring down the system cost and is a step towards making point of care diagnostics more accessible.

6 Acknowledgement

We gratefully acknowledge the funding from DST IMPRINT-2 (IMP/2018/001168).

References

- Aidukas, T., Eckert, R., Harvey, A.R., Waller, L., Konda, P.C.: Low-cost, submicron resolution, wide-field computational microscopy using opensource hardware. Scientific reports 9(1), 1–12 (2019)
- Bian, L., Suo, J., Zheng, G., Guo, K., Chen, F., Dai, Q.: Fourier ptychographic reconstruction using wirtinger flow optimization. Opt. Express 23(4), 4856-4866 (Feb 2015). https://doi.org/10.1364/OE.23.004856, http://www.opticsexpress. org/abstract.cfm?URI=oe-23-4-4856
- Bian, L., Zheng, G., Guo, K., Suo, J., Yang, C., Chen, F., Dai, Q.: Motion-corrected fourier ptychography. Biomedical optics express 7(11), 4543–4553 (2016)
- Boominathan, L., Maniparambil, M., Gupta, H., Baburajan, R., Mitra, K.: Phase retrieval for fourier ptychography under varying amount of measurements. arXiv preprint arXiv:1805.03593 (2018)
- Bostan, E., Soltanolkotabi, M., Ren, D., Waller, L.: Accelerated wirtinger flow for multiplexed fourier ptychographic microscopy. In: 2018 25th IEEE International Conference on Image Processing (ICIP). pp. 3823–3827. IEEE (2018)
- Breslauer, D.N., Maamari, R.N., Switz, N.A., Lam, W.A., Fletcher, D.A.: Mobile phone based clinical microscopy for global health applications. PloS one 4(7), e6320 (2009)
- Jiang, S., Guo, K., Liao, J., Zheng, G.: Solving fourier ptychographic imaging problems via neural network modeling and tensorflow. Biomed. Opt. Express 9(7), 3306-3319 (Jul 2018). https://doi.org/10.1364/BOE.9.003306, http://www.osapublishing.org/boe/abstract.cfm?URI=boe-9-7-3306
- Kappeler, A., Ghosh, S., Holloway, J., Cossairt, O., Katsaggelos, A.: Ptychnet: Cnn based fourier ptychography. In: 2017 IEEE International Conference on Image Processing (ICIP). pp. 1712–1716. IEEE (2017)
- Kellman, M., Bostan, E., Chen, M., Waller, L.: Data-driven design for fourier ptychographic microscopy. In: 2019 IEEE International Conference on Computational Photography (ICCP). pp. 1–8. IEEE (2019)
- Kim, J., Henley, B.M., Kim, C.H., Lester, H.A., Yang, C.: Incubator embedded cell culture imaging system (emsight) based on fourier ptychographic microscopy. Biomedical optics express 7(8), 3097–3110 (2016)
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image superresolution using a generative adversarial network (2017)
- Lee, K.C., Lee, K., Jung, J., Lee, S.H., Kim, D., Lee, S.A.: A smartphone-based fourier ptychographic microscope using the display screen for illumination. ACS Photonics 8(5), 1307–1315 (2021)
- 13. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
- Nguyen, T., Xue, Y., Li, Y., Tian, L., Nehmetallah, G.: Deep learning approach for fourier ptychography microscopy. Opt. Express 26(20), 26470-26484 (Oct 2018). https://doi.org/10.1364/OE.26.026470, http://www.opticsexpress.org/ abstract.cfm?URI=oe-26-20-26470
- Ou, X., Horstmeyer, R., Yang, C., Zheng, G.: Quantitative phase imaging via fourier ptychographic microscopy. Opt. Lett. 38(22), 4845–4848 (Nov 2013). https://doi.org/10.1364/OL.38.004845, http://ol.osa.org/abstract.cfm?URI= ol-38-22-4845

- 16 Atreyee Saha et al.
- 16. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024-8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library. pdf
- 17. of Pathology, I.D.: "iowa virtual slidebox", http://www.path.uiowa.edu/ virtualslidebox/
- Smith, Z.J., Chu, K., Espenson, A.R., Rahimzadeh, M., Gryshuk, A., Molinaro, M., Dwyre, D.M., Lane, S., Matthews, D., Wachsmann-Hogiu, S.: Cell-phone-based platform for biomedical device development and education applications. PloS one 6(3), e17150 (2011)
- Song, P., Jiang, S., Zhang, H., Huang, X., Zhang, Y., Zheng, G.: Full-field fourier ptychography (ffp): Spatially varying pupil modeling and its application for rapid field-dependent aberration metrology. APL Photonics 4(5), 050802 (2019)
- 20. Tian, L., Li, X., Ramchandran, K., Waller, L.: Multiplexed coded illumination for fourier ptychography with an led array microscope. Biomed. Opt. Express 5(7), 2376-2389 (Jul 2014). https://doi.org/10.1364/BOE.5.002376, http: //www.osapublishing.org/boe/abstract.cfm?URI=boe-5-7-2376
- 21. Tian, L., Liu, Z., Yeh, L.H., Chen, M., Zhong, J., Waller, L.: Computational illumination for high-speed in vitro fourier ptychographic microscopy. Optica 2(10), 904-911 (Oct 2015), http://www.osapublishing.org/optica/abstract. cfm?URI=optica-2-10-904, https://www.laurawaller.com/opensource/
- Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J.F., Mehri, S., Rostamzadeh, N., Bengio, Y., Pal, C.J.: Deep complex networks (2018)
- Zheng, G., Horstmeyer, R., Yang, C.: Wide-field, high-resolution fourier ptychographic microscopy. Nature photonics 7(9), 739–745 (2013)