



Object detection neural network improves Fourier ptychography reconstruction

FLORIAN STRÖHL,^{1,*}  SUYOG JADHAV,²  BALPREET S. AHLUWALIA,^{1,3}  KRISHNA AGARWAL,¹  AND DILIP K. PRASAD⁴ 

¹Department of Physics and Technology, UiT The Arctic University of Norway, NO- 9037 Tromsø, Norway

²Department of Electronics Engineering, Indian Institute of Technology (IIT ISM), 826004 Dhanbad, India

³Department of Clinical Science, Intervention and Technology, Karolinska Institute, 17177 Stockholm, Sweden

⁴Department of Computer Science, UiT The Arctic University of Norway, NO- 9037 Tromsø, Norway

*florian.strohl@uit.no

Abstract: High resolution microscopy is heavily dependent on superb optical elements and superresolution microscopy even more so. Correcting unavoidable optical aberrations during post-processing is an elegant method to reduce the optical system's complexity. A prime method that promises superresolution, aberration correction, and quantitative phase imaging is Fourier ptychography. This microscopy technique combines many images of the sample, recorded at differing illumination angles akin to computed tomography and uses error minimisation between the recorded images with those generated by a forward model. The more precise knowledge of those illumination angles is available for the image formation forward model, the better the result. Therefore, illumination estimation from the raw data is an important step and supports correct phase recovery and aberration correction. Here, we derive how illumination estimation can be cast as an object detection problem that permits the use of a fast convolutional neural network (CNN) for this task. We find that faster-RCNN delivers highly robust results and outperforms classical approaches by far with an up to 3-fold reduction in estimation errors. Intriguingly, we find that conventionally beneficial smoothing and filtering of raw data is counterproductive in this type of application. We present a detailed analysis of the network's performance and provide all our developed software openly.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Superresolution microscopy received its accolades with the Nobel price in chemistry in 2014 and has since been firmly established as a method of choice in biomedical imaging core facilities. A constant burden that goes along with ever higher resolution is the stark dependence on superb system alignment and performance of the employed optical elements. In practice, this is impossible to guarantee at all times and hence post-acquisition computational aberration correction has seen a rapid development recently [1–6].

Fourier ptychographic microscopy (FPM) [7] stands out from the superresolution family as it is a label-free technique. It is one of the latest microscopy methods to be developed and offers a range of benefits over conventional brightfield imaging. Its main features are (1) retrieval of the optical density of a sample without the need for interferometric detection, (2) correction of optical aberrations induced by the employed optics through recovery of the microscopes complex-valued coherent transfer function [8], (3) imaging with extraordinarily large space-bandwidth product, and (4) its ability to achieve resolution much larger than dictated by the microscope objective lens and thus allow even nanoscopic resolution [9,10].

FPM setups illuminate the sample from a multitude of directions sequentially and capture the scattered light using an objective lens that forms an image of the sample on a camera. Different illumination angles cause different features of the sample to be more pronounced. One can easily

verify this with a flashlight and a relief surface. Mathematically, the sample's complex scattering field gets phase-modulated (a rigorous derivation is provided as supplementary information). This modulation is non-linearly linked to both the pupil function and a sample-dependent phase-delay, typically called quantitative phase or simply phase. The pupil function describes optical aberrations and transmission strength of the imaging system, whereas the phase is a quantitative measure related to the sample's optical density.

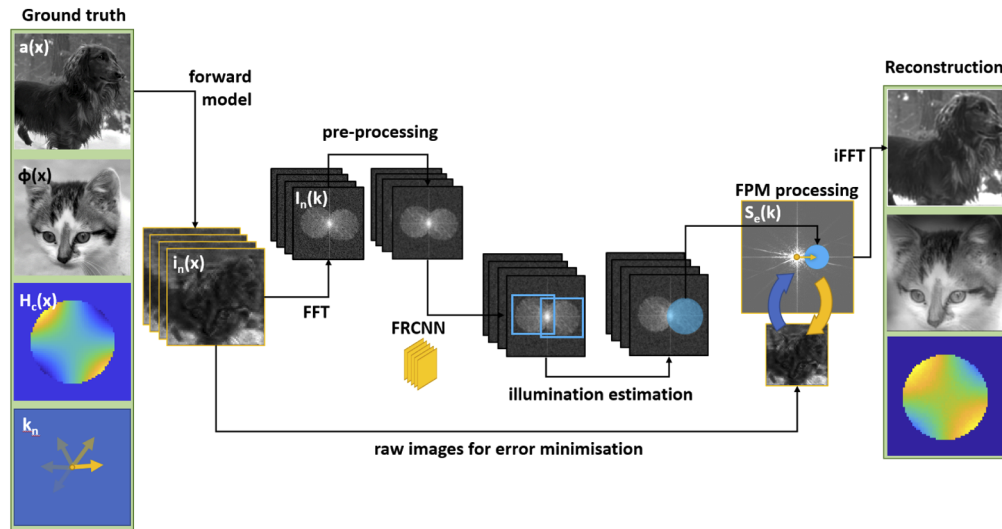


Fig. 1. Flowchart of the calibration and reconstruction process. Raw data is Fourier transformed and (optionally) pre-processed. Then the system parameters are extracted with the neural network (FRCNN) and processed via the FPM phase retrieval algorithm using the alternating projections method. Outputs are the sample's amplitude, phase, and the recovered system- (and patch-) specific pupil function.

To disentangle the pupil and phase components, FPM relies on a reconstruction algorithm (see Fig. 1) that minimises the error between the recorded frames and computer-generated 'raw' images based on a forward model of the imaging process. The parameters to optimise in this process are the sample phase and the pupil function, whereas knowledge about the illumination geometry serves as necessary constraint.

The accurate extraction and calibration of the illumination geometry from the raw data is the main focus of this article and we show how the problem of illumination estimation can be cast as an object detection problem (find and locate an image feature) in the Fourier domain, which permits the use of recently published high-performance object-detection neural networks. In the following, we explain how illumination calibration can be formulated in such a way and proceed by applying and evaluating a suitable neural network to the task. All developed code is freely available on GitHub (github.com/IAMsuyogJadhav/NN-Illumination-Estimation-FPM/).

2. Theory and methods

The illumination of a scattering sample with a plane wave is equivalent to shifting the object *spectrum* by the amount of the lateral illumination wave vector component, which is successively low-pass filtered by the objective's pupil function.

As derived rigorously in the supplementary document the oblique illumination in FPM introduces a spectral shift of the pupil function in the recorded spectrum that is directly proportional to the illumination wave vector (in + and - directions). These shifts are clearly

visible in the raw data and are a dominant feature that is largely independent of the sample (see Fig. 2(b) for an example). Detecting and locating one of these disks in the Fourier domain is thus equivalent to determining the illumination angle. In addition to the illumination angle, it is possible to retrieve the effective coherent cut-off frequency by accurately determining the radius of the displaced pupils. The FPM phase retrieval (with additional embedded pupil function recovery [8]) makes use of this information to disentangle the pupil from the underlying spectrum, which is governed by the combined extent of all raw image spectra. This is commonly implemented as an iteratively solved error minimisation problem (see Supplement 1 for further details). To summarise, the problem of calibrating the brightfield illumination in FPM can be simplified into an object detection problem in which a "noisy disk" needs to be automatically identified and accurately located.

This is a common task in computer vision and can be completed with impressive robustness and fidelity by specialised object-detection neural networks. Generally, neural networks (NNs) are more and more regularly used for computer vision and related machine learning tasks. In microscopy, a whole range of application areas for NNs has emerged [11], including denoising [12,13], digital staining [14–17], counting and labelling [18], tracking [19], image reconstruction [20–22], computational microscopy [23–26], virtual focusing [27,28], aberration estimation [29], and segmentation [30–32]. In the context of FPM, attempts have been made to perform the whole phase retrieval process with a neural network although using neural networks for the full FPM reconstruction pipeline is still an area of active research. So far, such approaches have found little use in practice, due to limited performance increases with respect to the classical phase retrieval [33] or due to impractically long reconstruction times in unsupervised networks [24]. Reliable and fast deep-learning-based phase imaging has been shown to be possible in networks with supervised training [23], but care must be taken when applying such networks to sample types they were not trained on. Moreover, supervised networks require training sets composed of raw data and conventionally reconstructed FPM images, highlighting the need for high-fidelity ground truth reconstructions. The performance thus still hinges on a reliable illumination estimation. Illumination calibration, in contrast, is largely independent of the sample and thus well-suited for neural network processing without training on vast sets of experimental ground truth data.

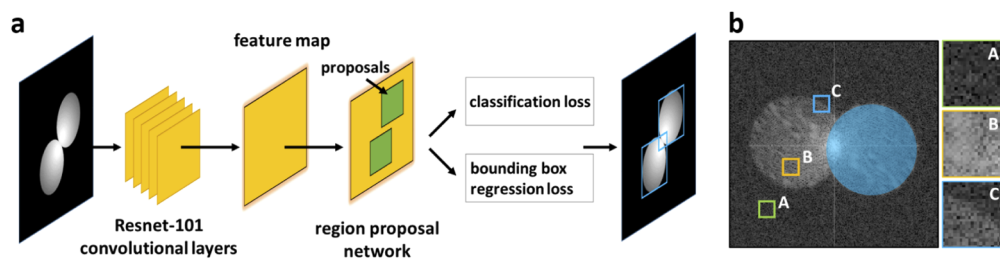


Fig. 2. (a) Block diagram of the Faster-RCNN (FRCNN). (b) Exemplary region propositions (RPs) on an FPM raw data spectrum. Panels (A,B) highlight RPs with limited value for object detection, whereas the RP in panel (C) is more beneficial. The blue disc is detected by the used FRCNN network.

Popular networks for this type of task are two-stage region-proposal convolution neural networks (RCNNs). A fast and very robust implementation of this network type is *faster* RCNN (FRCNN) [34], which is our architecture of choice. Its simplified architecture is shown in Fig. 2(a) to illustrate the two stages.

In brief, the network's first stage consists of the Resnet-101 unit, which contains a sequence of 33 residual blocks of four varieties (referred to as Conv2, Conv3, Conv4, and Conv5 for ease of reference in literature). The first 3 blocks are of type Conv2, the next 4 of Conv3, the next

23 of Conv4, and the remaining 3 of Conv5. Each of these blocks has three convolution layers connected serially and includes a residual link that skips all the three blocks such that the input may be directly passed to the output additively. The architectures of the three convolution layers of Conv2 to Conv5 are described in Table 1. Note that 1×1 convolutions of layer 1 and layer 3 are used to decrease the number of features. This is the most common application of this type of filter and hence these layers are often called feature map pooling layers. The Resnet-101 unit is followed by the region proposal network (RPN) unit. As the name suggests, it creates proposals for regions that likely contain the objects of interest. It takes in the feature map generated by Resnet-101 and simply creates several candidate anchor boxes (or region proposals) at each pixel in the feature map. The anchor boxes pass through a classifier and a regressor in parallel, such that the anchor boxes with good classification accuracy are identified by the classifier, and appropriate coordinates of the bounding boxes in the original image are identified by the regressor. The bounding boxes with good classification accuracy are therefore identified as the outputs.

Table 1. Details of the residual blocks used in Resnet-101.

Residual block	Layer 1	Layer 2	Layer 3
Conv2	64 kernels of 1x1 size	64 kernels of 3x3 size	256 kernels of 1x1 size
Conv3	128 kernels of 1x1 size	128 kernels of 3x3 size	512 kernels of 1x1 size
Conv4	256 kernels of 1x1 size	256 kernels of 3x3 size	1024 kernels of 1x1 size
Conv5	512 kernels of 1x1 size	512 kernels of 3x3 size	2048 kernels of 1x1 size

Our architecture choice is motivated by the following characteristics of our data. Consider the three region proposals (RPs) shown in Fig. 2(b), which may be generated by an object detection approach towards learning bounding box locations. It is evident that RPs A and B are ambiguous in conclusion of foreground or background while RP C may be much more useful towards detection. In this situation, among all the RPs created by an object detection approach, most will be ambiguous though and not useful for learning. Conventional single-stage approaches cannot deal with a poor ratio of useful RPs to those less meaningful. Two-stage approaches, on the other hand, use their second step of classification towards object detection, which allows them to handle even poorer ratios.

We used a pre-trained version of FRCNN [35] with ResNet-101 [36] as backbone that uses dilated convolutions [37] in Conv5 to benefit from transfer learning. This can be justified as low-level features are largely independent of the detection task at hand. Thus, using pre-trained networks shortens the training period for microscopy system-specific features like disks or rings. Furthermore, image *spectra* of natural images are commonly alike (decreasing amplitude with higher spatial frequencies). Furthermore, as microscopy systems are commonly designed to record at the Nyquist limit, the size of the apparent pupil in the raw data spectra relative to the image size is comparable between most microscopes. This limited search space is beneficial, as training on a limited range of cut-off values is sufficient to realise a universally applicable illumination finder. This leads to the conclusion that the application of the network to spectra of microscopy images is possible without the requirement for additional sample-specific training.

We trained the network on the magnitude spectra of computer-generated raw data obtained using the FPM forward model (see Supplement 1 for details). Note that since object detection with conventional NNs is not directly applicable to the complex-valued Fourier domain, we operated only on the Fourier transform *magnitudes*. This step can be justified as the imprint of the pupil function on the spectrum's phase contains only values between $\pm\pi$ and is thus more susceptible to the influence of the sample spectrum than the pupil magnitude.

3. Results and discussion

An example reconstruction using illumination calibration with FRCNN is shown in Fig. 3(a).

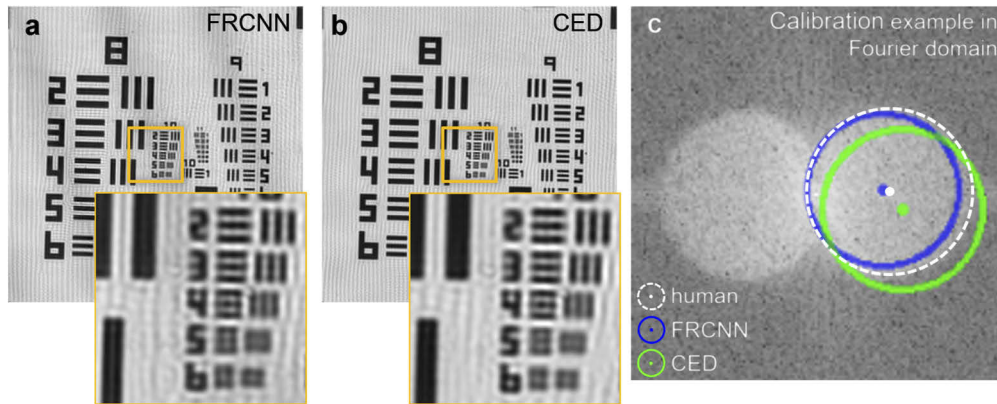


Fig. 3. (a) FPM amplitude reconstruction of a USAF target using illumination calibration with FRCNN. (b) FPM reconstruction using circular edge detection (CED) [38] for calibration. (c) Example of illumination calibration performance in Fourier space showing disk detection by a human operator, FRCNN, and CED. The data set is openly available as part of [38].

The reconstruction quality of FRCNN is on a par with a reconstruction for which illumination calibration was performed with the classical *circular edge detection* (CED) method (github.com/Waller-Lab/Angle_SelfCalibration) [38] (shown in panel b). However, looking at the disk detections in Fourier space (panel c), an improvement in favour of FRCNN can be seen (all frames of this data set are contained in [Supplement 1](#), Figure S4). To objectively quantify the performance of FRCNN with respect to CED, we performed disk detection with both techniques on over 1000 images (128×128 pixels) generated from ground truth images using the FPM forward model. The error in localisation of disks is shown in the violin plot [39] of Fig. 4(a). The violin plots (github.com/bastibe/Violinplot-Matlab) were generated in MATLAB and show each data point as well as an estimate of the probability density of the data. The mean absolute error for CED was 2.4 pixels while the the mean absolute error of FRCNN was 0.9 pixels ($\sim 3\times$ reduction in error).

Intriguingly, we find an almost bimodal distribution in CED, which indicates that a failure of the algorithm might be at times severe. Nevertheless, even the portion of "successful" disk estimations in CED are below the performance of FRCNN. Further, the error spread of FRCNN is smaller and only shows very few outliers, which speaks for its robustness.

Next we investigated the effect of image size. Because of spatially varying pupil aberrations, the imaging forward model can only be considered linear shift-invariant (LSI) in a small image area called an isoplanatic patch. Calibration (and reconstruction) should hence be performed on image sizes not larger than such a patch to fulfill the LSI imaging model to correctly estimate the pupil and illumination. However, a smaller image size makes illumination detection accuracy more difficult as Fourier space pixels become larger and overall less information is available to determine the illumination angle. Hence we compare FRCNNs trained separately on image sizes of 64×64 , 128×128 , 256×256 , and 512×512 pixels to find the best performing model given a certain isoplanatic patch size. When comparing localization error values, in pixels, from different image sizes, the error values from larger image sizes appear larger as a relatively small error in a large image corresponds to more pixels than a 'similar' error in a smaller image. To remove this inherent bias, we use as metric both the error in μm^{-1} as well as in pixels. The conversion formula is

$$\delta d_{\text{pxl}} = Np \delta d_{\mu\text{m}^{-1}}. \quad (1)$$

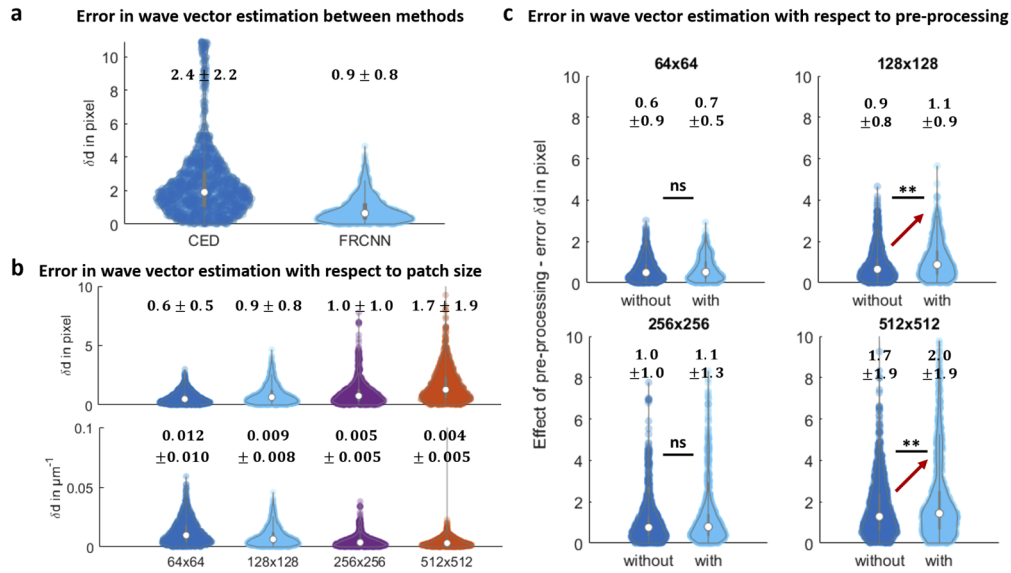


Fig. 4. (a) Violin plots of localization error distribution of CED [38] and FRCNN. (b) Error in terms of pixels and μm^{-1} between different patch sizes. (c) Effect of raw-data filtering on disk localisation performance of FRCNN. The plots show the error values for various patch sizes either without or with pre-processing. Difference between performance on raw and filtered raw data are measured using the Kolmogorov-Smirnoff test with outlier removal via the generalised extreme student deviate test (ns = not significant, ** = null hypothesis (no difference between distributions) rejected below the 1% significance level).

In Eq. (1), p is the effective camera pixel size in the sample plane. Since the conversion factor between pixels (δd_{pxl}) and μm ($\delta d_{\mu\text{m}^{-1}}$) scales with image size N , we get error values that conform to the same scale and thus can be compared over different patch sizes. As illustrated in Fig. 4(b), we find that increasing patch size reduces the error in terms of physical parameter (μm^{-1}), while in terms of pixel accuracy a larger patch size increases the error. The lower limit for necessary wave vector estimation precision is determined by the degree of spatial coherence of the illumination light [38] while the maximal image size is determined by the isoplanatic patch size. Therefore, it is possible to choose the image size that offers both a suitable isoplanatic patch size and achieves high illumination estimation precision. Note that we also used CED on these patch sizes, which performed always at least three times less accurate than FRCNN (results contained in Supplement 1).

Thirdly, we explored the effect of pre-processing of the *Fourier spectra* before feeding them to the neural network. In standard machine learning tasks, pre-processing of the raw data improves results and smoothing and denoising an image is generally deemed beneficial for algorithms to detect image features. On the other hand altering the image spectrum might obscure the actual location of the shifted pupil. The effects of pre-processing on the disk center estimation are summarised in the violin plots in Fig. 4(c), which compares the localisation error (in pixels) between un-processed raw spectra and filtered spectra using the full pre-processing pipeline. We find a limited or even at times adverse effect of pre-processing (mirrored by the violin plots of localisation errors in Fig. 4(c)). A detailed analysis (see Supplement 1) shows that bilateral filtering can have a small positive effect (yet not significant) on some patch sizes, whereas other filters for image smoothing can be highly disadvantageous. In contrast to most deep-learning application in computer vision, our experiments thus indicate that pre-processing provides no

advantage when applied to object detection tasks in Fourier space, but might indeed worsen the performance while further increasing computation time.

Lastly, we investigated the generalisability of illumination calibration via neural networks. We used both conventional refractive objective data ("normal") and reflective objective data possessing a prominent obscuration ("obscuration") as input of three differently trained FRCNNs. Network 1 had seen only "normal" objective data, network 2 was trained on reflective objective data only, and network 3 was trained on both types of data. Note that the network architecture was the same in all cases and only the obscuration was modelled additionally in the forward model for networks 2 and 3. The patch size was 256x256 and no Fourier space pre-processing was applied. As is evident in Table 2, the presence of the obscuration significantly worsened the performance of the network that had never seen it during training.

Vice versa, the network trained on obscuration data had a steep decline in calibration accuracy when applied to "normal" data. Two other observations can be made though: Firstly, the presence of an obscuration is beneficial for illumination calibration (the overall error is smaller). We assume that any additional feature of the pupil would have this effect as more useful RPs would be present for the network to work with. Secondly, the performance of a more broadly trained network is largely on a par with a specialised one. Intriguingly, this extension of illumination calibration to reflective objectives only required a small adaption of the forward model, which would not have been feasible in such a simple and straightforward way with conventional approaches like CED.

Table 2. Mean error of center estimation (in pixels) for three differently trained NNs.

Tested on	Trained on		
	normal	obscuration	both
normal	1.02	4.51	1.01
obscuration	7.81	0.51	0.69

4. Summary

It is interesting that illumination estimation can be posed as an object detection problem which is a common deep learning task, albeit not in Fourier space where it has not been conventionally tried to the best of our knowledge. We observe that use of faster RCNN for illumination estimation shows improvements over traditional methods like CED [38] with a 3-fold reduction in disk localisation error. Further, deep learning allows us to design tailor-made algorithms unique to particular microscopy setups with different isoplanatic patch sizes. The increased degree of abstraction in neural networks further eliminates the need for devising dedicated feature detection routines for distinct microscopy setups - the network can adapt to any pupil shape, as for example found in reflective microscope objectives. This can also help mitigate the loss of precision usually observed when an algorithm is tried on a type of data that is substantially different from the data it was designed to be used on. This also renders our approach highly user-friendly, as it is free from user-set parameters for successful illumination estimation. We additionally investigated the effect of pre-processing. Contrary to common knowledge in the context of many other image processing tasks involving neural networks, where such pre-processing proves very useful, we found that it is less viable on small image patch sizes in FPM illumination calibration. Finally, with the progress made in computational hardware in recent years, deep learning showed to be computationally feasible and could provide much more precise estimations with less computational overhead compared to CED by a factor of 2.

Looking ahead, using neural networks for the full FPM reconstruction pipeline is also an area of active research. An interesting approach in this respect is the combination of classical reconstruction routines and neural networks, where only the first FPM image of a time-series is reconstructed classically and serves as the sole training set for a neural network [40]. Given a

long enough time-series, training of the network and application to consecutive frames is then considerably faster than classical reconstruction of each frame, while maintaining equivalent image quality. Moreover, as fewer raw frames are required for reconstruction with neural networks, the overall frame-rate can be increased tremendously. In the end though, it hinges on a reliable illumination estimation.

5. Author contributions

FS conceived and supervised the project, implemented the forward model, analysed data, and wrote the manuscript. SJ implemented and trained the networks, and performed simulations. DKP provided expertise on neural networks. KA and BSA provided guidance and research tools. All authors assisted in writing the manuscript.

Funding

Publication fund of UiT The Arctic University of Norway; Norges Forskningsråd (285571); European Research Council (336716, 804233, 836355).

Acknowledgements

The authors would like to thank Deanna L. Wolfson and Ida S. Opstad for provision of fluorescence, brightfield, and phase microscopy images that were used as part of the data set for ground truth generation in the FPM forward model. Further, we would like to thank the publication fund of UiT The Arctic University of Norway for covering the publication charges for this article.

Disclosures

The authors declare no conflicts of interest.

See [Supplement 1](#) for supporting content.

References

1. F. Xu, D. Ma, K. P. MacPherson, S. Liu, Y. Bu, Y. Wang, Y. Tang, C. Bi, T. Kwok, A. A. Chubykin, P. Yin, S. Calve, G. E. Landreth, and F. Huang, "Three-dimensional nanoscopy of whole cells and tissues with in situ point spread function retrieval," *Nat. Methods* **17**(5), 531–540 (2020).
2. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica* **5**(4), 458 (2018).
3. J. Demmerle, C. Innocent, A. J. North, G. Ball, M. Müller, E. Miron, A. Matsuda, I. M. Dobbie, Y. Markaki, and L. Schermelleh, "Strategic and practical guidelines for successful structured illumination microscopy," *Nat. Protoc.* **12**(5), 988–1010 (2017).
4. S. V. Koho, E. Slenders, G. Tortarolo, M. Castello, M. Buttafava, F. Villa, E. Tcarenkova, M. Ameloot, P. Bianchini, C. J. Sheppard, A. Diaspro, A. Tosi, and G. Vicidomini, "Two-photon image-scanning microscopy with spad array and blind image reconstruction," *Biomed. Opt. Express* **11**(6), 2905–2924 (2020).
5. L.-H. Yeh, S. Chowdhury, and L. Waller, "Computational structured illumination for high-content fluorescence and phase microscopy," *Biomed. Opt. Express* **10**(4), 1978 (2019).
6. M. Chen, Z. F. Phillips, and L. Waller, "Quantitative differential phase contrast (DPC) microscopy with computational aberration correction," *Opt. Express* **26**(25), 32888 (2018).
7. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field high-resolution fourier ptychographic microscopy," *Nat. Photonics* **7**(9), 739–745 (2013).
8. X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for fourier ptychographic microscopy," *Opt. Express* **22**(5), 4960–4972 (2014).
9. F. Ströhl, I. S. Opstad, J.-C. Tinguely, F. T. Dullo, I. Mela, J. W. Osterrieth, B. S. Ahluwalia, and C. F. Kaminski, "Super-condenser enables label-free nanoscopy," *Opt. Express* **27**(18), 25280–25292 (2019).
10. C. Pang, J. Li, M. Tang, J. Wang, I. Mela, F. Ströhl, L. Hecker, W. Shen, Q. Liu, X. Liu, Y. Wang, H. Zhang, M. Xu, X. Zhang, X. Liu, Q. Yang, and C. F. Kaminski, "On-chip super-resolution imaging with fluorescent polymer films," *Adv. Funct. Mater.* **29**, 1900126 (2019).

11. L. von Chamier, J. Jukkala, C. Spahn, M. Lerche, S. Hernández-Pérez, P. K. Mattila, E. Karinou, S. Holden, A. C. Solak, A. Krull, T.-O. Buchholz, F. Jug, L. A. Royer, M. Heilemann, R. F. Laine, G. Jacquemet, and R. Henriques, “Zerocostd4mic: an open platform to simplify access and use of deep-learning in microscopy,” *bioRxiv* (2020).
12. A. Krull, T.-O. Buchholz, and F. Jug, “Noise2void-learning denoising from single noisy images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2019), pp. 2129–2137.
13. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, “Content-aware image restoration: pushing the limits of fluorescence microscopy,” *Nat. Methods* **15**(12), 1090–1097 (2018).
14. M. E. Kandel, Y. R. He, Y. J. Lee, T. H.-Y. Chen, K. M. Sullivan, O. Aydin, M. T. A. Saif, H. Kong, N. Sobh, and G. Popescu, “Pics: Phase imaging with computational specificity,” *arXiv preprint arXiv:2002.08361* (2020).
15. Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, “Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning,” *Light: Sci. Appl.* **8**(1), 23 (2019).
16. Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydin, J. E. Zuckerman, T. Chong, A. E. Sisk, L. M. Westbrook, W. D. Wallace, and A. Ozcan, “Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning,” *Nat. Biomed. Eng.* **3**(6), 466–477 (2019).
17. C. Ounkomol, S. Seshamani, M. M. Malekar, F. Collman, and G. R. Johnson, “Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy,” *Nat. Methods* **15**(11), 917–920 (2018).
18. T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, and O. Ronneberger, “U-net: deep learning for cell counting, detection, and morphometry,” *Nat. Methods* **16**(1), 67–70 (2019).
19. S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, and A. Kreshuk, “ilastik: Interactive machine learning for (bio) image analysis,” *Nat. Methods* **16**(12), 1226–1232 (2019).
20. C. N. Christensen, E. N. Ward, P. Lio, and C. F. Kaminski, “MI-sim: A deep neural network for reconstruction of structured illumination microscopy images,” *arXiv preprint arXiv:2003.11064* (2020).
21. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydin, L. A. Bentolila, C. Kural, and A. Ozcan, “Deep learning enables cross-modality super-resolution in fluorescence microscopy,” *Nat. Methods* **16**(1), 103–110 (2019).
22. W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, “Deep learning massively accelerates super-resolution localization microscopy,” *Nat. Biotechnol.* **36**(5), 460–468 (2018).
23. Y. Xue, S. Cheng, Y. Li, and L. Tian, “Reliable deep-learning-based phase imaging with uncertainty quantification,” *Optica* **6**(5), 618–629 (2019).
24. K. C. Zhou and R. Horstmeyer, “Diffraction tomography with a deep image prior,” *Opt. Express* **28**(9), 12872–12896 (2020).
25. H. Zhang, C. Fang, X. Xie, Y. Yang, W. Mei, D. Jin, and P. Fei, “High-throughput high-resolution deep learning microscopy based on registration-free generative adversarial network,” *Biomed. Opt. Express* **10**(3), 1044–1063 (2019).
26. E. Nehmei, L. E. Weiss, T. Michaelii, and Y. Shechtman, “Deep-storm: super-resolution single-molecule microscopy by deep learning,” *Optica* **5**(4), 458–464 (2018).
27. Y. Wu, Y. Rivenson, H. Wang, Y. Luo, E. Ben-David, L. A. Bentolila, C. Pritz, and A. Ozcan, “Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning,” *Nat. Methods* **16**(12), 1323–1331 (2019).
28. H. Pinkard, Z. Phillips, A. Babakhani, D. A. Fletcher, and L. Waller, “Deep learning for single-shot autofocus microscopy,” *Optica* **6**(6), 794–797 (2019).
29. E. Bostan, R. Heckel, M. Chen, M. Kellman, and L. Waller, “Deep phase decoder: Self-calibrating phase microscopy with an untrained deep neural network,” *arXiv preprint arXiv:2001.09803* (2020).
30. M. G. Haberl, C. Churas, L. Tindall, D. Boassa, S. Phan, E. A. Bushong, M. Madany, R. Akay, T. J. Deerinck, S. T. Peltier, and M. H. Ellisman, “Cdeep3m–plug-and-play cloud-based deep learning for image segmentation,” *Nat. Methods* **15**(9), 677–680 (2018).
31. A. Arbellé and T. R. Raviv, “Microscopy cell segmentation via adversarial neural networks,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, (2018), pp. 645–648.
32. A. Sekh, I.-S. Opstad, A. Birgisdottir, T. Myrnel, B. Ahluwalia, K. Agarwal, and D. K. Prasad, “Learning nanoscale motion patterns of vesicles in living cells,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), pp. 1–10.
33. L. Boominathan, M. Maniparambil, H. Gupta, R. Baburajan, and K. Mitra, “Phase retrieval for fourier ptychography under varying amount of measurements,” *arXiv preprint arXiv:1805.03593* (2018).
34. S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems*, (2015), pp. 91–99.
35. S. Ren, K. He, R. B. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017).
36. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR abs/1512.03385* (2015).
37. F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” *arXiv preprint arXiv:1511.07122* (2015).
38. R. Eckert, Z. F. Phillips, and L. Waller, “Efficient illumination angle self-calibration in fourier ptychography,” *Appl. Opt.* **57**(19), 5434–5442 (2018).

39. J. L. Hintze and R. D. Nelson, "Violin plots: a box plot-density trace synergism," *The Am. Stat.* **52**, 181–184 (1998).
40. N. Thanh, Y. Xue, Y. Li, L. Tian, and G. Nehmetallah, "Deep learning approach to fourier ptychographic microscopy," *Opt. Express* **26**(20), 26470–26484 (2018).