Direct Image Classification from Fourier Ptychographic Microscopy Measurements without Reconstruction

Navya Sonal Agarwal[*], Jan Philipp Schneider[†], Kanchana Vaishnavi Gandikota[†],
Syed Muhammad Kazim[‡], John Meshreki[‡], Ivo Ihrke[‡], and Michael Moeller[†]
[*] Kalinga Institute of Industrial Technology, India.
[†] Chair for Computer Vision, University of Siegen, Germany.
[‡] Chair for Computational Sensing and Communications Engineering, University of Siegen, Germany.

Abstract—The computational imaging technique of Fourier Ptychographic Microscopy (FPM) enables high-resolution imaging with a wide field of view and can serve as an extremely valuable tool, e.g. in the classification of cells in medical applications. However, reconstructing a high-resolution image from tens or even hundreds of measurements is computationally expensive, particularly for a wide field of view. Therefore, in this paper, we investigate the idea of classifying the image content in the FPM measurements directly without performing a reconstruction step first. We show that Convolutional Neural Networks (CNN) can extract meaningful information from measurement sequences, significantly outperforming the classification on a single band-limited image (up to 12 %) while being significantly more efficient than a reconstruction of a high-resolution image. Furthermore, we demonstrate that a learned multiplexing of several raw measurements allows maintaining the classification accuracy while reducing the amount of data (and consequently also the acquisition time) significantly.

I. Introduction

Fourier Ptychographic Miscroscopy (FPM) [1] has become increasingly popular in recent years due to its ability to reconstruct super-resolved amplitude and phase images with a wide field of view. This opens up numerous image enhancement and application possibilities in areas like biology and medicine. FPM can be implemented with low hardware costs by replacing the illumination unit of the microscope with an LED array to illuminate the specimen from different lighting angles. However, considering that hundreds of images of 10 or more megapixels need to be combined to a high-resolution gigapixel image via an optimization method, FPM requires large computational resources.

In this paper, we propose to mitigate computational demands in applications where the image content needs to be categorized by directly feeding the measurements, i.e. the images recorded using different illumination directions, into a CNN. We demonstrate that the resulting classification significantly improves over the accuracy obtained on a single on-axis brightfield illumination image only. Our approach can be particularly attractive in applications where small regions of interest are scattered in the field of view, e.g. in cytopathology: It allows identifying regions of interest which can then be reconstructed and further examined by smaller measurement patches. While reconstructions on small measurement patches are common in FPM (c.f. [1]), our work is, to the best of our knowledge, the first to consider the direct classification of direct measurement data.

Furthermore, we propose to incorporate *multiplexing* [2, 3], a technique that combines measurements by illuminating the probe with multiple LEDs simultaneously. We demonstrate that although one has to expect a loss of image quality in FPM reconstruction, the direct classification from raw measurements remains remarkably unaffected by the multiplexing, allowing to reduce the number of recorded images by up to 60% without significant loss in classification accuracy.

II. Background

As the core idea of FPM is to record individual components of the specimen's diffraction spectrum by changing the illumination angle with several LEDs, the data formation model [1]

$$f^k = |\mathcal{F}^{-1}\mathcal{M}^k\mathcal{F}u| + \text{noise}, \quad (1)$$

can be used to represent the image formation of a single band-limited measurement $f^k \in \mathbb{R}^{W \times H}$. Further, $u \in \mathbb{C}$ denotes the true high-resolution specimen to reconstruct, $\mathcal{F}$ denotes the Fourier transform, $\mathcal{M}^k$ is an operator that models the pupil function and the Limiting Aperture (LA). In the simplest case, the operator $\mathcal{M}^k$ can be modeled as a cropping operation that sets all values outside of a circle with LED-position dependent center and a radius proportional to $(2\pi/\lambda)$NA to zero, where λ denotes the illumination wavelength, and NA the Numerical Aperature (NA) of the objective lens. In the absence of a physical setup, the radius and position of the cropping operator $\mathcal{M}^k$ can be chosen arbitrarily, allowing us to simulate various NA sizes, led positions, and alignments. An example of simulated FPM measurements is provided in fig. 1, which illustrates the cropping operation at different regions in Fourier space corresponding to LED positions, and the respective raw FPM measurements.

The data formation model (1) gives rise to an inverse problem frequently being solved by optimizing

$$\arg\min_u \left(\sum_{k=1}^{K} H(f^k, |\mathcal{F}^{-1}\mathcal{M}^k\mathcal{F}u|) + \alpha R(u) \right), \quad (2)$$

for a suitable measure of discrepancy H (e.g. an ℓ^1 norm) and a suitable regularizer R, e.g. the total variation of u. Considering that each measurement f^k commonly is an image of several megapixels and that K commonly is in the order of 100, Eq. (2) is expensive to solve.

To at least reduce the measurement time, the original FPM formulation of [1] was extended in [2] by using multiplexing, as well as learning the specific LED pattern and illumination intensities of the multiplexing in [4]. Such approaches can be modeled by a weighted linear combination of the measurements that one would have obtained from a single LED illumination, i.e.

$$f^k = \sum_l \beta_l^k |\mathcal{F}^{-1}\mathcal{M}_l^k\mathcal{F}u| + \text{noise}, \quad (3)$$

for non-negative weights β_l^k representing the illumination intensity and $\mathcal{M}_l^k$ representing the operators arising from different (simultaneously active) LEDs.

III. NUMERICAL EXPERIMENTS

Considering the high computational demands of first solving (2) and subsequently feeding the reconstructed high-resolution image into a classifier, we investigate the idea of training a classifier on direct measurement data, i.e., on the collection of all f^k stacked into an image with K channels.

We perform synthetic experiments using three popular image recognition datasets, each of which have 10 classes: MNIST [5] with grey scale images of digits of resolution 28×28 pixels, and the color image datasets CIFAR 10 [6] and Imagenette [7] with images resized to resolution 100×100 pixels. We simulate FPM measurements of images from these datasets by applying (1) considering 25 crops in the Fourier space for different values of NA. We conduct experiments using Resnet18 [8] comparing the following scenarios: (i) providing only the central measurement, denoted by CC (central crop) in table I, (ii) providing a stack of all the low-resolution measurements, denoted by SC (stacked crop), (iii) providing the reconstructed images obtained through a gradient descent-based energy minimization scheme using TV regularization, denoted by R, and (iv) providing the ground truth images as an upper bound (UB).

We modify the first layer of the network to reflect the increased number of measurements for both the stacked crop and the multiplexed setting. The networks are trained with a cross-entropy loss using the Adam optimizer [9] with learning rates of $1e-3$ for MNIST, $1e-4$ for CIFAR 10 and Imagenette for 20 epochs. We train networks for CIFAR10 and MNIST from scratch and finetune a pretrained Resnet on Imagenette. We report the classification accuracy of the models at the final epoch.

The results of this experiment are provided in table I. While using a single FPM measurement (CC) alone results in comparatively low accuracy, using all the 25 low-resolution measurements improves the accuracy by 1-1.5% in MNIST, 10-12% in CIFAR10 and 0.5-6% in Imagenette. In our simple experiment, the accuracy of a proper reconstruction (R) almost reaches the upper bound as we are using the same operator for simulation and reconstruction, and assume the prior knowledge of the image to be reconstructed having a zero phase. The fact that the stacked raw-data classification remains inferior to a full reconstruction is to be expected, as learning the backward model of FPM in a purely data-driven way using similar zero-shot supervised formulations is empirically observed to be a challenging task [10]. However, as the reconstruction involves significant computational costs, pre-classification using the direct measurements, as proposed in our SC setting, is very attractive to determine regions of interest. Variations in the cropping radius of the pupil (corresponding to varying limiting apertures) yield surprisingly inconclusive results with SC-accuracies not always being monotonically increasing with increasing pupil radius.

Additionally, we considered simulating multiplexing through multiple-LED illumination by learning a linear combination of measurements implemented using a 1×1 convolution layer whose non-negative weights are learned along with the classification network weights. The results of this experiment are shown in table II, where we simulated multiplexing up to 25 different LEDs (with different intensities) into 5, 10, 15 and 20 measurements, which are provided as inputs to the classification network. The results show only a small reduction in accuracy for low-resolution datasets ($\sim 3\%$ on CIFAR10) even when the number of multiplexed measurements is reduced to 5 in comparison to using all 25 band-limited measurements. With higher number of multiplexed measurements (≥ 10),
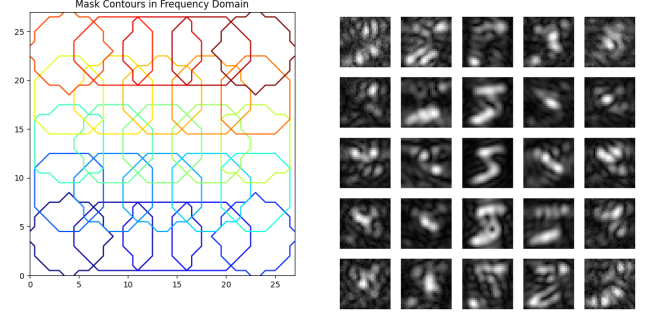

Fig. 1. Illustration of FPM measurements: Each colored contour represents one cropped spectrum in the frequency domain (left) and the corresponding band-limited FPM measurement (right).

the classifier accuracy is maintained. For the Imagenette dataset, we trained the network for 40 epochs instead of 20, as the learned weights for multiplexing were not trained well enough at 20 epochs. For a fair comparison, we repeated the SC and CC also for 40 epochs, which, however, resulted in lower accuracies due to overfitting. Even on Imagenette, we observe that the learned multiplexing results in improved accuracies compared to single measurements, with only a small drop in accuracy, indicating the benefit of learning the illumination pattern to record fewer measurements and further reduce the computational demands.

Dataset	Pupil Radius	Accuracy (%)			
		CC	SC	R	UB
MNIST 28×28	3	97.76	99.33		
	4	98.59	99.35	99.42	99.47
	5	99.06	99.02		
CIFAR10 32×32	3	60.70	72.32		
	4	63.38	75.83	87.27	95.52
	5	65.64	75.14		
Imagenette 100×100	15	75.08	81.2		
	20	77.73	76.99	91.20	93.52
	25	75.62	76.10		

TABLE I
COMPARISON OF ACCURACIES OF A RESNET18 TRAINED ON THE CENTRAL MEASUREMENT (CC), A STACKING OF ALL MEASUREMENTS (SC), ON RECONSTRUCTED IMAGES (R) AND GROUND TRUTH (UB).

Dataset	Radius	SC	CC	Number of multiplexed measuremets			
				5	10	15	20
MNIST	5	99.25	99.06	99.22	99.37	99.50	99.28
CIFAR10	5	75.14	65.64	72.32	75.37	74.88	75.53
Imagenette	15	81.2	75.08	76.99	79.21	78.42	81.07

TABLE II
CLASSIFICATION ACCURACIES USING ALL FPM MEASUREMENTS VS. MULTIPLEXED MEASUREMENTS

IV. CONCLUSIONS

In this paper, we investigated the utility of direct FPM measurements in deep learning-based classification. When the number of band-limited measurements is increased, a performance increase of up to 12 % can be observed compared to a center crop (SC), highlighting the ability of the networks to extract information from the additional frequency spectrum covered. We also found that the number of FPM measurements can be compressed and reduced using a learned linear combination with a negligible loss in classification accuracy compared to the full measurement. These results indicate the potential of directly using FPM measurements in extracting potential regions of interest in Gigapixel scale images, reducing the computational overhead associated with their analysis.

V. ACKNOWLEDGMENTS

REFERENCES

[1] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nature Photonics*, vol. 7, p. 739–745, 09 2013.

[2] L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for fourier ptychography with an led array microscope," *Biomed. Opt. Express*, vol. 5, no. 7, pp. 2376–2389, Jul 2014. [Online]. Available: https://opg.optica.org/boe/abstract.cfm?URI=boe-5-7-2376

[3] M. Kellman, 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, pp. 1–8.

[4] M. R. Kellman, E. Bostan, N. A. Repina, and L. Waller, "Physics-based learned design: optimized coded-illumination for quantitative phase imaging," *IEEE Transactions on Computational Imaging*, vol. 5, no. 3, pp. 344–353, 2019.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[6] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[7] J. Howard. (2019) imagenette. [Online]. Available: https://github.com/fastai/imagenette/

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[10] L. Zhao, X. Zhou, X. Lu, H. Tong, and H. Fang, "Transformer-based reconstruction for fourier ptychographic microscopy," *IEEE Access*, 2023.