Title Page

Title	O-Net: A fast and precise deep learning architecture for computational super-resolved phase-modulated optical microscopy
Running Head	A deep-neural network (DNN) based framework for super-resolving phase contrast and DIC images with existing optics
Authors	Shiraz S. Kaderuppan[1,2], Wai Leong Eugene Wong[1,2,3], Anurag Sharma[1,2,3] and Wai Lok Woo[2,4]
Authors' Institutional Affiliations	[1]Newcastle Research and Innovation Institute Pte Ltd, (NewRIIS), Devan Nair Institute for Employment and Employability, Singapore 609607 [2]Faculty of Science, Agriculture and Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K. [3]Newcastle University in Singapore, SIT Building @ Nanyang Polytechnic, Singapore 567739 [4]Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K.
Primary Institution for Research	Newcastle Research and Innovation Institute Pte Ltd, (NewRIIS), Devan Nair Institute for Employment and Employability, Singapore 609607
Corresponding Author	Shiraz S. Kaderuppan
Corresponding Author's Mailing Address	Newcastle Research and Innovation Institute Pte Ltd, (NewRIIS), Devan Nair Institute for Employment and Employability, Singapore 609607

Corresponding Author's E-mail	E-mail: S.S.O.Kaderuppan@newcastle.ac.uk

O-Net: A fast and precise deep learning architecture for computational super-resolved phase-modulated optical microscopy

Abstract

We present a fast and precise deep-learning architecture, which we term O-Net, for obtaining super-resolved images from conventional phase-modulated optical microscopical techniques, such as phase contrast microscopy (PCM) and differential interference contrast (DIC) microscopy. O-Net represents a novel deep convolutional neural network that can be trained on both simulated or experimental data, the latter of which is being demonstrated in the present context. The present study demonstrates the ability of the proposed method to achieve super-resolved images even under poor signal-to-noise ratios and does not require prior information on the point spread function or optical character of the system. Moreover, unlike previous state-of-the-art deep neural networks (such as U-Nets), the O-Net architecture seemingly demonstrates immunity to network hallucination, a commonly cited issue caused by network overfitting when U-Nets are employed. Models derived from the proposed O-Net architecture are validated through empirical comparison with a similar sample imaged via scanning electron microscopy (SEM) (the gold standard in high-resolution microscopical imaging) and are found to generate ultra-resolved images which came close to that of the SEM micrograph.

Introduction

Convolutional deep neural networks (DNNs) have previously been characterized [according to Ronneberger *et al.* (2015), Girshick *et al.* (2014) and Krizhevsky *et al.* (2012)] to play a vital role in feature detection and object recognition in computer vision applications. Nonetheless, DNNs only rose to prominence in these aspects during the recent decade, despite their

presence in the field of machine learning for a considerably long period of time (LeCun *et al.*, 1989) – a bottleneck primarily attributable to the large amounts of training data required (Sze *et al.*, 2017) coupled with the need for wide *and* deep networks to realize potential gains in DNN-derived results. Circumvention of this hindrance may be postulated to be partly driven by the recent and ongoing developments in computing hardware and multicore GPU architectures such as the GeForce RTX 30 Series (n.d.) & AMD Radeon™ RX Graphics Cards (n.d.), coupled with the widespread availability of cloud computing platforms [e.g. AWS Deep Learning AMIs (n.d.), Welcome To Colaboratory (n.d.) & IBM Watson products (n.d.)], which have resulted in declining costs and increased accessibility of the masses to such advanced high-performance computing (HPC) platforms. Consequently, this has allowed for the proliferation of new DNN architectures such as R-CNN (Girshick *et al.*, 2014), AlexNet (Krizhevsky *et al.*, 2012) & ResNet (He *et al.*, 2016) for the realization of goals previously unfathomable in research, coupled with the increasing deployment of artificial intelligence (AI) algorithms in industries as diverse as healthcare [where DNNs are being exploited to identify diseased tissue for cancer diagnosis (Zhongyi *et al.*, 2017), amongst others], education [for student profiling, assessment, lesson personalization and tutoring, a field described as Artificial Intelligence in Education (AIEd) in (Zawacki-Richter *et al.*, 2019)], and even in the financial and e-commerce sectors [to predict the likelihood of a loan being defaulted (Bayraci & Susuz, 2019) or cross-selling of products (Cohn, 2015)]. Considering the ubiquitous prevalence of DNNs & their widespread applications in various sectors in the current economic landscape, it would be apt to dive deeper into the technical underpinnings underlying these DNNs, in particular U-Nets (Ronneberger *et al.*, 2015), which represent one of the most widely utilized DNN architectures today.

U-Nets are a popular DNN architecture often utilized in feature detection and identification, with a couple of prominent commercially available applications being Artificial Intelligence for 3D Visualization and Analysis Software (n.d.) & AIVIA (n.d.) (amongst others). For instance, in AIVIA (n.d.), U-Nets, Generative Adversarial Networks (GANs) and Residual Channel Attention Networks (RCANs) may be harnessed to segment 3D electron micrographs of neurons, while also aiding in the identification and categorization of specific cell types (based on key structural features) and suggesting new phenotypes (where these are not explicitly described by the user). Moreover, U-Nets have also shown great promise in numerous international deep learning competitions, such as the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks and the Cell Tracking Challenge at ISBI 2015, amongst others (Ronneberger *et al.*, 2015). A figure illustrating the U-Net architecture is depicted as follows:

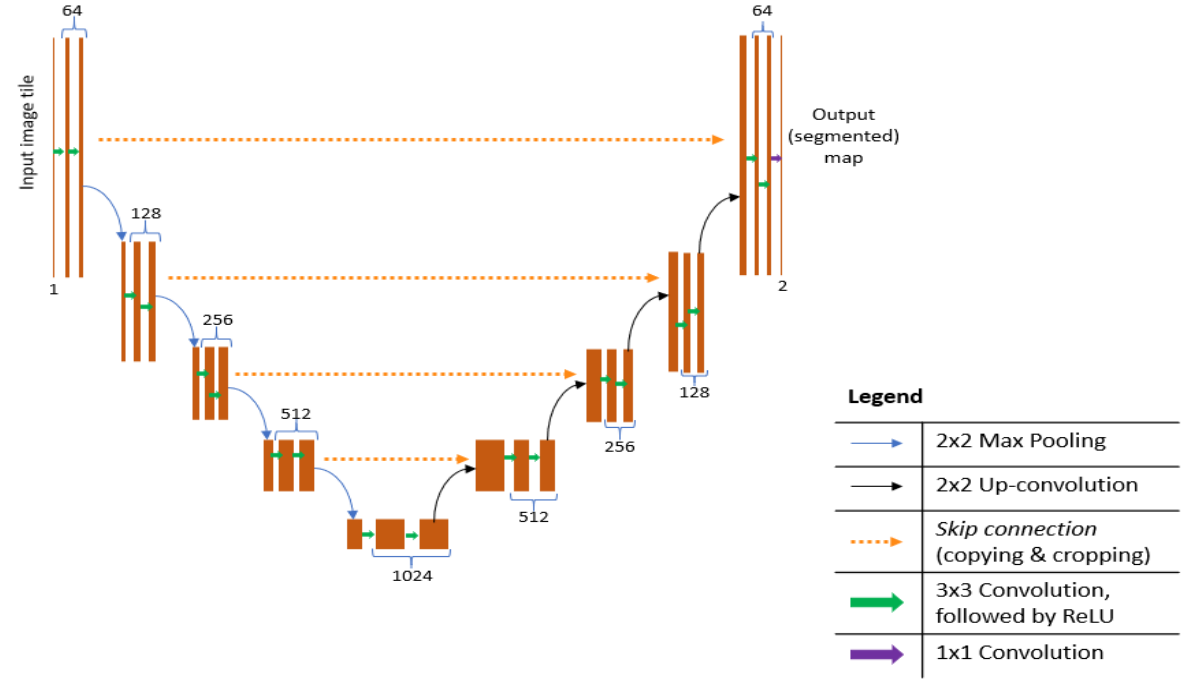


Figure 1: An example of the U-Net architecture as described in Ronneberger *et al.* (2015). Notice the presence of a descending limb, comprising of convolution & max pooling

(*downsampling*) operations and an ascending limb, comprising of transposed convolution (*upsampling*) operations. Skip-connections are also present to allow the output of each convolution layer to transit immediately to its corresponding transposed convolution layer. Figure adapted from Ronneberger *et al.* (2015).

In addition to being utilized for object detection & identification in images, U-Nets have also been successfully exploited in super-resolving images acquired through epifluorescence microscopy. Some seminal work done in this field include Wang *et al*. (2019) and Ouyang *et al*. (2018), the latter developing a modified U-Net architecture (which they termed A-Net). Nonetheless, the U-Net architecture poses a ***severe*** limitation in the field of *in silico* super-resolution (SR) imaging, namely the prediction of non-existent features in the image, resulting in the generation of artifacts – a dilemma characterized by Belthangady & Royer (2019) as *network hallucination*. It was for this reason that Hoffman *et al*. (2021) and Akst (2018) highlighted significant scepticism expressed on the reliability of the results obtained through the application of GANs in super-resolving fluorescence microscopical images. We postulate that this issue may be attributed to the discard of valuable information encoded by the pixels in an image, triggered by the initial use of the convolution function in the U-Net architecture. Originally, this was regarded and proposed by the U-Net developers as a *structural consideration* for capturing context in the training dataset (Ronneberger *et al.*, 2015). U-Net was thus developed to remove redundancy in a dataset, compressing it to *only* the key features required for the corresponding deep learning task via repeated mathematical convolutions until a bottleneck was reached. Subsequently, the network used up-sampling (*transpose convolution*) to expand the dataset thereby including additional information into the learnt dataset. The U-Net developers deemed this to achieve precise localization

(Ronneberger *et al.*, 2015), although image up-sampling through the introduction of pixels into an image (as in the present context of SR imaging) would result in spurious/falsified image details being incorporated into the resultant image. In this regard, we propose an alternative DNN architecture (termed O-Net) as a more accurate means of achieving image super-resolution in phase contrast & DIC microscopies with minimal network hallucinations. Figure 2 below illustrates our proposed O-Net architecture:

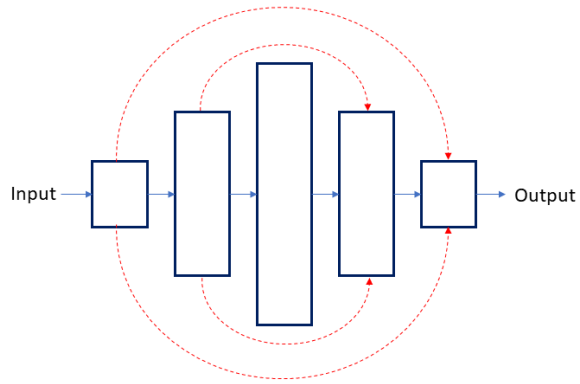


Figure 2: A semblance of the O-Net architecture, which we are proposing in the present context. Contrary to the U-Net architecture (Ronneberger *et al.*, 2015) as shown in Figure 1 previously, O-Net seeks to expand the dataset initially (thereby increasing its sparsity) before compressing it. As with U-Net (Ronneberger et al., 2015), skip-connections may also be present in O-Net (exemplified by the red arrows in the Figure) to allow the output of each transposed convolution layer to transit immediately to its corresponding convolution layer.

In the subsequent sections, we show that O-Net holds significant promise for *in silico* super-resolution microscopy, as evidenced by its ability to create models capable of super-resolving the poroids in the striae of *P. dactylus* var. *dariana*, which are generally separated by a distance of 72nm – 89nm (Barone, n.d.). In addition, we also demonstrate that O-Net holds significant potential for extrapolations of its depth and scope in the near future, fuelled by the ongoing technological advancements in GPU and RAM capacities.

Materials and Methods

1. Data Acquisition & Preparation (Light Microscopy)

The data source used for training the networks comprised of images which were gleaned from a variety of different commercially-prepared biological samples, using a Leica N PLAN L 20X/0.4 Corr Ph1 objective (Leica P/N: 506058) [for acquisition of low-resolution (LR) images] and a Leica HCX PL Fluotar L 40X/0.60 Corr Ph2 objective (Leica P/N: 506203) [for acquisition of high-resolution (HR) images] on a Leica DM4000M microscope equipped with a CMOS camera (RisingCam® E3ISPM12000KPA, RisingTech) having a pixel size of 1.85µm x 1.85µm, an EK 14 Mot motorized stage (Märzhäuser Wetzlar GmbH & Co. KG) and a self-developed stage controller. Identical regions of interest (ROIs) from the samples were imaged under both diascopic phase contrast and differential interference contrast (DIC) microscopy, and the acquired images registered, cropped, and subsequently split into 256x256 RGB tiles using MATLAB R2020a (© 1984-2020, The MathWorks, Inc). Shifts encountered in the ROIs (where evident) were resolved through multi-layer image cropping using Corel PHOTO-PAINT X7 (© 2015 Corel Corporation) post-registration in MATLAB R2020a (© 1984-2020, The MathWorks, Inc), prior to being split into 256x256 RGB image tiles in MATLAB R2020a (© 1984-2020, The MathWorks, Inc). The image tiles formed were then concatenated in LR-HR image pairs for *each* imaging modality used (with similar ROIs from each objective being paired), resulting in a total of 3944 image pairs for each dataset, before being transformed into separate NumPy arrays for training the respective Pix2Pix networks in Python 3.8 (2 separate image datasets were thus formulated – one each for DIC and PCM respectively, for training the individual network architectures). Here, the LR images constituted the **Source** images (to be transformed by the network), while the HR images represented the **Expected** (*target*) images, which the network utilized as ground truth for the training.

Test images were acquired from a prepared slide of *Pinnularia dactylus* var. *dariana* (*P. dactylus* var. *dariana*) (A.W.F.Schmidt) Cleve 1895 (made at Diatom Shop, Diatom Lab) using a Leica HCX PL Apo 100X/1.4 Oil Ph3 CS objective (Leica P/N: 506211) on the same imaging setup, and the models generated previously used to super-resolve these images for comparison against a SEM standard. The *P. dactylus* var. *dariana* slide was used as a test slide in this context, as the spacing between the poroids was 72nm - 89nm apart (Barone, n.d.), which was considerably smaller than the diffraction limit of the optical microscope, i.e. ~200nm (Kaderuppan *et al.*, 2020). Being able to resolve these poroids thus provided conclusive evidence on the ability of these networks to achieve label-free computational nanoscopy.

2. Pix2Pix GAN Architecture

Pix2Pix represents a type of conditional generative adversarial network (cGAN) architecture (Isola *et al.*, 2017) which incorporates a generator and discriminator network depicted simply in Figure 3:

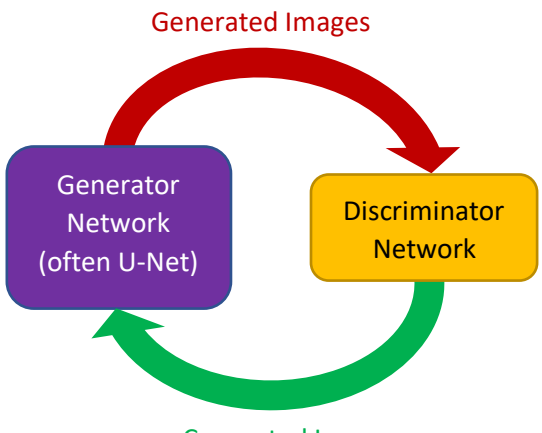


Figure 3: A simplified schematic of the Pix2Pix GAN/conditional GAN (cGAN) architecture as described in Isola *et al.* (2017). A constant flow of model generation, refinement & image comparison via two distinct networks (the generator & discriminator) is being shown here, where the images generated are compared by the

discriminator against the ground truth (*expected*) images. The computed error is fed back to the generator to refine the layers responsible for generating the said image, thereby improving model accuracy. The process continues, until all training epochs are completed.

Here, the generator attempts to devise simulated images in order to successfully confound the discriminator into conceptualizing the generated image as the desired target image. According to pix2pix: Image-to-image translation with a conditional GAN (2021), the generator utilizes a (i) sigmoid (binary) cross-entropy loss $\in$ between the fabricated images and an array of ones, (ii) the mean absolute error (MAE) loss ℓ_1 between the ground truth and generated images, as well as (iii) a constant λ assigned a value of 100, the latter two being implemented by the developers of this architecture (Isola *et al.*, 2017) as well. The equations describing each of these losses, as well as their composition for derivation of an overall loss $\mathcal{G}$ are defined as follows:

$$\in = \frac{\sum_{i=1}^{n}[(b_i-1)\log(1-\hat{b}_i)-b_i\log\hat{b}_i]}{n} \qquad (1)$$

where b_i is the label and $\hat{b}_i$ represents the probability of b_i = 1 [derived from (Murphy K. , 2012) & (Godoy, 2018)] (this is utilized in the discriminator losses d_R & d_G as well),

$$\text{MAE } (\ell_1) \text{ loss} = \frac{\sum_{i=1}^{n}|a_i-b_i|}{n} \qquad (2)$$

where n is the number of pixels in the image, a_i refers to the target value & b_i refers to the estimated value of the assayed parameter [e.g. pixel RGB (or HSL) intensities at pixel i] and

$$\text{Total generator loss, } \mathcal{G} = \in + (\lambda \bullet \ell_1) \qquad (3)$$

where λ = 100 (pix2pix: Image-to-image translation with a conditional GAN, 2021).

In most instances, the generator is based on a U-Net architecture (Ronneberger *et al.*, 2015), while the discriminator utilizes a convolutional PatchGAN classifier (Isola *et al.*, 2017). The U-Net generator consists of individual encoder blocks in the descending (*downsampling*) arm

linked (via *skip connections*) to decoder blocks in the ascending (*upsampling*) arm, each block in the former comprising convolution, batch normalization & Leaky ReLU operations, while the latter blocks are composed of transposed convolution, batch normalization, dropout (for the first 3 layers) & ReLU operations (as shown in Figure 1 previously) (Ronneberger *et al.*, 2015).

In the present study, a Pix2Pix GAN incorporating a generator based on our proposed O-Net was utilized to achieve image super-resolution, and this was contrasted against the traditional U-Net GAN. For performance comparison between the O-Net and U-Net DNN architectures, similar activation functions, number & size of convolution kernels (as well as the number of encoder & decoder blocks) were used for both the generator and discriminator networks. In addition, both models were generally trained more than 95 epochs. Specific details on the U-Net & O-Net architectures assimilated in the present study are described in the accompanying Supplement – (i) **Supplementary Tables ST1**, **ST2** & **ST3** and **Supplementary Figures SF1**, **SF2** & **SF3** (for U-Net), and (ii) **Supplementary Tables ST4**, **ST5** & **ST6** and **Supplementary Figures SF4**, **SF5** & **SF6** (for O-Net).

3. O-Net Architecture

Our proposed framework (O-Net) incorporates upsampling (*transposed convolution*) operations in the initial input segment of the network, followed by downsampling (*convolution*) operations in the subsequent arm of the network. In the present study, we developed a 5-layer O-Net architecture, described as follows:

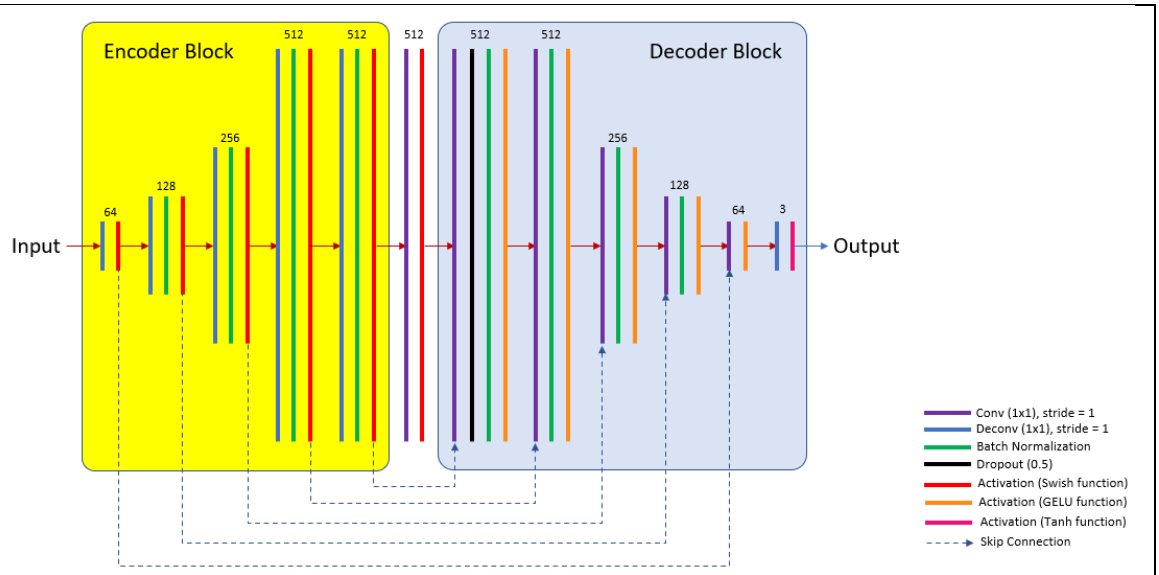


Figure 4: The O-Net architecture as utilized in the current study and contrasted with the U-Net architecture (depicted in Figure 1 previously). Here, an encoder block (consisting of 5 transposed convolution layers) is concatenated with a decoder block (having 5 convolution layers). Skip-connections (*concatenations*) are also present to link conjugate layers in the encoder block with those in the decoder block. A key difference between the U-Net and O-Net architectures lies in the constitution of the encoder and decoder blocks – in U-Net, the encoder block comprises of convolution operations, while the decoder block consists of transposed convolution operations. Conversely, in O-Net, the encoder block is made up of transposed convolution layers, while the decoder block is composed of convolution layers.

For the O-Net models assessed in the current context, we have thus replaced the U-Net architecture in the generator of a Pix2Pix GAN with the proposed O-Net scaffold (as described in Figure 4), while the activation functions used in the encoder & decoder blocks are Swish (Ramachandran *et al.*, 2017) & GELU (Hendrycks & Gimpel, 2020) respectively (which are also used in the U-Net model for comparison purposes). Further image analysis of the generated images was then performed in MATLAB R2020a (© 1984-2020, The MathWorks, Inc), using

well-established objective image quality metrics [such as the peak signal-to-noise ratio (PSNR), signal-to-noise ratio (SNR), image mean-squared error (IMSE) & structural similarity index (SSIM)].

The codes and models used for obtaining the images shown in the present study are supplied in the accompanying Supplement.

4. Image Denoising

Separately, we have also sought to assess the capability of our proposed framework (O-Net) against that of U-Net for models which have ***not*** been explicitly trained for image denoising. An artificial representation of salt-and-pepper noise was infused into the validation images using MATLAB R2020a (© 1984-2020, The MathWorks, Inc) and the models trained for image SR evaluated in parallel for their capabilities in denoising the noise-infused images.

Results

We sought to compare the models developed using our proposed O-Net architecture against the popularized U-Net architecture to attain super-resolved optical microscopical images of samples acquired via both PCM and DIC microscopy. Both of these imaging modalities translate phase variations across different regions of a sample into changes in signal amplitude (exemplified by image brightness), resulting in 'halos' and pseudo-relief features observed in PCM and DIC respectively [where significant phase gradient changes are encountered spatially (Murphy *et al.*, n.d.), (Bagnell, Jr., 2012)]. Conversely, the trained models seek to alleviate these artifacts, while uncovering information concealed within the phase variations of the image, the latter representing higher spatial frequencies in the modulation transfer function (MTF) plot (which correspond to sub-resolved structures) (Introduction to Modulation Transfer Function, n.d.). The results obtained from these experiments are detailed as follows:

1. DIC Imaging

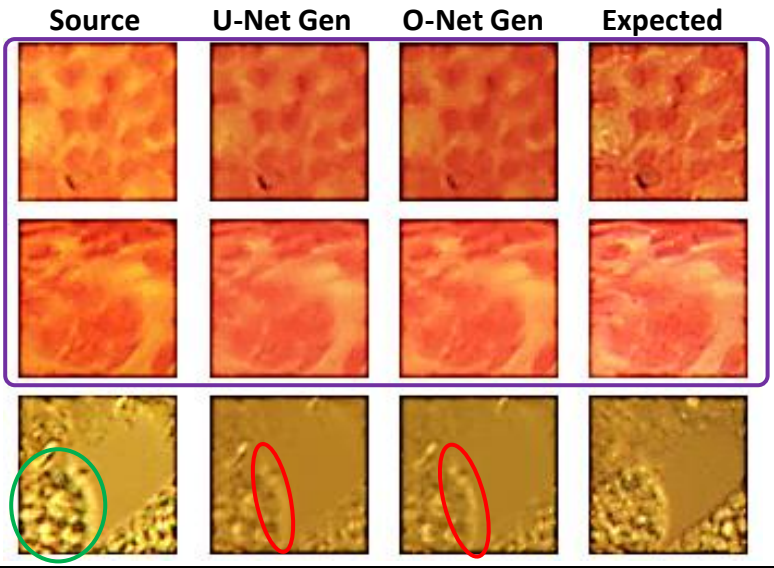


Figure 5: Image super-resolution attained through application of models employing the U-Net and O-Net architectures. In both instances, training of the models was conducted over 101 epochs. The **Source** image (input) was acquired via the 20X/0.40 Ph1 objective, the **U-Net Gen** and **O-Net Gen** images refer to the images generated by the respective DNNs, while the **Expected** image was taken to be the ground truth image for DNN training (in this context, being acquired under the 40X/0.60 Ph2 objective). Interestingly, a comparison between these 3 images reveals a greater similarity between the DNN-generated (i.e. both the **U-Net Gen** and **O-Net Gen**) images and the **Source** images (as indicated within the purple rectangular region), as compared to the **Expected** (ground truth) images. In addition, stark contrasting effects (triggered by steep phase gradients as indicated in the green ellipse) seem to be mitigated by the DNN. On a separate note, details present in the **O-Net Gen** image seemingly surpass that of the **U-Net Gen** image (as shown within the red ellipses), although this is not very apparent from the present Figure. [**N.B.*: All images have been further enhanced (equally) in Microsoft Word (for comparison purposes)].

		PSNR	SNR	IMSE	SSIM
U-Net	**Row 1**	29.6617	24.4121	70.2925	0.9899
	Row 2	25.0388	20.9465	203.7974	0.9832
	Row 3	25.4305	19.5920	186.2237	0.9654
	Average	26.7103	21.6502	153.4379	0.9795
O-Net	**Row 1**	26.8589	21.6093	134.0255	0.9861
	Row 2	26.3047	22.2124	152.2681	0.9848
	Row 3	25.5297	19.6912	182.0172	0.9652
	Average	26.2311	21.1710	156.1036	0.9787

Table 1: Common image quality metrics for each of the DNN-generated DIC images as shown in Figure 5 previously. Here, the metrics displayed are the PSNR, SNR, IMSE & SSIM. In the computation of these values, the **Expected** image is taken as the baseline for comparison against the respective DNN-generated images.

From Table 1, it may be observed that the U-Net generated images generally depict slightly higher PSNR, SNR & SSIM values than the O-Net generated images, while having a lower average IMSE score of 153.4379 (as compared to O-Net's IMSE of 156.1036). All of these may suggest that U-Net generally produces images of a *higher* quality than O-Net (at least as indicated by these statistical quantifiers of image quality). However, a closer inspection of these metrics (and the methodology adopted to quantify these values) will reveal that the U-Net images may have higher scores in these domains as they *exclude* higher spatial frequencies in the MTF plot – a common cause of reduced image quality perceived by these metrics as these are often regarded as '*noise*'. Excluding such detail (which is fundamental to

super-resolution imaging) in the hope of making the generated images less 'noisy' and thus capable of garnering a higher score according to these widely used metrics (to assess image quality) would inadvertently be unviable in the present context. It would be noteworthy to highlight at this juncture the difference between *pseudo* noise and *true* noise – the former constituting higher spatial frequencies encoding useful information which is often misperceived by the user as *true* noise (Kaderuppan *et al.*, 2020). This is not intended to nullify the importance of these metrics in assessing image quality, but merely as a note of caution on deductions drawn through the use of such statistical quantifiers in computational super-resolution imaging (which should be treated with careful consideration).

We consequently sought to deploy models trained under each of these DNN architectures on photomicrographs of *P. dactylus* var *dariana* acquired using DIC with the 100X/1.4 Oil Ph3 objective, to obtain the super-resolved images of the diatom striae. The generated images were compared against the SEM micrograph of the poroids [accessible online at Barone (n.d.)][1], with the results being depicted in Figure 6 below:

[1] Due to image copyright issues, the SEM micrograph cannot be reproduced in the current manuscript, although the interested reader is encouraged to refer to the afore-mentioned source [i.e. Barone (n.d.)], where the SEM micrograph is depicted.

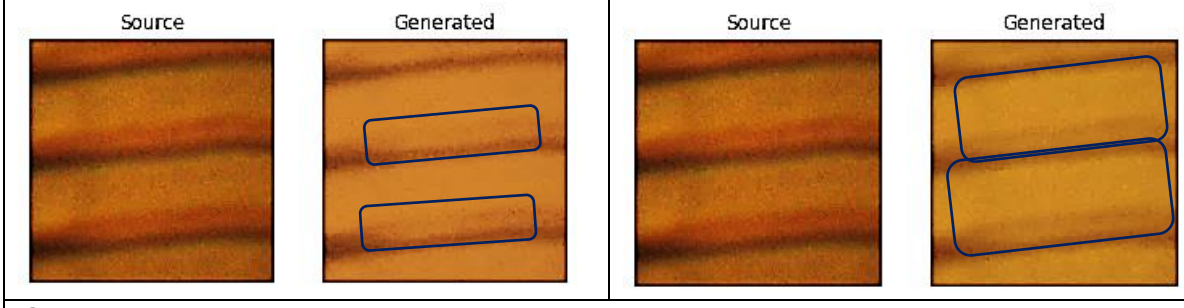


Figure 6: DIC micrographs of *P. dactylus* var *dariana* super-resolved through implementation of **A** the U-Net architecture, and **B** the O-Net architecture in the Pix2Pix generator framework. As in Figure 5 previously, the models employed here were trained over 101 epochs. Notice the clear delineation in the inner periphery of the striae in the O-Net generated micrograph (Panel **B**), which is less evident in the U-Net generated micrograph (Panel **A**). Nonetheless, individual poroids are *not* easily resolved in this context (for both the U-Net & O-Net generated images), as the optical path difference (OPD) between the poroids and their surrounding meshwork is relatively small, resulting in indistinct contrast variations experienced during DIC microscopy. [****N.B.***: All images have been further enhanced (equally) in Microsoft Word (for comparison purposes)].

Further evaluation on the relative accuracies of both the U-Net & O-Net generated images was then conducted, and the results shown in Figure 7 as follows:

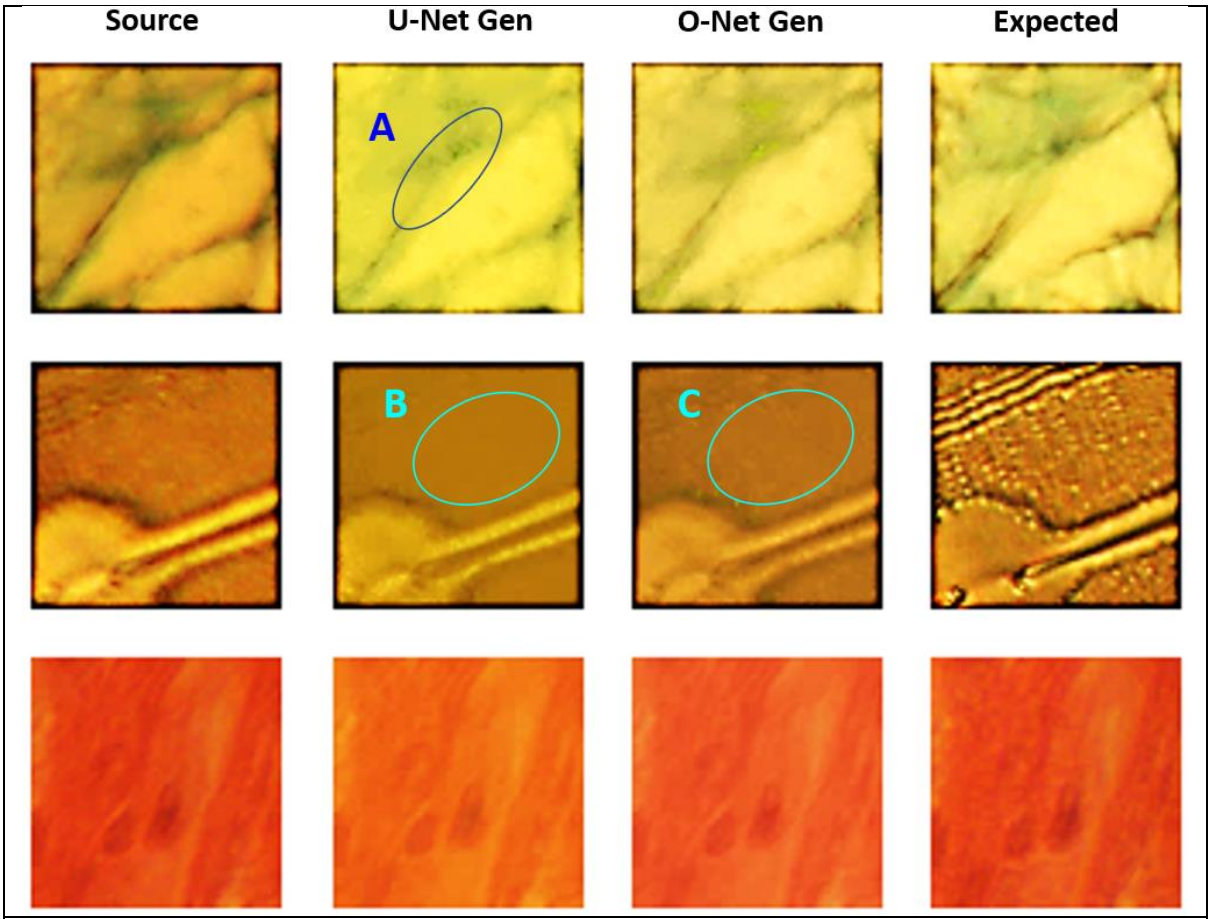
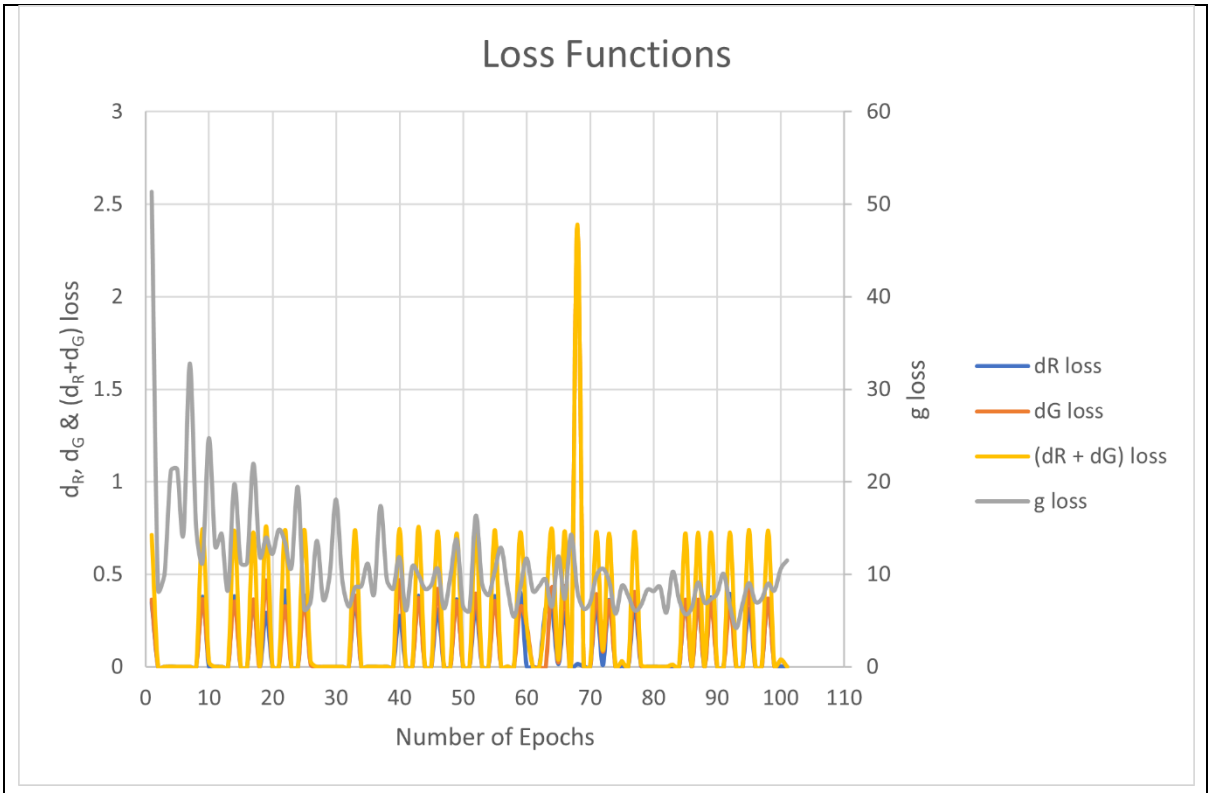


Figure 7: Problems faced with super-resolved images obtained through utilization of the U-Net architecture as compared to O-Net. Here, both the U-Net & O-Net models used were trained over 97 epochs. 2 major issues faced by employing the U-Net models in this context are circled as shown – **A** the presence of extraneous structures in the super-resolved **U-Net Gen** image [a potential consequence of *network hallucination*, as highlighted in (Belthangady & Royer, 2019)], and **B** the considerable loss in detail of some features in the **U-Net Gen** image, when compared to the **Source** (input) or the **Expected** (ground truth) images. The **O-Net Gen** images, however, do not manifest such anomalies, although mitigation of the pseudo-relief effects (present in DIC) may cause some features to be less obvious (as circled in **C**).

As observed from Figure 7, the images generated by U-Net are prone to potential network hallucination and loss of features during *in silico* super-resolution. U-Net de-emphasizes

certain features of interest as a consequence of its initial convolution (*downsampling*) operations (putatively misinterpreting these as noise), while overemphasizing on features which are learnt, mapping these learnt features to the input image and thus generating potential artifacts in the process – a flaw described in Belthangady & Royer (2019) as *network hallucination*. As both issues pose relatively significant impacts on the model's intended deployment (especially with regards to computational nanoscopy applications in the healthcare/research sectors, where the emergence of artifacts may lead to incorrect deductions being drawn), we surmise that U-Net may not be well-suited for computational nanoscopy. O-Net (on the other hand) has not been characterized thus far to exhibit such issues (as indicated by the images produced from a similar O-Net model trained over 97 epochs in Figure 7), since the incorporation of transposed convolution layers prior to convolution in the O-Net architecture allows for redundant features learnt during the DNN training to be dropped from the model during network refinement. From a computational perspective, the plots of the respective loss functions [namely the discriminator losses on real samples (d_R) & generated samples (d_G), as well as the generator loss g] coupled with the compounded discriminator loss ($d_R + d_G$) for each of the models trained under these DNN frameworks are shown in Figure 8 as follows:

A



B

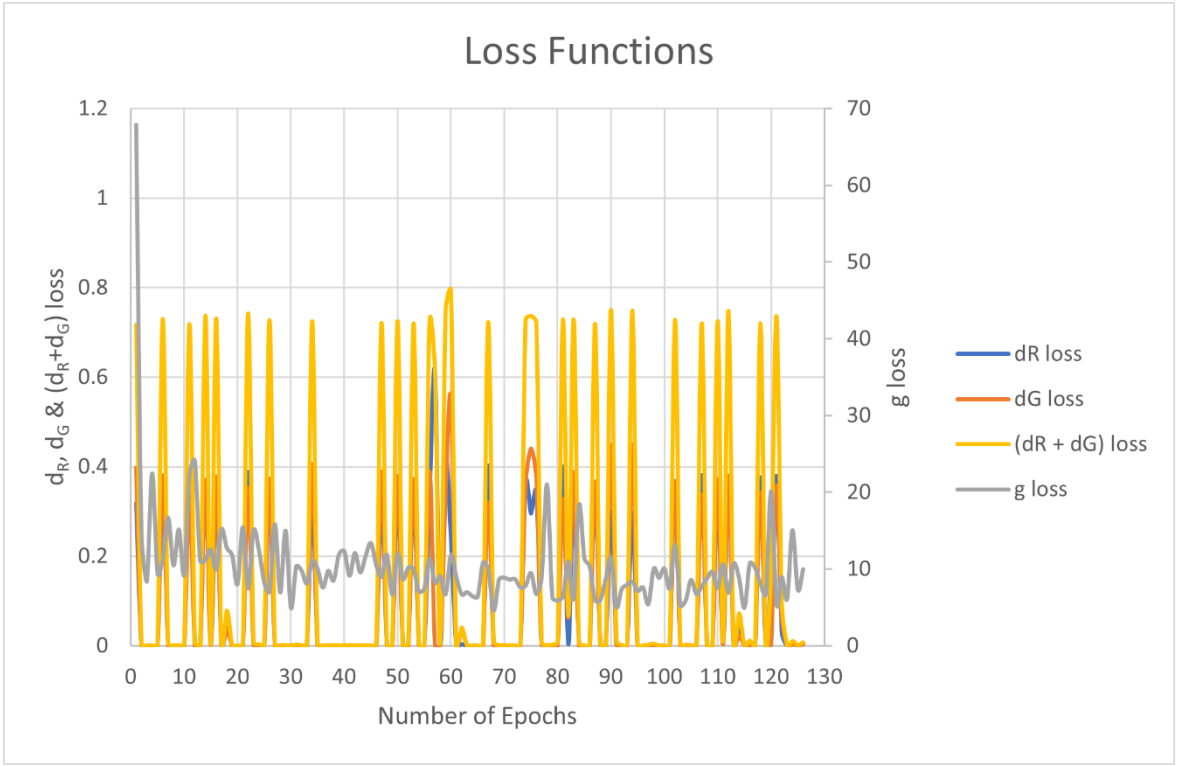


Figure 8: Loss functions for DIC micrographs imaged under **A** the U-Net architecture, and **B** the O-Net architecture in the Pix2Pix generator framework. Notice the spikes in the plot

of the loss functions, which might be attributed to the numerous intermittent training cycles which the models for each framework were subjected to. Here, we take the validation loss to be represented as (d_R + d_G) (with the coupled trailing loss observed as a steep decline in the spike), although it would be useful to mention that when training and evaluating GANs, there is ***no*** objective loss function which can be used to do this (Brownlee, 2019). Specific details of the equations underpinning the individual loss functions are described in the accompanying **Supplementary Information** (for the interested reader).

A similar set of experiments was then conducted for images obtained through PCM, with the findings for these experiments described in the following section:

2. PCM Imaging

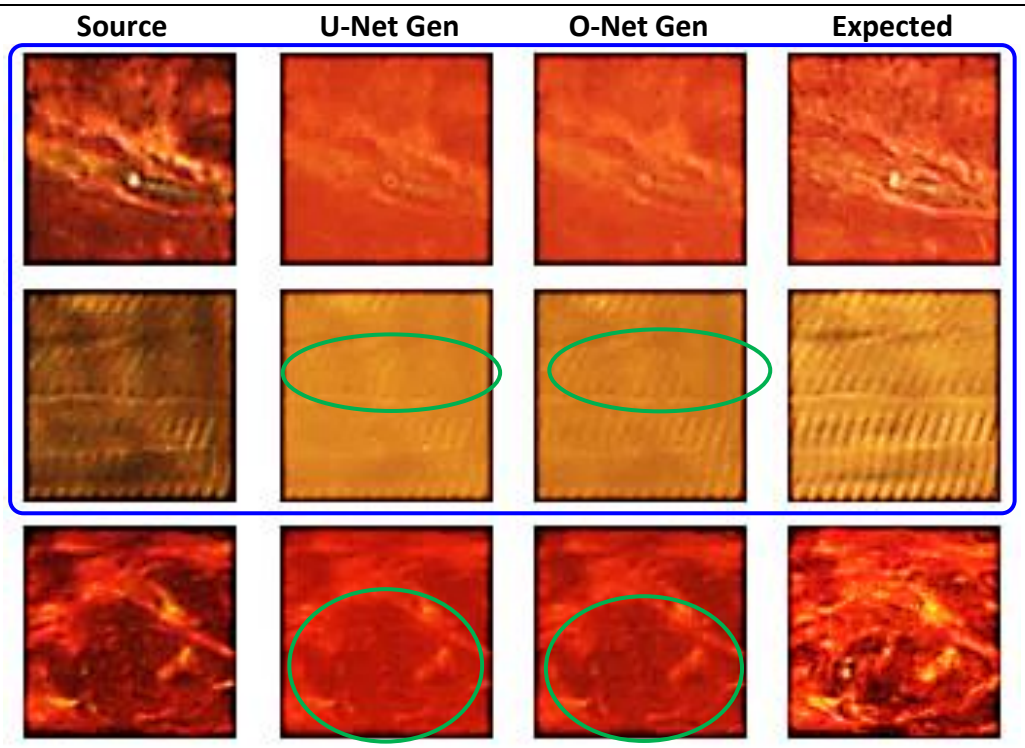


Figure 9: Assimilation of the U-Net and O-Net architectures for PCM computational nanoscopy. In both instances, the models employed were trained over 101 epochs. The **Source** image (input) was acquired via the 20X/0.40 Ph1 objective, the **U-Net Gen** and **O-Net Gen** images refer to the images generated by the respective DNNs, while the **Expected** image was taken to be the ground truth image for DNN training (in this context, being acquired using the 40X/0.60 Ph2 objective). However, unlike Figure 5 previously, here, both the **U-Net** (and **O-Net**) **Gen** images differ substantially from the **Source** image, approaching the **Expected** image more closely (as exhibited within the blue rectangular region). This is partly attributed to the alleviation of the 'halo' and 'shading-off' effects present in the **Source** images within *both* the DNN-generated images, although it is also evident that the **O-Net Gen** images depict more pronounced details when compared to their U-Net counterparts (as described within the green ellipses). [****N.B.***: All images have been further enhanced (equally) in Microsoft Word (for comparison purposes)].

		PSNR	SNR	IMSE	SSIM
U-Net	**Row 1**	26.5040	20.7330	145.4375	0.9837
	Row 2	26.3334	20.9422	151.2641	0.9793
	Row 3	20.7062	14.7384	552.6645	0.9410
	Average	23.5198	18.8045	283.1220	0.9680
O-Net	**Row 1**	27.9498	22.1787	104.2566	0.9869
	Row 2	24.1370	18.7457	250.8319	0.9732
	Row 3	20.0521	14.0842	642.5029	0.9299
	Average	24.0463	18.3362	332.5305	0.9633

Table 2: PSNR, SNR, IMSE & SSIM values for the PCM images shown in Figure 9 previously. As with Table 1 previously, the **Expected Image** is taken as the baseline for comparison against the **Target Image** (to obtain these values).

From the values shown in Table 2, we can say that although O-Net seems to have a higher average PSNR of 24.0463 (as compared to U-Net's 23.5198), U-Net surpasses O-Net in terms of the SNR, IMSE and SSIM values (a lower IMSE score is indicative of a better correlation with the reference image, in this case, the **Expected** images as shown in Figure 9). Nonetheless (as discussed previously), in the context of super-resolution, *none* of these statistical metrics can be taken as definitive quantifiers of image quality, since a lack of very fine details present in the image might be interpreted as an absence of 'noise' [such as Poisson-Gaussian noise (Yang & Lee, 2015) or Speckle noise (Ren *et al.*, 2019)], resulting in elevated scores. In this respect, there is thus a need to distinguish between *pseudo* noise (caused by fine details in the image) and *true* noise (undesired background 'signals' present in the image), as expounded in Kaderuppan *et al.* (2020).

As with Figure 6 previously, employment of both U-Net & O-Net trained models on PCM photomicrographs of *P. dactylus* var *dariana* (captured via the 100X/1.4 Oil Ph3 objective) was performed, with the findings presented in the following Figure 10:

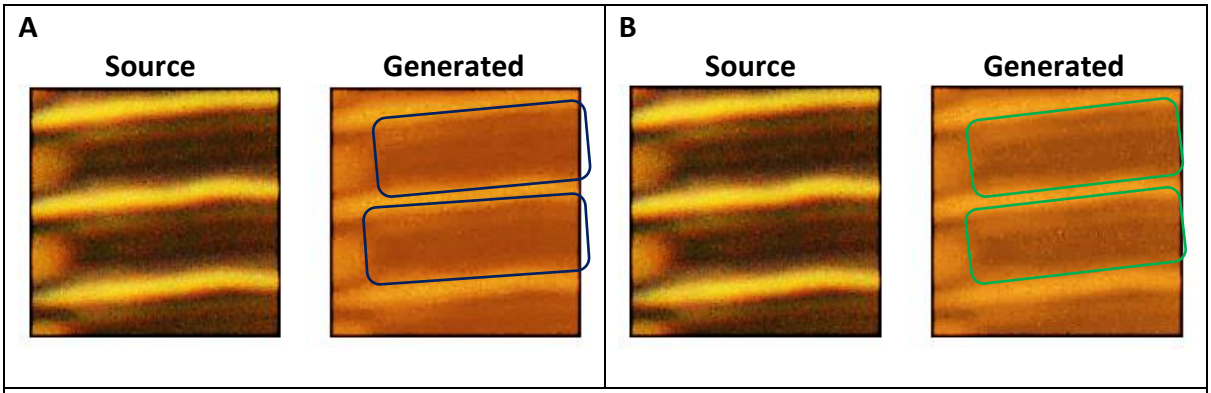


Figure 10: PCM micrographs of *P. dactylus* var *dariana* super-resolved through application of **A** the U-Net architecture, and **B** the O-Net architecture in the Pix2Pix generator framework. As in the previous figures, the models employed here were trained over 101 epochs. From these images, the O-Net generated micrograph depicts the poroid *distributions* more distinctly than the corresponding U-Net generated micrograph (visualized macroscopically as dark double bands in the diatom striae). In addition, upon closer inspection, the individual poroids in the O-Net micrograph (Panel **B**) may be identified as bright specks observed in the green rectangular regions (although these are not as obvious in the present Figure). [****N.B.***: All images have been further enhanced (equally) in Microsoft Word (for comparison purposes)].

In order to visualize the individual poroids more distinctly, we attempted to further train our O-Net model over an additional 20 epochs, with the results being demonstrated in Figure 11 as follows:

Source	Generated

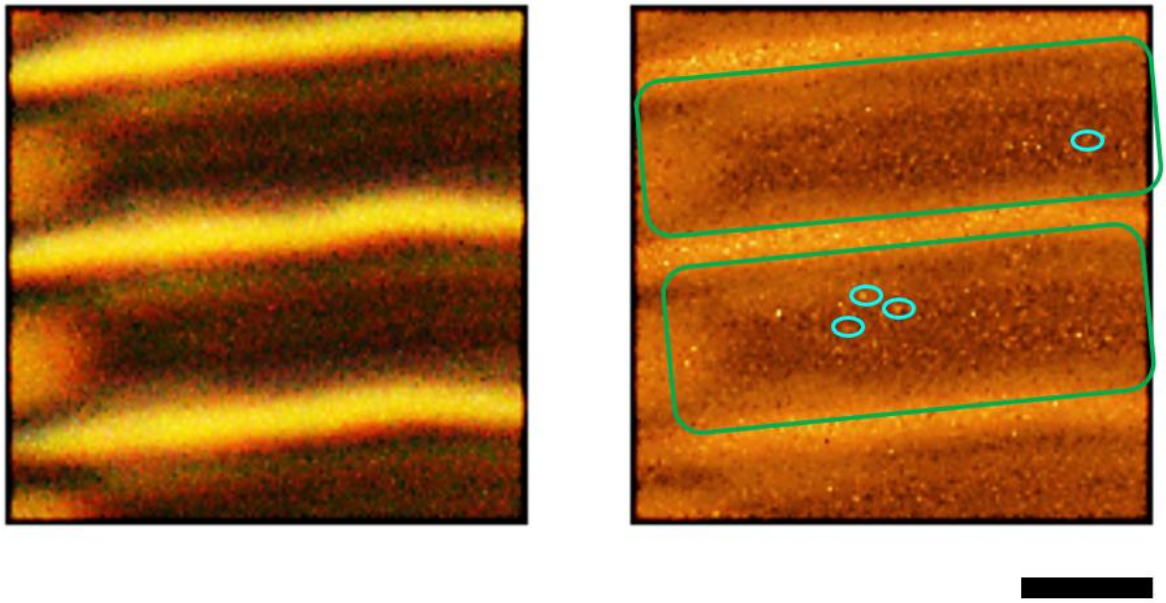


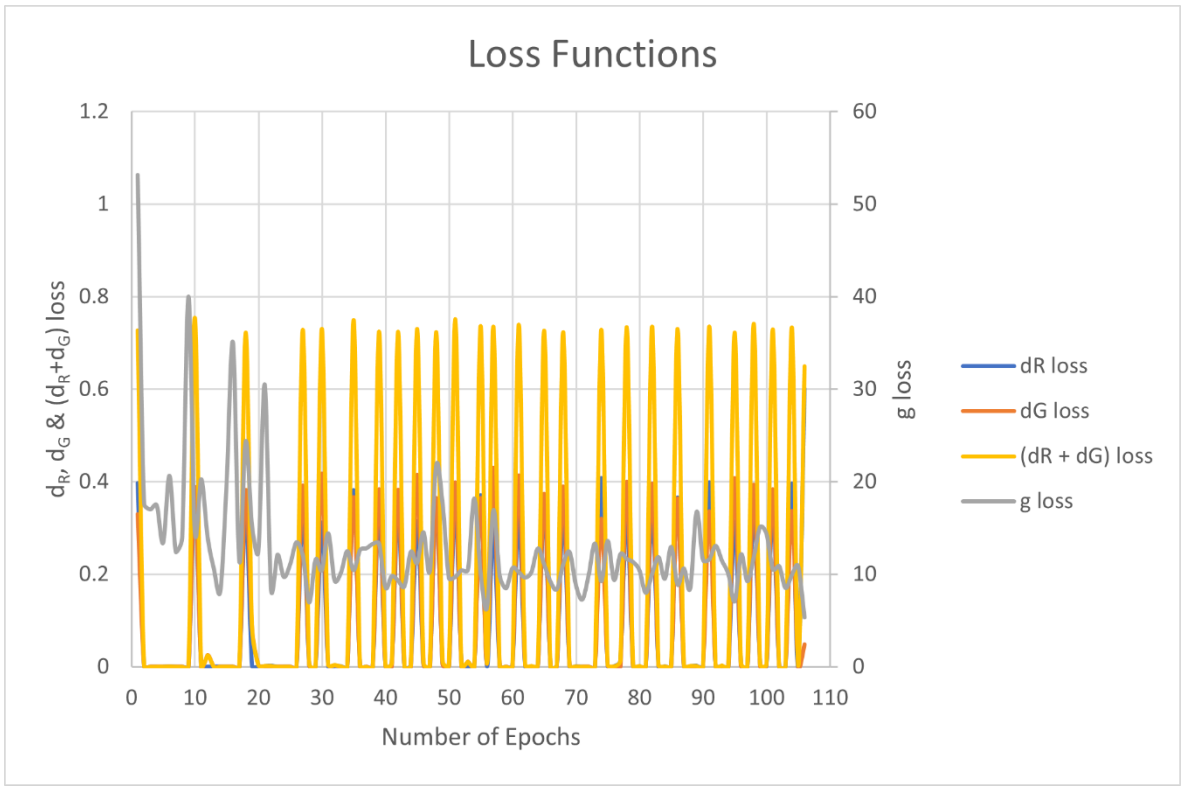
Figure 11: A generated PCM micrograph of *P. dactylus* var *dariana* super-resolved through application of the O-Net architecture in the Pix2Pix generator framework. Here, the O-Net model used was trained over 120 epochs. In this image tile (256x256), the individual poroids become clearly evident as bright granules (4 of which are circled in cyan here) randomly interspersed within a darker meshwork in the green rectangular regions. For comparison purposes, a reference SEM micrograph of the poroids depicting their arrangement in the striae is available online at Barone (n.d.). Scale Bar: 497.96nm. [****N.B.***: All images have been further enhanced (equally) in Microsoft Word (for comparison purposes)].

From the above Figure 11, one may note that the poroids of *P. dactylus* var *dariana* (separated by a mean distance of 80.5nm ± 8.5nm) become easily resolved through employment of our proposed O-Net framework on the source image acquired with the 100X/1.4 Oil Ph3 objective (Leica P/N: 506211). The poroids here appear as bright specks against a darker meshwork, as image contrast in PCM is determined by the phase shift experienced by the light rays as these rays traverse different points in the sample – the rays transmitted through the meshwork

separating the poroids undergo destructive interference at the image plane (unlike that passing through the poroids), making the poroids appear brighter than the meshwork of the striae (Abramowitz & Davidson, n.d.-b). This is in stark contrast to the SEM micrograph [accessible at Barone (n.d.)], where the poroids show up as darker regions within a slightly brighter meshwork of the striae.

Similarly, the plots of the d_R, d_G, ($d_R + d_G$) & g loss functions for each of the models developed under U-Net & O-Net are shown in Figure 12 below:

A



B

Figure 12: Loss functions for PCM micrographs imaged under **A** the U-Net architecture, and **B** the O-Net architecture in the Pix2Pix generator framework. As illustrated in Figure 8 previously, several spikes are evident in the plots of all the loss functions (d_R, d_G, $d_R + d_G$ & g loss) in both the U-Net & O-Net architectures. These spikes are likely triggered by the approach used to train the models. Further details underlying this anomaly are surfaced in the Limitations section of this paper. In addition (and as also described previously in Figure 8), we are using the compounded loss ($d_R + d_G$) as the validation loss for the model in this context (the trailing loss being represented as the steep decline in the validation loss spike), although there is (in fact) ***no*** objective measure for assessing the performance of a GAN, as discussed in Brownlee (2019).

3. Image Denoising

A simple salt-and-pepper noise algorithm was used to inject noise into the source image in MATLAB, and image denoising was separately performed in Python. The results for each of

the network-generated images (as compared to the source & ground truth images) are depicted in Figure 13 below[SSK(1)]:

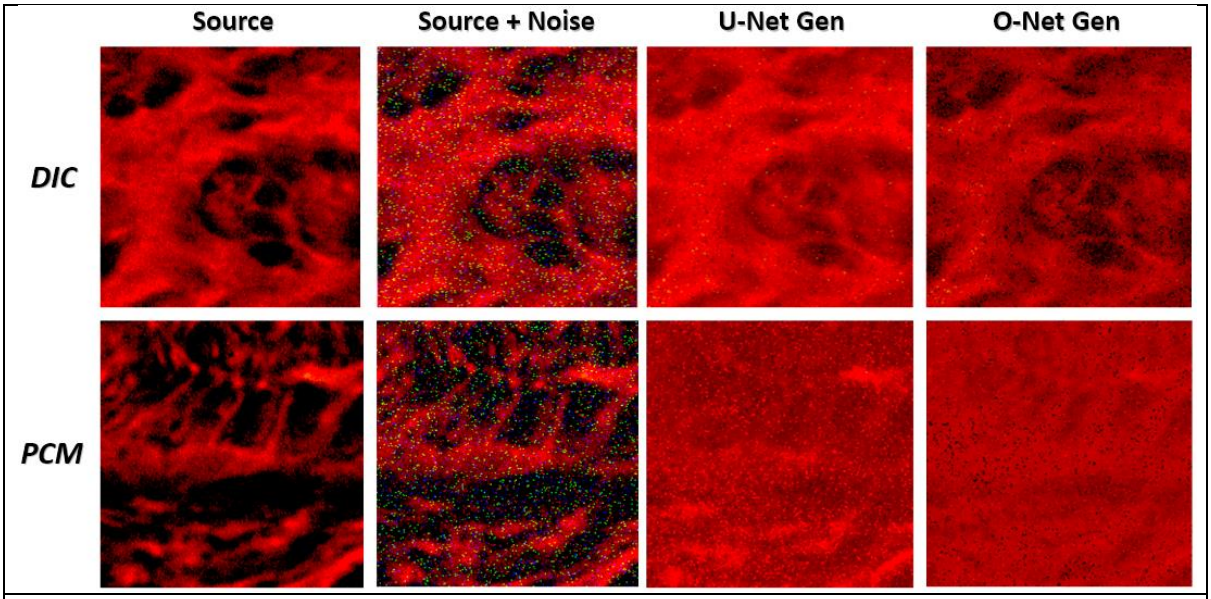


Figure 13: A couple of salt-and-pepper noise-injected micrographs (labelled **Source + Noise**) which have been imaged under both DIC and PCM. These noisy images have been subsequently processed using the U-Net & O-Net models, each of which was trained over 101 epochs. The original **Source** and **Source + Noise** images are included for comparison with the U-Net (and O-Net) generated images. From this Figure, we may observe that U-Net images generally include the noisy green pixels (especially for the PCM image) as bright 'hot' pixels, while the O-Net generated images tend to 'burn in' the noisy pixels to blend with the background. It is essential at this juncture to inform that the models were ***not*** trained to denoise the input images, hence any implication intended to do this (i.e. image denoising) in the current context is not anticipated. Nonetheless, it would also be useful to highlight that although both models do not eliminate noise from the input image (since they were not specifically trained for this context), the U-Net generated images tend to over-emphasize the noise as prominent 'hot' pixels in the image, unlike O-Net-generated

image which provides a more realistic view of the specimen by blending the noise pixels with the background, although image contrast in the U-Net generated PCM image generally surpasses that of its O-Net counterpart. [***N.B.***: All images have been further enhanced (equally) in Microsoft Word (for comparison purposes)].

4. Computational Complexity & Load

Other than the evaluation of images generated from models employing either network architectures, we have also attempted to evaluate the computational complexity and load imposed by training models under both the U-Net & O-Net frameworks. Here, the average time required for the execution of a single training iteration was utilized as a metric for model training efficacy, the results of which are tabulated in Table 3 below:

Framework	GPU	CPU	RAM (GB)	Initial Iteration Time / s	*(Average Time / Iteration) / s
5-layer U-Net	NVIDIA GeForce 920MX	Intel® Core™ i5-7200U CPU @ 2.50GHz	20	46.647	42.429
	NVIDIA Tesla K80	2 * Intel® Xeon® Platinum 8170 CPU @ 2.10GHz	95.7	16.845	2.185
	NVIDIA GeForce RTX 3090	Intel® Core™ i9-10920X CPU @ 3.50GHz	128	2.334	0.299
5-layer O-Net	NVIDIA GeForce 920MX	Intel® Core™ i5-7200U CPU @ 2.50GHz	20	45.310	42.369
	NVIDIA Tesla K80	2 * Intel® Xeon® Platinum 8170 CPU @ 2.10GHz	95.7	7.985	2.386
	NVIDIA GeForce RTX 3090	Intel® Core™ i9-10920X CPU @ 3.50GHz	128	2.391	0.296

Table 3: Training iteration times (in seconds) indicated for models trained under each of the assayed frameworks (i.e. U-Net & O-Net) ***within the same session*** employing a 5-layer network architecture in each of the descending & ascending

limbs. The **Initial Iteration Time** refers to the time taken for the first iteration, while the **(Average Time/Iteration)** refers to the average time *per* iteration for iterations after the first iteration.

From the values shown in Table 3, we may deduce that the average computational loads & complexity (as measured by the iteration times) for each of the U-Net and O-Net architectures appear to be relatively similar in systems employing either the NVIDIA GeForce 920MX or NVIDIA RTX 3090 GPUs. However, in the NVIDIA Tesla K80 GPU system, O-Net actually demonstrates a slightly improved training performance with an almost 2-fold lower initial iteration time. ***N.B.: These results demonstrate the training time per iteration for models trained using either framework within the same session (workspace variables being cleared before commencing the training of the subsequent model).*** Schematically, the O-Net architecture which introduces sparsity in the model training (thereby resulting in an initial increase in the parameter size as depicted in **Supplementary Table ST6**) would be expected to exhibit a greater computational load than U-Net (which seeks to reduce the initial input parameter size as shown in **Supplementary Table ST3**) albeit the number of parameters being identical for both models, but this is clearly not the case from the results in Table 3. Instead, the training of O-Net based models seem comparatively similar to that of U-Net (with regards to training efficacy). An important caveat however exists here – that of training the models within the same session (although the workspace variables may be cleared prior to the training of the next model).

Discussion

From the results gleaned thus far, one may realize the capabilities of O-Net in computational nanoscopy for phase-modulated microscopical imaging techniques (such as DIC and PCM).

Here, we have shown that O-Net is capable of super-resolving both PCM and DIC images (as depicted in Figures 6, 10 and 11) while also alleviating several common artifacts faced when acquiring images using these techniques [such as the 'halo' and 'shading-off' effects in PCM (Murphy *et al.*, n.d.), or the pseudo-3D effect in DIC (Bagnell, Jr., 2012)]. This has also led us to surmise that the proposed O-Net models seek to *narrow* the optical *point spread function* (PSF), which may be expressed mathematically as follows:

Assuming ideal image formation (i.e., where noise & optical aberration effects are absent), then we have the following:

$$(f \circledast g) \xrightarrow{j} (h \circledast g) \qquad (4)$$

where *f* refers to the PSF of the optical system (when using the 20X/0.4 Ph1 objective), *g* refers to the actual, unconvolved (ground truth) image of the specimen and *h* refers to the PSF of the optical system (when using the 40X/0.6 Ph2 objective). Here, the O-Net model endeavours to learn a suitable function *j* to reduce the impact of the PSF *f* on the ground truth image *g* (since *h* has a smaller blurring kernel size as compared to *f*). When the O-Net model has learnt *j*, this value of *j* is then applied as follows:

$$(k \circledast g) \xrightarrow{j} \underbrace{(m \circledast g)}_{\text{super-resolved image}} \qquad (5)$$

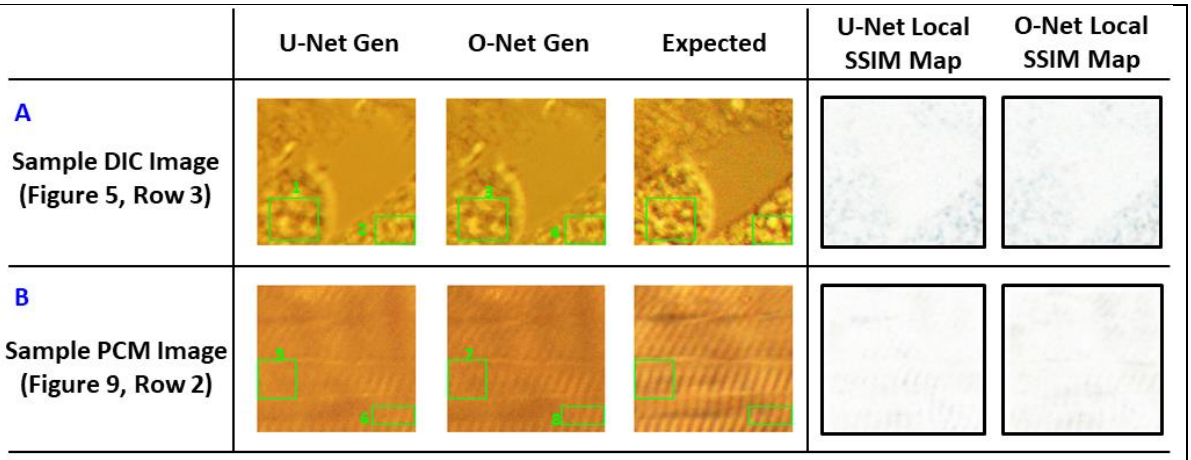
where *k* refers to the PSF of the optical system (when using the 100X/1.4 Ph3 objective), *g* refers to the actual, unconvolved (ground truth) image of the specimen and *m* refers to the *computed* PSF of the optical system (*if* using a super-high NA objective). The PSF *m* is thus a *computed* (i.e. hypothetical) PSF, deduced from reducing *k* by the same mapping function *j* used to reduce *f* to *h*. Through this, we can thus obtain the super-resolved image $(m \circledast g)$.

Moreover, unlike U-Net which constricts the input dataset to a condensed feature space with each successive convolution, thereby reaching a bottleneck before expanding on the learnt

features (Kızrak, 2019), O-Net introduces *sparsity* into the input dataset initially (through transposed convolution operations) before removing redundant features which may be introduced during the learning process (via the successive convolution layers). This explains why O-Net is less prone to artifact generation such as network hallucination, as discussed in Belthangady & Royer (2019) and Ouyang *et al.* (2018). In addition (& unlike previously-employed approaches such as that in Ouyang *et al.*, 2018), our models have been trained with a diverse range of images obtained from plant, animal and protist samples, thereby being unrestricted to the specimen type and less prone to network-overfitting.

On a separate note, an analysis of the loss function plots described in Figure 12 previously may lead one to concur that the O-Net PCM models have ***not*** been optimally trained as compared to their U-Net counterparts, since the *g* loss for U-Net is considerably smaller than those of O-Net. This observation may also be partially corroborated through the metrics in Table 2, which also seemingly indicate a better overall performance (higher SSIM score) of U-Net as compared to O-Net. Nonetheless, it would be essential to highlight here that in the domain of super-resolution microscopy, one seeks to retrieve very high spatial frequencies (corresponding to very fine details) within the image, which may often be obscured in the background and misconstrued as noise [a phenomenon termed *pseudo* noise, as discussed in Kaderuppan *et al.* (2020)]. A generated image which thus scores very high in these commonly utilized image metrics (PSNR, SNR & SSIM) may be regarded to have very little noise coupled with high overall structural similarity to the target image, although this lack of noise may also include the absence of very fine structures in the said image (the said structures being perceived by these algorithms as 'noise'), as highlighted previously. Similarly, a model exhibiting a lower *g* loss should not be conceived as being better-trained in this context, since the metrics used to compute these losses [namely the mean absolute error (MAE) and binary

cross-entropy] are also founded on the principle of *universal homology* between the generated and ground truth images, although these differ from the mean squared error (MSE), which constitutes the basis of IMSE & PSNR in Tables 1 & 2 (as described in the accompanying Supplementary Information). For this reason, despite the *g* losses for the O-Net PCM models being considerably higher than that of U-Net (as shown in Figure 12), one cannot assume that the O-Net models have not been sufficiently trained at this juncture (unless this assumption may be verified through other means), since the *g* losses for O-Net might have potentially equilibrated at a higher value than that of U-Net (for the afore-stated reasons). In a similar context, the difference between the values of these metrics (i.e., the PSNR, SNR & SSIM) for both the O-Net & U-Net models (considering the DIC & PCM images) is relatively small (< 1), implying a presumably insignificant variation, which may be attributed to the reduction of 'noise' in the U-Net generated images. For this reason, we have sought to evaluate the *local* SSIM values for *both* the U-Net (& O-Net) generated images (when matched against their respective ground truth reference standards) using MATLAB R2020a (© 1984-2020, The MathWorks, Inc). A representation of the results obtained is depicted in Figure 14 below:



Panel	Image Type	ROI	Global Values	SSIM	PSNR	SNR	IMSE
A (Fig 5, R3)	U-Net Generated DIC	1	PSNR: 25.4305 SNR: 19.592 IMSE: 186.2237 SSIM: 0.9654	0.94997	23.4433	16.2587	294.271
		2		0.92647	20.8153	13.9698	538.9536
	O-Net Generated DIC	3	PSNR: 25.5297 SNR: 19.6912 IMSE: 182.0172 SSIM: 0.9652	0.94918	23.4234	16.2388	295.6224
		4		0.92712	21.0013	14.1558	516.3626
B (Fig 9, R2)	U-Net Generated PCM	5	PSNR: 26.3334 SNR: 20.9422 IMSE: 151.2641 SSIM: 0.9793	0.97738	26.3308	19.4644	151.3566
		6		0.97649	27.5279	20.1521	114.8919
	O-Net Generated PCM	7	PSNR: 24.137 SNR: 18.7457 IMSE: 250.8319 SSIM: 0.9732	0.97288	24.2071	17.3407	246.8115
		8		0.97781	26.9132	19.5374	132.3612

Figure 14: Local SSIM maps for individual pixels (& SSIM values for indicated ROIs) in sample **A** DIC and **B** PCM images, obtained *via* both U-Net (& O-Net) based models (the DIC image being the same as **Figure 5 Row 3** & the PCM image being **Figure 9 Row 2**). The darkened regions represent areas which have low local SSIM values [i.e. implicating differences between the Generated and Expected (target) images], while the lighter regions are indicative of areas having higher SSIM scores. The increased density & extent of lighter regions thus contributes to generally elevated SSIM scores for U-Net models (as described in Tables 1 & 2 previously). Further details on the other parameters (including the global & local SSIM, IMSE, PSNR & SNR) are shown in the Table (values shown in green indicate the better-scoring image amongst a pair of assessed ROIs).

From Figure 14, we may deduce that U-Net models generally depict slightly greater overall similarity to the target (Expected) image, based on their global SSIM scores. However, on closer inspection, we notice that the local SSIM (for the individual ROIs) *vary* across the image, with both the U- (& O-) Net models depicting somewhat equivalent efficacy in super-resolving the images, at least where the local SSIM is considered (the U-Net model surpasses that of O-Net in 4 out of 8 ROIs, while a similar trend is observed for O-Net surpassing U-Net as well). These findings suggest that the global SSIM score (though often exploited for assessing image similarity to the ground truth) may ***not*** be a suitable metric for evaluating the image SR quality, a deduction corroborated by manual visual inspection of the generated images in Figure 14, as well as in **Supplementary Figures SF7** & **SF8** (where the features which are more clearly discernible in the O-Net output than U-Net are circled in the coloured ellipses). In addition, the differences between the global SSIM scores for the models are largely insignificant, being (on average) <0.015. Nonetheless, an interesting anomaly is noted when comparing ROIs #5 & #7 (ROI #5 has a higher local SSIM score of **0.97738**, as opposed to ROI #7 with a local SSIM score of **0.97288**), although the striations in ROI #7 appear to be more distinct than that in ROI #5. Through closer inspection of the said ROIs, we may surmise that the greater density of matched *background* pixels in ROI #5 might have contributed to the elevated local SSIM scores, implying that SSIM (even when applied locally as in the current context) might not be a suitable predictor of image SR quality.

In light of the measured SSIM anomaly depicted previously, we have also sought to compute other parameters (such as the PSNR, SNR & IMSE) for comparison amongst the different ROIs, the details of which are exemplified in **Supplementary Tables ST7** & **ST8**. Interestingly, these other assayed metrics support the findings gleaned from SSIM comparisons (i.e. that of ROIs #5, #1 & #4 being predominantly 'better' than ROIs #7, #3, #2 respectively), although they

also purport that ROI #6 is 'superior' to ROI #8 in terms of image SR quality, a finding which contradicts that of the local SSIM metric for these ROIs.

Expectedly, a putative avenue for future work in the present context might explore the incorporation of some regularization approaches into the loss functions to allow a greater alignment of the loss value with the model performance *locally* (rather than globally). In addition [and as highlighted by the anomalies surfaced in Figure 14 and **Supplementary Figures SF7** & **SF8** (as well as **Supplementary Tables ST7** & **ST8**) previously], a suitable metric which more closely aligns with the visually-discerned output of the generated images might need to be developed in this respect, facilitating both model validation & evaluation.

It would be noteworthy to mention that previous work done by other researchers in this domain of computational nanoscopy [such as Ouyang *et al.* (2018), Wang *et al.* (2019), Nehme *et al.* (2018) & Rivenson *et al.* (2017)] has also convincingly demonstrated how one may be able to obtain computational super-resolved *fluorescent* microscopical images through application of a DNN. One of these studies (Ouyang *et al.*, 2018) describes a self-developed DNN (termed A-Net), which proved successful in generating images comparable to that of photoactivated localization microscopy (PALM) (Betzig *et al.*, 2006), an approach named ANNA-PALM by Ouyang *et al.* (2018). Here, the developers of ANNA-PALM cited key advantages of their proposed network, including the ability of performing high-throughput super-resolution fluorescent microscopy, being indifferent to experimental fluctuations & yet adaptable to biological structures which may be very different from the training dataset (Ouyang *et al.*, 2018). Nonetheless, Ouyang *et al.* (2018) were also aware of the limitations of ANNA-PALM, such as the need to be trained on structures similar to those which were being elucidated and the imminent failure of A-Net, when required to generate images based on a random molecular distribution. Separately, Wang *et al.* (2019) have also developed a U-Net-

based GAN architecture for super-resolving confocal fluorescence microscopical images, which demonstrated significant similarity to the STED images. Similarly (and as expounded in the current study), our proposed O-Net model has demonstrated significant potential in reconstructing a super-resolved image based on a *single* input image, without being provided prior information on the PSF of the system used. As a *blind* approach to obtaining super-resolved optical microscopical images without using nanobeads to decipher the PSF of the optical train [such as is commonly used in non-blind image deconvolution protocols (Scientific Volume Imaging, n.d.)], O-Net thus signifies a novel method of obtaining accurate (and potentially hallucination-free) super-resolved images for ***phase-modulated optical microscopical modalities***, namely PCM and DIC microscopy. In this context, additional key advantages which may be gained through our currently proposed approach include the removal of the halo and shading-off artifact often encountered in native PCM & DIC imaging respectively, as well as the *reduction* (but ***not*** elimination) of inherent noise (such as salt-and-pepper noise) present in the micrographs, through blending the 'noisy' pixels into the background (as illustrated in Figure 13 previously). It is essential to emphasize on the importance of this, since *eliminating* the noise entirely may result in a potential loss of high-frequency information encoded in the image (if the noise was actually *pseudo* noise). This (to us) revolutionizes the applicability for each of these techniques, extending their use-cases to cover more subjects (especially from a biological perspective) and without the need for introducing fluorescent probes (which may inadvertently disrupt the native *in vivo* environment).

Delving deeper into image denoising (although this does not represent the focus of our present study & trained models), it would be noteworthy to highlight that although Tables 1 and 2 depict higher SNR & lower IMSE scores of U-Net over O-Net (implying an improved

performance of U-Net in its ability to denoise images & a potentially unoptimized O-Net architecture being utilized in the present context), the DNN-denoised results demonstrated in Figure 13 seem to suggest otherwise. Here, the O-Net generated images 'burn in' the noisy pixels, while the U-Net model *emphasizes* the noise as 'hot pixels' (where the images are subjected to salt-and-pepper noise). This (to us) surfaces an interesting anomaly, which may be postulated to be attributed to the manner in which metrics such as SNR & IMSE are computed. These metrics likely penalize the non-suppression of *pseudo* noise features more heavily than true noise, implying that as U-Net generally suppresses high frequency *pseudo* noise features more than O-Net, it scores better in these measures, while the under-rated performance of O-Net in these aspects is due to the fact that it does not suppress such high frequency information (unlike U-Net). In any case, the difference in these metrics between the U-Net & O-Net models is relatively small and don't appear to be statistically significant.

From a computational standpoint, the generation of images from O-Net-derived models takes <1 min (on average) on an Intel® Core™ i5-7200U CPU laptop (equipped with 20GB DDR3 RAM and a NVIDIA GeForce 920MX GPU), making it a quick & efficient way of obtaining super-resolved PCM & DIC images. In addition, training of O-Net models demonstrate relatively similar training efficacies as that of U-Net models, depicted by their average training times per iteration in Table 3 previously.

Limitations of the Current Study & its Potential for Future Work

Training of the DNNs in the present study were performed intermittently (due to hardware limitations and the batch size required to train the models). The intermittent training resulted in spikes observed in the loss function plots for the discriminator network (as shown in Figures 8 and 12), which we presume may be due to a random seed being selected for the discriminator each time a training run is executed. In this regard, we have sought to compare

(as part of a separate study) the precision of the results obtained via intermittent *vs* continuous training, for further verification of the model performance & the losses obtained. We demonstrate (as described in Kaderuppan *et al*., 2022) that intermittent training of the DNN generally results in images bearing a closer similarity to the ground truth (Expected) images, as opposed to their continuously-trained counterparts.

A second limitation in the current study was that the models utilized in this context were trained on data acquired under a Leica DM4000M microscope in conjunction with a RisingCam® CMOS camera. This implies that the said models were trained to recognize and compensate for the PSF experienced in the optical train for *this* particular microscope and camera setup, while the use of other microscope and camera models may necessitate a *re-training* of the models (to compensate for the different aberrations present in the optical train for these microscope models). This may be further accentuated through the use of different specimen preparation protocols, varied illumination intensities and objective types, all of which contribute to fundamental variations in the PSF. Circumvention of this limitation may thus be probable with the use of multiple datasets (acquired from different microscope and camera combinations) for training the models. Nonetheless, the functionality of the proposed O-Net architecture (for construction of these models to do super-resolution imaging) is seemingly assured, as depicted through our present study.

Thirdly, the current study seeks to explore the functionality of O-Net in super-resolving images acquired via two popular phase-modulated microscopical imaging modalities (i.e., PCM and DIC microscopy). In these techniques alone, we have demonstrated the capability of O-Net to derive substantially accurate SR images. However, it would be prudent to emphasize that besides these methods, there exist several other phase-modulated microscopical techniques [such as oblique illumination (Chambers *et al.*, n.d.) and Hoffman

modulation contrast (Abramowitz & Davidson, n.d.-a)] as well as other popular optical microscopical methods [including darkfield (Chambers *et al.*, n.d.) and polarization contrast (Robinson & Davidson, n.d.)], which we have not assessed the accuracy of O-Net to super-resolve. In this regard, future investigations on the use of O-Net in super-resolution microscopy may involve its employment in various other optical microscopical imaging approaches such as oblique illumination, polarization contrast or darkfield microscopy.

Another possible limitation which may be addressed through future work include the development of a suitable image metric for assessing the 'super-resolution' quality of an image. Such a metric should consider the key difference between the features and the background of an image (as well as between *pseudo* noise & *true* noise), while also potentially incorporating a penalty for network hallucination effects, which (as severe training artifacts) may lead to ill-informed decisions. This metric may then also be employed for automated optimization of the O-Net architecture for SR imaging (currently, all optimization of the models trained is done *manually*, which is tedious and not easily implemented). In addition, GAN optimization currently poses a challenge on its own, while utilising a widely-employed metric such as the loss function for optimizing the GAN does not guarantee that the GAN has been similarly optimized under other hyperparameters (such as SSIM, IMSE, PSNR or SNR). A fine balance would thus need to be established between optimizing the GAN based on its loss function and other image metrics. Alternatively, the use of persistent homology diagrams as a means for defining pixel connectivity coupled with potential artefacts caused by introducing sparsity (in the developed O-Net architecture) for quantifying some of these structural details may also be regarded as useful avenues for exploration in future work.

Other potential avenues which may be considered for extrapolations of the present study include the use of higher-powered objectives for training the models, deeper O-Net

architectures, identifying the minimum size of features that an O-Net-trained model can effectively resolve, as well as the potential integration of deconvolution and/or denoising algorithms into the proposed O-Net framework. All these thus spell an exciting & promising future for O-Net as the *de facto* framework for computational nanoscopy.

Conclusions

In the presented work, the performance of our proposed DNN architecture (O-Net) was assessed against the widely utilized and state-of-the-art DNN architecture [U-Net (Ronneberger *et al.*, 2015)] with respect to its accuracy in super-resolving PCM & DIC micrographs computationally. Our proposed model depicts a relatively high level of accuracy while being simultaneously potentially immune to issues such as network hallucination (a common dilemma faced by U-Net trained models). Moreover (& as previously highlighted), our models have been trained using a wide variety of images acquired from a multitude of samples (plants, protists, animal tissue, etc). This makes them generally resistant to network over-fitting, unlike past studies (such as Ouyang *et al.*, 2018) which emphasized on specific learnt traits (e.g. microtubule filaments) for super-resolution mapping. Nonetheless, it would be noteworthy to highlight at this juncture that we are unable to use popular statistical measures such as the Receiver Operating Characteristic (ROC) curve or SSIM to quantitatively determine the relative accuracies of the super-resolved images generated by our proposed O-Net architecture, since there are currently *no* optical super-resolved microscope images which may be utilized as a basis for comparing with the super-resolved PCM and DIC images - we manually assessed the super-resolution capabilities of O-Net and U-Net against the SEM image of *P. dactylus* var *dariana* in the current work. One may seek to utilize the statistical metrics described in Tables 1 and 2 (such as the PSNR, SSIM and ISME) as a means of gauging the accuracy of the O-Net generated images against a known standard (in this case, the

images obtained from the 40X/0.6 Ph2 objective as shown in Figures 5 & 9), but we do *not* regard this as a viable comparison in the present context, for the reasons explicated previously. Moreover, some of the higher-resolution images differ significantly from that of the lower-resolution images, while having a high SSIM score in this context (as shown in Tables 1 and 2) might inadvertently be misinterpreted by the reader as a sterling performance exhibited by the DNN (in resembling the ground truth data), although this might (in fact) be indicative of network *over-fitting*, which *may* consequently result in potential network hallucination issues. In this regard, we believe that a new local image metric should be developed (potentially in future studies) to assess the quality of the super-resolved images, as opposed to the highly popular global image metrics (such as SSIM, PSNR and IMSE) employed currently. Despite this, we can confidently postulate on the potentiality of O-Net as a novel DNN architecture for wide-scale deployment in computer vision applications, in particular computational nanoscopy.

Acknowledgments

The authors would like to extend their sincere appreciations towards past studies conducted by other researchers in the field who have laid some of the groundwork for attaining computational nanoscopy. In addition, the authors would also like to thank the Editorial staff & reviewers of *Microscopy and Microanalysis* for reviewing this manuscript and suggesting revisions to improve it.

Competing Interests

The authors declare none.

References

Abramowitz, M., & Davidson, M. W. (n.d.-a). *Hoffman Modulation Contrast Basics*. (EVIDENT|OLYMPUS CORPORATION) Retrieved 4 7, 2021, from https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/hoffman/

Abramowitz, M., & Davidson, M. W. (n.d.-b). *Introduction to Phase Contrast*. (EVIDENT|OLYMPUS CORPORATION) Retrieved 19 4, 2021, from https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/phasecontrast/phase/

AIVIA. (n.d.). (Leica Microsystems) Retrieved 15 7, 2021, from https://www.aivia-software.com/aivia

Akst, J. (2018). *AI Networks Generate Super-Resolution from Basic Microscopy*. Retrieved 14 10, 2021, from https://www.the-scientist.com/news-opinion/ai-networks-generate-super-resolution-from-basic-microscopy-65219

AMD Radeon™ RX Graphics Cards. (n.d.). (Advanced Micro Devices, Inc) Retrieved 4 7, 2021, from https://www.amd.com/en/graphics/radeon-rx-graphics

Artificial Intelligence for 3D Visualization and Analysis Software. (n.d.). (Thermo Fisher Scientific Inc.) Retrieved 15 7, 2021, from https://www.thermofisher.com/us/en/home/electron-microscopy/products/software-em-3d-vis/3d-visualization-analysis-software/artificial-intelligence.html

AWS Deep Learning AMIs. (n.d.). (Amazon Web Services, Inc.) Retrieved 4 7, 2021, from https://aws.amazon.com/machine-learning/amis/

Bagnell, Jr., C. R. (2012). Chapter 11 - *Differential Interference Contrast Microscopy*. Retrieved 2 11, 2019, from https://www.med.unc.edu/microscopy/files/2018/06/lm-ch-11-dic.pdf

Barone, S. (n.d.). *Diatom Shop*. (Diatom Lab) Retrieved 4 7, 2021, from http://www.diatomshop.com/

Bayraci, S., & Susuz, O. (2019). A Deep Neural Network (DNN) based classification model in application to loan default prediction. *Theoretical and Applied Economics, XXVI* **4**(621), 75-84.

Belthangady, C., & Royer, L. A. (2019). Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature Methods* **16**(12), 1215-1225.

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., & Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**(5793), 1642-1645.

Brownlee, J. (26 8, 2019). *How to Evaluate Generative Adversarial Networks*. (Machine Learning Mastery Pty. Ltd.) Retrieved 7 2, 2022, from https://machinelearningmastery.com/how-to-evaluate-generative-adversarial-networks/

Chambers, W., Fellers, T. J., & Davidson, M. W. (n.d.). *Darkfield Illumination*. (Nikon Instruments Inc) Retrieved 4 7, 2021, from https://www.microscopyu.com/techniques/stereomicroscopy/darkfield-illumination

Chambers, W., Fellers, T. J., & Davidson, M. W. (n.d.). *Oblique Illumination*. (Nikon Instruments Inc) Retrieved 4 7, 2021, from https://www.microscopyu.com/techniques/stereomicroscopy/oblique-illumination

Cohn, C. (2015). *A Beginner's Guide To Upselling And Cross-Selling*. (Forbes) Retrieved 4 7, 2021, from https://www.forbes.com/sites/chuckcohn/2015/05/15/a-beginners-guide-to-upselling-and-cross-selling/?sh=1a725d182912

GeForce RTX 30 Series. (n.d.). (NVIDIA Corporation) Retrieved 4 7, 2021, from https://www.nvidia.com/en-us/geforce/graphics-cards/30-series/

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580-587. Columbus, OH, USA.

Godoy, D. (22 11, 2018). *Understanding binary cross-entropy / log loss: a visual explanation*. Retrieved from Towards Data Science: https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778. Las Vegas, NV, USA.

Hendrycks, D., & Gimpel, K. (2020). Gaussian Error Linear Units (GELUs). *arXiv:1606.08415*.

Hoffman, D. P., Slavitt, I., & Fitzpatrick, C. A. (2021). The promise and peril of deep learning in microscopy. *Nature Methods* **18**, 131-132.

IBM Watson products. (n.d.). (IBM) Retrieved 4 7, 2021, from https://www.ibm.com/watson/products-services

Introduction to Modulation Transfer Function. (n.d.). (Edmund Optics Inc.) Retrieved 4 7, 2021, from https://www.edmundoptics.com/knowledge-center/application-notes/optics/introduction-to-modulation-transfer-function/

Isola, P., Zhu, J.-Y., Zhu, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967-5976. Honolulu, HI, USA.

Kaderuppan, S. S., Wong, E. W. L., Sharma, A., & Woo, W. L. (2020). Smart Nanoscopy: A Review of Computational Approaches to Achieve Super-Resolved Optical Microscopy. *IEEE Access* **8**, 214801-214831.

Kaderuppan, S. S., Wong, E. W. L., Sharma, A., & Woo, W. L. (2022). Impact analysis of deep neural network training methodology on computational nanoscopy. In *Focus on Microscopy (FOM) 2022 – Online*. Retrieved April 19, 2022, from https://www.focusonmicroscopy.org/2022-program-online/?event_id=27.

Kızrak, A. (2019). Deep Learning for Image Segmentation: U-Net Architecture. (Medium) Retrieved 4 7, 2021, from https://heartbeat.fritz.ai/deep-learning-for-image-segmentation-u-net-architecture-ff17f6e4c1cf

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS)* 2012 **25**, 1097-1105. Lake Tahoe, NV, USA.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* **1**(4), 541-551.

Murphy, D. B., Oldfield, R., Schwartz, S., & Davidson, M. W. (n.d.). Introduction to Phase Contrast Microscopy. (Nikon - MicroscopyU) Retrieved 24 Apr, 2019, from https://www.microscopyu.com/techniques/phase-contrast/introduction-to-phase-contrast-microscopy

Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective.* Cambridge, Massachusetts, London, England: The MIT Press.

Nehme, E., Weiss, L. E., Michaeli, T., & Shechtman, Y. (2018). Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **5**(4), 458-464.

Ouyang, W., Aristov, A., Lelek, M., Hao, X., & Zimmer, C. (2018). Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology* **36**(5), 460-468.

pix2pix: Image-to-image translation with a conditional GAN. (19 6, 2021). (TensorFlow) Retrieved 4 7, 2021, from https://www.tensorflow.org/tutorials/generative/pix2pix

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Swish: a Self-Gated Activation Function. *arXiv:1710.05941v1*.

Ren, R., Guo, Z., Jia, Z., Yang, J., Kasabov, N. K., & Li, C. (2019). Speckle Noise Removal in Image-based Detection of Refractive Index Changes in Porous Silicon Microarrays. *Sci Rep* **9**(15001).

Rivenson, Y., Göröcs, Z., Günaydin, H., Zhang, Y., Wang, H., & Ozcan, A. (2017). Deep learning microscopy. *Optica* **4**(11), 1437-1443.

Robinson, P. C., & Davidson, M. W. (n.d.). Polarized Light Microscopy. (Nikon Instruments Inc.) Retrieved 4 7, 2021, from https://www.microscopyu.com/techniques/polarized-light/polarized-light-microscopy

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. (Computer Vision Group, Freiburg) Retrieved 30 Apr, 2019, from https://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a/

Scientific Volume Imaging. (n.d.). Huygens PSF Distiller. (Scientific Volume Imaging B.V) Retrieved 4 7, 2021, from https://svi.nl/Huygens-PSF-Distiller

Sze, V., Chen, Y.-H., Yang, T.-J., & Elmer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE* **105**(12), 2295-2329.

Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Günaydın, H., Bentolila, L. A., Kural, C. & Ozcan, A. (2019). Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat Methods* **16**, 103–110.

Welcome To Colaboratory. (n.d.). (Google Research) Retrieved 4 7, 2021, from https://colab.research.google.com/notebooks/intro.ipynb?utm_source=scs-index

Yang, S., & Lee, B.-U. (2015). Poisson-Gaussian Noise Reduction Using the Hidden Markov Model in Contourlet Domain for Fluorescence Microscopy Images. *PLoS ONE* **10**(9), e0136964.

Zawacki-Richter, O., Marín, V., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education - where are the educators? *International Journal of Educational Technology in Higher Education* **16**(39).

Zhongyi, H., Benzheng, W., Yuanjie, Z., Yilong, Y., Kejian, L., & Shuo, L. (2017). Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. *Scientific Reports* **7**, 4172.