

# Deep learning for single-shot autofocus microscopy

*Henry Pinkard  
Zachary Phillips  
Arman Babakhani  
Daniel Fletcher  
Laura Waller*

Electrical Engineering and Computer Sciences  
University of California at Berkeley

Technical Report No. UCB/EECS-2019-161

<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-161.html>

December 1, 2019



Copyright © 2019, by the author(s).  
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

#### Acknowledgement

The authors thank S.Y. Liu for providing the tissue sample and BIDS and its personnel in providing physical space, general logistical and technical support.

# Deep learning for single-shot autofocus microscopy

Henry Pinkard<sup>1,2,3,4\*</sup>, Zachary Phillips<sup>5</sup>, Arman Babakhani<sup>6</sup>, Daniel A. Fletcher<sup>7,8</sup>, and Laura Waller<sup>2,3,5,8</sup>

<sup>1</sup>Computational Biology Graduate Group, University of California, Berkeley, CA 94720, USA

<sup>2</sup>Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA

<sup>3</sup>Berkeley Institute for Data Science

<sup>4</sup>University of California San Francisco Bakar Computational Health Sciences Institute

<sup>5</sup>Graduate Group in Applied Science and Technology, University of California, Berkeley, CA 94720, USA

<sup>6</sup>Department of Physics, University of California, Berkeley, CA 94720, USA

<sup>7</sup>Department of Bioengineering and Biophysics Program, University of California, Berkeley, CA 94720, USA

<sup>8</sup>Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

\*Correspondence: [henry.pinkard@gmail.com](mailto:henry.pinkard@gmail.com)

May 13, 2019

**Maintaining an in-focus image over long time scales is an essential and non-trivial task for a variety of microscopy applications. Here, we describe a fast and robust auto-focusing method that is compatible with a wide range of existing microscopes. It requires only the addition of one or a few off-axis illumination sources (e.g. LEDs), and can predict the focus correction from a single image with this illumination. We designed a neural network architecture, the fully connected Fourier neural network (FCFNN), that exploits an understanding of the physics of the illumination in order to make accurate predictions with 2-3 orders of magnitude fewer learned parameters and less memory usage than existing state-of-the-art architectures, allowing it to be trained without any specialized hardware. We provide an open-source implementation of our method, in order to enable fast and inexpensive autofocus compatible with a variety of microscopes.**

Many biological experiments involve imaging samples in a microscope over long time periods or large spatial scales, making it difficult to keep the sample in focus. For example, when observing a sample over time periods of hours or days, thermal fluctuations can induce focus drift [8]. Or, when scanning and stitching together many fields-of-view (FoV) to form a high-content high-resolution image, a sample that is not sufficiently flat necessitates refocusing at each position [28]. Since it is often experimentally impractical or cumbersome to manually maintain focus, an automatic focusing mechanism is essential.

A variety of solutions have been developed for autofocus. Broadly, these methods can be divided into two classes: hardware-based schemes that attempt to directly measure the distance from the objective lens to the sample [1, 2, 7, 4, 29], and software-based methods that take one or more out-of-focus images and use them to deter-

mine the optimal focal position [19, 27, 11, 10]. The former usually require hardware modifications to the microscope (e.g. an infrared laser interferometry setup, additional cameras or optical elements), which can be expensive and place constraints on other aspects of the imaging system. Software-based methods, on the other hand, can be slow or inaccurate. For example, a software-based method might require a full focal stack, then use some measure of image sharpness to compute the ideal focal plane [19]. More advanced methods attempt to reduce the number of images needed to compute the correct focus [27], or use just a single out-of-focus image [11, 10]. However, existing single-shot autofocus methods either rely on nontrivial hardware modifications such as additional lenses and sensors [10] or are limited in their application to specialized regimes (i.e. can only correct defocus in one direction within a certain range) [11].

Here, we demonstrate a new computational imaging-based single-shot autofocus method that does not suffer from the limitations of previous methods. The only hardware modification it requires is the addition of one or more off-axis LEDs as an illumination source, from which we correct defocus based on a single out-of-focus image. Alternately, it can be used with no hardware modification on existing coded-illumination setups, which have been demonstrated for super-resolution [30, 14, 22], quantitative phase [30, 23, 14], and multi-contrast microscopy [31, 12].

The central idea of our method is that a neural network can be trained to predict how far out of focus a microscope is, based on a single image taken at arbitrary defocus under spatially coherent illumination. A related idea has recently been used to achieve fast, post-experimental digital refocusing in digital holography [25, 18]. Our work addresses autofocus in more general microscope systems, with both incoherent and coherent illumination. Intuitively, we believe this works because coherent illumination yields images with sharp features even when the

sample is out of focus. Thus, there is sufficient information in the out-of-focus image that an appropriate neural network can learn a function that maps these features to the correct defocus distance, regardless of the structural details of the sample. To test this idea we collected data using a Zeiss Axio Observer microscope ( $20\times$ , 0.5 NA) with the illumination source replaced by a programmable quasi-dome LED array [16]. The LED array provides a flexible means of source patterning, but is not necessary to implement this technique (see Note S1).

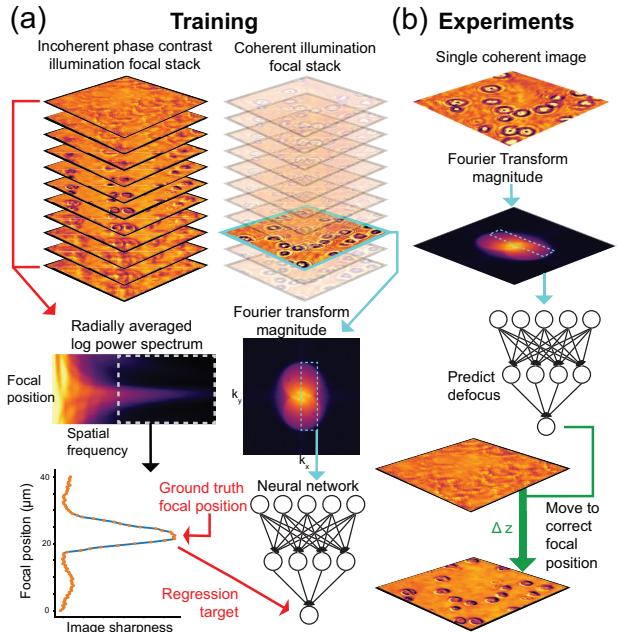
Though our experimental focus prediction requires only one image, we do need to collect focal stacks for training and validation. We use Micro-Magellan [17] for software control of the microscope, collecting focal stacks over  $60\ \mu\text{m}$  with  $1\ \mu\text{m}$  spacing, distributed symmetrically around the true focal plane. For each part of the sample, we collect focal stacks with two different types of illumination: spatially coherent (i.e. a single LED) and (nearly) spatially incoherent (i.e. many LEDs at once).

The incoherent focal stack is used for computing the ground truth focal position, since the reduced coherence results in sharp images only when the sample is in focus. Sharpness can be quantified for each image in the stack by summing the high-frequency content of its radially averaged log power spectrum. The maximum of the resultant curve was chosen as the ground truth focal position for the stack (Fig. 1a, left). Because this ground truth value is calculated by a deterministic algorithm, this paradigm scales well to large amounts of training data. For transparent samples, the incoherent image stack was captured with asymmetric illumination in order to create phase contrast [13]. In our case, this was achieved by using the LED array to project a half annulus source pattern [23]; however, any asymmetric source pattern should suffice.

The coherent focal stack is used one image at a time as the input to the network, which is trained to predict the ground truth focal position (Fig. 1). Since the network only takes a single image as its input, each image in the stack represents a separate training example. In our case, the coherent focal stack was captured by illuminating the sample with a single off-axis LED. In the case of arbitrary illumination control (e.g. with an LED array) different illumination angles or patterns may perform differently for a given amount of training data. Supplementary Fig. S1 compares performance for varying single-LED illumination angles as well as multi-LED patterns. For simplicity, here we consider only the case of a single LED positioned at an angle of 24 degrees relative the optical axis.

Our neural network architecture for predicting defocus (described in detail in Note S3), which we call the fully connected Fourier neural network (FCFNN), differs substantially from the convolutional neural networks (CNNs) typically used in image processing tasks [18, 25, 26] (Note S4). We reasoned that singly-scattered light would contain the most useful information for defocus prediction, and thus we designed the FCFNN to exclude parts of the captured image’s Fourier transform that are outside the single-scattering region for off-axis illumination (Fig. S2).

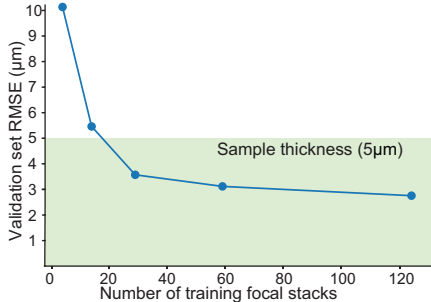
This results in 2-3 orders of magnitude fewer free parameters and memory usage during training than state-of-the-art CNNs (Table S1). Hence, our network can be trained on a desktop CPU in a few hours with no specialized computing hardware, making our method more reproducible, without sacrificing quality.



**Figure 1: Training and defocus prediction.** a) Training data consists of two focal stacks for each part of the sample, one with incoherent (phase contrast) illumination, and one with off-axis coherent illumination. Left: The high spatial frequency part of each image’s power spectrum from the incoherent stack is used to compute a ground truth focal position. Right: For each coherent image in the stack, the central pixels from the magnitude of its Fourier transform are used as input to a neural network that is trained to predict defocus. The full set of training examples is generated by repeating this process for each of the coherent images in the stack. b) After training, experiments need only collect a single coherent image, which is fed through the same pipeline to predict defocus. The microscope’s focus can then be adjusted to correct defocus.

Briefly, the FCFNN (Fig. 1a, right) begins with a single coherent image. This image is Fourier transformed, and the magnitude of the complex-valued pixels in the central part of the Fourier transform are reshaped into a single vector, which is used as the input to a trainable fully connected neural network. After the network has been trained, it can be used to correct defocus during an experiment by capturing a single image at an arbitrary defocus under the same coherent illumination. The network predicts defocus distance, then the microscope moves to the correct focal position (Fig. 1b).

Training with 440 focal stacks took 1.5 hours on a desk-

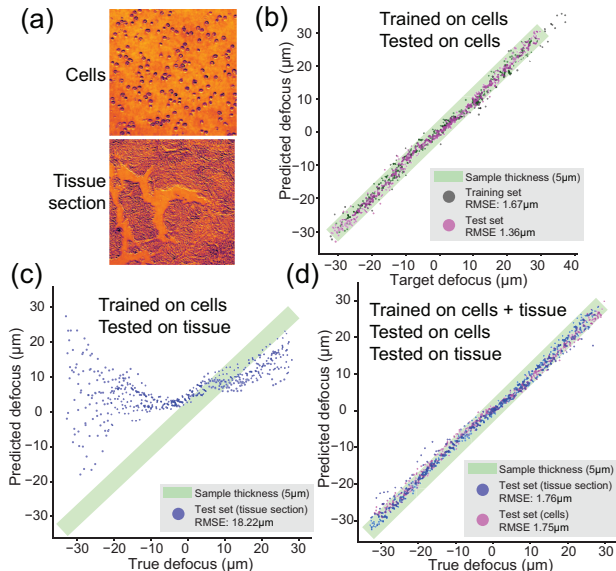


**Figure 2: Performance vs. amount of training data.** Defocus prediction performance (measured by validation RMSE) improves as a function of the number of focal stacks used during the training phase of the method.

top CPU or 30 minutes on a GeForce GTX 1080 Ti GPU, in addition to 2 minutes per focal stack for pre-computing ground truth focal planes and Fourier transforms. A single prediction from a 2048x2048 image takes  $\sim 50$  ms on a desktop CPU. We were able to train FCFNNs capable of predicting defocus with root-mean-squared error (RMSE) smaller than the axial thickness of the sample (cells). Figure 2 shows how this performance varies based on the number of focal stacks used to train the network, where each focal stack contained 60 planes spaced  $1 \mu\text{m}$  apart, distributed symmetrically around the true focal plane. Note that this curve could be quite different depending on the sample type and quality of training data.

To test the performance of our method across different samples, we collected data from two different sample types (Fig. 3a): white blood cells attached to coverglass, and an unstained  $5 \mu\text{m}$  thick mounted histology tissue section. When the network is *trained* on images of cells, then *tested* on different images of cells, it performs very well (Fig. 3b). However, when the network is trained on images of cells, then tested on a different sample type (tissue), it performs poorly (Fig. 3c). Hence, the method does not inherently generalize to new sample types. To solve this problem, we diversify the training data. We add a smaller amount of additional training data from the new sample type (here 130 focal stacks of tissue data, in addition to the 440 stacks of cell data it was originally trained on). With this training, the network performs well on both tissue and cell samples. Hence, our method can generalize to other sample types, without sacrificing performance on the original sample type (Fig. 3d). The best performing neural networks in other domains are typically trained on large and varied datasets [9]. Thus, if the FCFNN is trained on defocus data from a variety of sample types, it should generalize to new types more easily.

Empirically, we discovered that discarding the phase of the Fourier transform and using only the magnitude as the input to the network dramatically boosted performance. To illustrate, Fig. 4a compares networks trained using the Fourier transform magnitude as input vs. those trained on the argument of the Fourier transform phase.

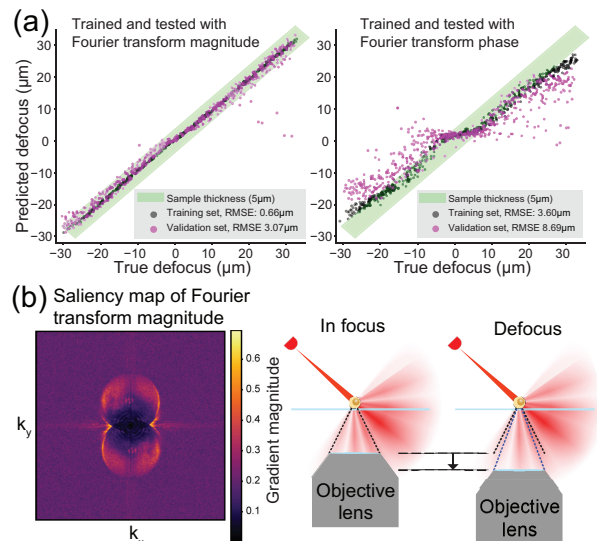


**Figure 3: Generalization to new sample types.** a) Representative images of cells and tissue section samples. b) A network trained on focal stacks of cells predicts defocus well in other cell samples, c) but fails at predicting defocus in tissue sections. d) After adding limited additional training data on tissue section samples, however, the network can learn to predict defocus well in both sample types.

Not only were networks using magnitude able to better fit the training data, they also generalized better to a validation set. This suggests useful information for predicting defocus in a coherent intensity image is relatively more concentrated in the magnitude compared to the phase of its Fourier transform. We speculate that this is because the phase of the intensity image generally relates more to spatial position of features (which is unimportant for focus prediction), whereas the magnitude contains more information about how they are transformed by the imaging system.

In order to understand what features of the images the network learns to make predictions from, we compute a saliency map for a network trained using the entire uncropped Fourier transform, shown in Fig. 4b. The saliency map attempts to identify which parts of the input the network is using to make decisions, by visualizing the gradient of a single unit within the neural network with respect to the input [20]. The idea is that the output unit is more sensitive to features with a large gradient and thus these have a greater influence on prediction. In our case, the gradient of the output (i.e. the defocus prediction) was computed with respect to the the Fourier transform magnitude. Averaging the magnitude of the gradient image over many examples clearly shows that the network recognizes specific parts of the the overlapping two-circle structure (Fig. 4b) that is typical for an image formed by coherent off-axis illumination (Fig. S2) [5]. In particular,

the regions at the edges of the circles have an especially large gradient. These areas correspond to the highest angles of light collected by the objective lens. Intuitively, this makes sense because changing the focus will lead to proportionally greater changes in the light collected at the highest angles (Fig. 4b).



**Figure 4: Understanding how the network predicts defocus.** a) A network trained on the magnitude of the Fourier transform of the input image performs better than one trained on the argument of the phase of the Fourier transform. b) Left: a saliency map (the magnitude of the defocus prediction’s gradient with respect to the Fourier transform magnitude) shows the edges of the object spectrum have the strongest influence on defocus predictions. Right: edges correspond to high-angle scattered light, which may not be captured off-focus, providing significant changes in the input image with defocus.

To summarize, we have demonstrated a method for training and using neural networks for single-shot autofocus, with analysis of design principles and practical trade-offs. The method works with different sample types and is simple to implement on a conventional transmitted light microscope, requiring only the addition of off-axis illumination and no specialized hardware for training the neural network. We introduced the FCFNN, a neural network architecture that incorporates knowledge of the physics of the imaging system into its design, thereby making it orders of magnitude more efficient in terms of parameter number and memory requirements during training than general state-of-the-art approaches for image processing.

See Supplement 1 for supporting content. The code needed to implement this technique and reproduce all figures in this manuscript can be found in the Jupyter notebook: 1. H. Pinkard, "Single-shot autofocus microscopy using deep learning-code," (2019), <https://doi.org/10.6084/m9.figshare.7453436.v1>. Due to its large size, the corresponding data is available upon

request.

This project was funded by Packard Fellowship and Chan Zuckerberg Biohub Investigator Awards to Laura Waller and Daniel Fletcher, STROBE: A NSF Science & Technology Center under Grant No. DMR 1548924, a NIH R01 grant to Daniel Fletcher, a NSF Graduate Research Fellowship awarded to Henry Pinkard, and a Berkeley Institute for Data Science/UCSF Bakar Computational Health Sciences Institute Fellowship awarded to Henry Pinkard with support from the Koret Foundation, the Gordon and Betty Moore Foundation through Grant GBMF3834 and the Alfred P. Sloan Foundation through Grant 2013-10-27 to the University of California, Berkeley.

The authors thank S.Y. Liu for providing the tissue sample and BIDS and its personnel in providing physical space, general logistical and technical support.

## Note S1: Practical aspects of implementing on a new microscope

**Hardware/Illumination** In order to generate data using our method, the microscope must be able to image samples with two different contrast modalities: one with spatially incoherent illumination for computing ground truth focal position from image blur, and a second coherent or nearly coherent illumination (i.e. one or a few LEDs) as input to the neural network. The incoherent illumination can be accomplished with the regular bright-field illumination of a transmitted light microscope in the case of samples that absorb light. In the case of transparent phase-only samples (like the ones used in our experiments), incoherent phase contrast can be created by using any asymmetric illumination pattern. We achieved this by using a half-annulus pattern on a programmable LED illuminator, but this specific pattern is not necessary. The same effect can be achieved by blocking out half of the illumination aperture of a microscope condenser with a piece of cardboard[13] or other means of achieving asymmetric illumination. The asymmetric incoherent illumination is only needed for the generation of training data, so it does not need to be permanent.

For the spatially coherent illumination, a single LED pointed at the sample from an oblique angle (i.e. not directly above) generates sufficient contrast, as shown in the main paper. However, our experiments with different multi-LED patterns (see note S2) indicate that a series of LEDs arranged in a line might be even better for this purpose.

**Software** Our implementation used a stack of open source acquisition control software based on Micro-Manager [6] and the plugin for high-throughput microscopy, Micro-Magellan [17]. Both are agnostic to specific hardware, and can thus be implemented on any microscope to easily collect training data. Automated LED illumination in Micro-Manager can be con-

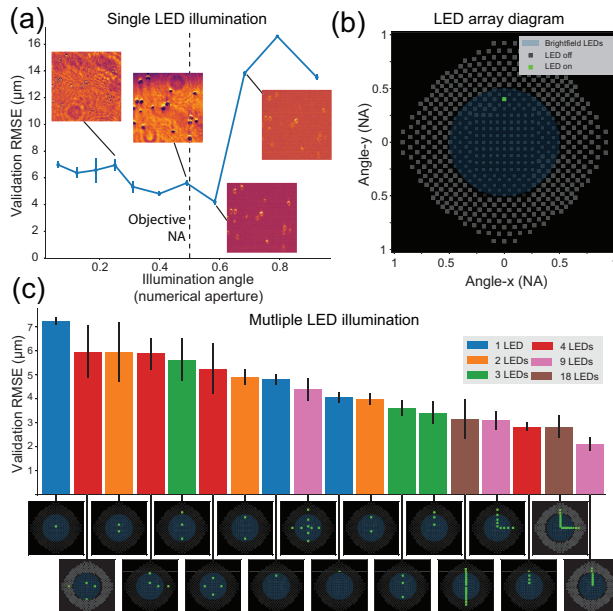
figured using a simple circuit connected to an Arduino and the Micro-Manager device adapter to control digital IO. Large numbers of focal stacks can be collected in an automated way using the 3D imaging capabilities of Micro-Magellan, and a Python reader for Micro-Magellan data allows for easy integration of data into deep learning frameworks. Examples of this can be seen in the Jupyter notebook: 1. H. Pinkard, "Single-shot autofocus microscopy using deep learning-code," (2019), <https://doi.org/10.6084/m9.figshare.7453436.v1>.

**Other imaging geometries** Although we have demonstrated this technique on a transmitted light microscope with LED illumination, in theory there is no reason why it couldn't be applied to other coherent illuminations and geometries. For example, using a laser instead of an LED as a coherent illumination source should be possible with minimal modification. We've also demonstrated the technique using relatively thin samples. Autofocusing methods like ours are generally not directly applicable to thick samples, since it is difficult to define the ground truth focal plane of a thick sample in a transmitted light configuration. However, in principle it is possible that these methods could be used in a reflected light geometry, where the "true" focal plane corresponds to the top of the sample.

## Note S2: Choosing an illumination pattern

Although the network is capable of learning to predict defocus from images taken under the illumination of a single off-axis LED, as shown in the main paper, different angles or combinations of angles of illumination might contain more useful information for prediction. Better performance can make the prediction task more accurate, easier and able to be learned with less training data. Since our experimental setup uses a programmable LED array quasi-dome as an illumination source [16], we can choose the source patterns at will to test this. First, restricting the analysis to one LED at a time, we tested how the angle of the single-LED illumination affects performance (Fig. S1a). We found that performance improves with increasing angle of illumination, up to a point where performance rapidly degrades. This drop-off occurs in the 'darkfield' region (where the illumination angle is larger than the objective's NA), likely due to the low signal-to-noise ratio (SNR) of the higher-angle darkfield images (see inset images in Fig. S1a). This drop in SNR could plausibly be caused by either a decrease in the number of photons hitting the sample from higher-angle LEDs, or a drop in the content of the sample itself at higher frequencies. To rule out the first possibility, we compensated for the expected number of photons incident on a unit area of the sample, which is expected to fall off approximately proportional to  $\frac{1}{\cos(\theta)}$ , where  $\theta$  is the angle of illumination relative to the optical axis [15]. The dataset

used here increases exposure time in proportion to  $\cos(\theta)$  in order to compensate for this. Thus, the degradation of performance at high angles is most likely due to the amount of high frequency content in the sample itself at these angles and therefore might be sample-specific.

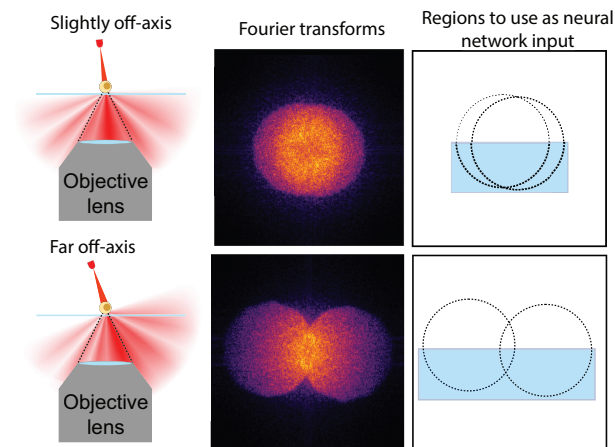


**Figure S1: Illumination design.** a) Increasing the numerical aperture (NA) (i.e. angle relative to the optical axis) of single-LED illumination increases the accuracy of defocus predictions, up to a point at which it degrades. b) Diagram of LED placements in NA space for our LED quasi-dome. c) Defocus prediction performance for different illumination patterns. Patterns with multiple LEDs in an asymmetric line show the lowest error.

Next, we tested 18 different single or multi-LED source patterns chosen from within the distribution of  $x$  and  $y$  axis-aligned LEDs available on our quasi-dome (Fig. S1b,c). Since the light from any two LEDs is mutually incoherent, single-LED images can be added digitally to synthesize the image that would have been produced with multiple-LED illumination. This enabled us to computationally experiment with different illumination types on the same sample. Figure S1c shows the defocus prediction performance of various patterns of illumination. The best performing patterns were those that contained multiple LEDs arranged in a line. Given that specific parts of the Fourier transform contain important information for defocus prediction and that these areas will move to different parts of Fourier space with different angles of illumination, we speculate that the line of LEDs helps to spread relevant information for defocus prediction into different parts of the spectrum. Although this analysis demonstrates more and higher angle LED patterns seem to yield superior performance, there are potential caveats: In the former case, it could fail to hold when applied to a denser sample (i.e. not a sparse distribution of cells).

In the latter, there is the cost of the increase in exposure time needed to acquire such images.

### Note S3: Fully connected Fourier neural network architecture



**Figure S2: Fourier Transform regions to use as network input.** Off-axis illumination with a coherent point source at an angle within the numerical aperture of the collection objective produces a characteristic two-circle structure in the log magnitude of the Fourier transform of the captured image. As the angle of illumination increases, these circles move further apart. Information about single-scattering events is confined within these circles. The blue regions represent the pixels that should be cropped out and fed into the neural network architecture.

The fully connected Fourier neural network (FCFNN) begins with a single coherent intensity image captured by the microscope. This image is Fourier transformed, and the magnitude of the complex-valued pixels in the central part of the Fourier transform are reshaped into a single vector. The useful part of the Fourier transform is directly related to the angle of coherent illumination (Fig. S2). A coherent illumination source such as an LED that is within the range of brightfield angles for the given objective (i.e. at an angle less than the maximum captured angle as determined the objective’s NA) will display a characteristic 2-circle structure in its Fourier transform magnitude. The two circles contain information corresponding to the singly-scattered light from the sample and move farther apart as the angle of the illumination increases. The neural network input should consist of half of the pixels in which these circles lie, because as the saliency map in Fig. 4b of the main text demonstrates, they contain the useful information for predicting defocus. Only half the pixels are needed because the Fourier transform of a real-valued input (i.e. an intensity image)

has symmetric magnitudes, so the other half contain redundant information. These circles move with changing illumination angle, so they angle of illumination and relevant pixels must be selected together.

After cropping out the relevant pixels and reshaping them into a vector, the vector is normalized to have unit mean in order to account for differences in illumination brightness, and it is then used as the input layer of a neural network trained in TensorFlow [3]. The learnable part of the FCFNN consists of a series of small (100 unit) fully connected layers, followed by a single scalar output (the defocus prediction).

We experimented with several hyperparameters and regularization methods to improve performance on our training data. The most successful of these were: 1) Changing the number and width of the fully connected layers. We started small and increased both until this ceased to improve performance, which occurred with 10 fully connected layers of 100 units each. 2) Applying dropout [21] to the vectorized Fourier transform input layer (but not other layers) to prevent overfitting to specific parts of the Fourier transform. 3) Dividing the input image into patches and averaging the predictions over each patch. This gave best performance when we divided the 2048x2048 image into 1024x1024 patches. 4) Using only the central part of the Fourier transform magnitude as an input vector. We manually tested how much of the edges to crop out. 5) Early stopping - when loss on a held out validation set ceased to improve - helped test performance.

In general, we observed better performance training on noisier and more varied inputs (i.e. cells at different densities, particularly lower densities, and different exposure times). This is consistent with other results in deep learning, where adding noise to training data improves performance [24].

### Note S4: Comparison of FCFNNs and CNNs

The FCFNN differs substantially from the convolutional neural networks (CNNs) used as the state-of-the-art in image processing tasks. Typically, to solve a many-to-one regression task of predicting a scalar from an image, as in the defocus prediction problem here, CNNs first use a series of convolutional blocks with learnable weights to learn to extract relevant features from the image and then often will pass those features through a series of fully connected layers to generate a scalar prediction [9]. Here, we have replaced the feature learning part of the network with a deterministic feature extraction module that uses only the physically-relevant parts of the Fourier Transform.

Deterministically downsampling images into feature vectors early in the network reduces the required number of learnable weights and the memory used by the backpropagation algorithm to compute gradients during



**Table S1: Comparison of number of learnable weights and memory usage**

| Architecture           | Image size | # of learnable weights | memory training per example (MB or KB) |
|------------------------|------------|------------------------|--|
| FCFNN (ours)           | 1024x1024  | $6.3 \times 10^6$      | 18 KB                                  |
| CNN (Ren et. al[18])   | 1000x1000  | $2.5 \times 10^8$      | 111 MB                                 |
| FCFNN (ours)           | 1000x1000  | $6.0 \times 10^6$      | 17 KB                                  |
| CNN (Yang et. al [26]) | 84x84      | $2.9 \times 10^7$      | 1.3 MB                                 |
| FCFNN (ours)           | 84x84      | $3.6 \times 10^4$      | 3 KB                                   |

training by 2-3 orders of magnitude. Table S1 shows a comparison between our FCFNN and two CNNs used for comparable tasks. The architecture used by Ren et al. is used for post-acquisition defocus prediction in digital holography and the architecture of Yang et al. is used for post-acquisition classification of images as in-focus or out-of-focus. Both use the conventional CNN paradigm of a series of convolutional blocks followed by one or more fully connected layers.

Similar to CNNs, our FCFNN can also incorporate information from different parts of the full image. CNNs do this with a series of convolutional blocks that gradually expand the size of the receptive fields. The FCFNN does this inherently by use of the Fourier transform. Each pixel in the Fourier transform corresponds to a sinusoid of a certain frequency and orientation in the original image, so its magnitude draws information from every pixel.

## References

[1] Nikon perfect focus. <https://www.microscopyu.com/applications/live-cell-imaging/nikon-perfect-focus-system>. Accessed: 2019-01-23.

[2] Zeiss definite focus. <https://www.zeiss.com/microscopy/us/products/light-microscopes/axio-observer-for-biology/definite-focus.html>. Accessed: 2019-01-23.

[3] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg Cor-

rado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Man, Rajat Monga, Sherry Moore, Derek Murray, Jon Shlens, Benoit Steiner, Ilya Sutskever, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Oriol Vinyals, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Technical report, Google Research, 2015.

[4] M. Bathe-Peters, P. Annibale, and M. J. Lohse. All-optical microscope autofocus based on an electrically tunable lens and a totally internally reflected IR laser. *Optics Express*, 26(3):2359, 2018.

[5] Regina Eckert, Zachary F. Phillips, and Laura Waller. Efficient illumination angle self-calibration in Fourier ptychography. *Applied optics*, 57(19):5434–5442, 2018.

[6] Arthur Edelstein, Nenad Amodaj, Karl Hoover, Ron Vale, and Nico Stuurman. Computer Control of Microscopes Using Micro-Manager. *Current Protocols in Molecular Biology*, (14):1–17, 2010.

[7] Kaikai Guo, Jun Liao, Zichao Bian, Xin Heng, and Guoan Zheng. InstantScope: a low-cost whole slide imaging system with instant focal plane detection. *Biomedical Optics Express*, 6(9):3210, 2015.

[8] Marko Kreft, Matja Stenovec, and Robert Zorec. Focus-drift correction in time-lapse confocal imaging. *Annals of the New York Academy of Sciences*, 1048:321–330, 2005.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2 2011.

[10] J Liao, L Bian, Z Bian, Z Zhang, C Patel, K Hoshino, Y C Eldar, and G Zheng. Single-frame rapid autofocusing for brightfield and fluorescence whole slide imaging. *Biomedical Optics Express*, 7(11):4763–4768, 2016.

[11] Jun Liao, Yutong Jiang, Zichao Bian, Bahareh Mahrou, Aparna Nambiar, Alexander W Magsam, Kaikai Guo, Shiyao Wang, Yong ku Cho, and Guoan Zheng. Rapid focus map surveying for whole slide imaging with continuous sample motion. *Opt. Lett.*, 42(17):3379–3382, 2017.

[12] Ziji Liu, Lei Tian, Sijia Liu, and Laura Waller. Real-time brightfield, darkfield, and phase contrast imaging in a light-emitting diode array microscope. *Journal of Biomedical Optics*, 19(10):106002, 2014.

- [13] Shalin B Mehta and Colin J R Sheppard. Quantitative phase-gradient imaging at high resolution with asymmetric illumination-based differential phase contrast. *Optics Letters*, 34(13):1924, 2009.
- [14] Xiaoze Ou, Roarke Horstmeyer, Changhui Yang, and Guoan Zheng. Quantitative phase imaging via Fourier ptychographic microscopy. *Optics letters*, 38(22):4845–8, 2013.
- [15] Zachary F. Phillips, Michael V. D’Ambrosio, Lei Tian, Jared J. Rulison, Hurshal S. Patel, Nitin Sadras, Aditya V. Gande, Neil a. Switz, Daniel a. Fletcher, and Laura Waller. Multi-Contrast Imaging and Digital Refocusing on a Mobile Microscope with a Domed LED Array. *Plos One*, 10(5):e0124938, 2015.
- [16] Zachary F. Phillips, Regina Eckert, and Laura Waller. Quasi-Dome: A Self-Calibrated High-NA LED Illuminator for Fourier Ptychography. In *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*, volume IW4E.5, 2017.
- [17] Henry Pinkard, Nico Stuurman, Kaitlin Corbin, Ronald Vale, and Matthew F Krummel. MicroMagellan: open-source, sample-adaptive, acquisition software for optical microscopy. *Nature Methods*, 13(10):807–809, 9 2016.
- [18] Zhenbo Ren, Zhimin Xu, and Edmund Y. Lam. Learning-based nonparametric autofocusing for digital holography. *Optica*, 5(4):337, 2018.
- [19] Feimo Shen, Louis Hodgson, and Klaus Hahn. Digital Autofocus Methods for Automated Microscopy. In *Methods in Enzymology*, volume 414, pages 620–632. Academic Press, 2006.
- [20] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv preprint*, 1312.6034, 12 2013.
- [21] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.
- [22] Lei Tian and Laura Waller. 3D intensity and phase imaging from light field measurements in an LED array microscope. *Optica*, 2(2):104–111, 2015.
- [23] Lei Tian and Laura Waller. Quantitative differential phase contrast imaging in an LED array microscope. *Optics Express*, 23(9):11394, 2015.
- [24] Pascal Vincent and Hugo Larochelle. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pierre-Antoine Manzagol. *Journal of Machine Learning Research*, 11:3371–3408, 2010.
- [25] Yichen Wu, Yair Rivenson, Yibo Zhang, Zhensong Wei, Harun Günaydin, Xing Lin, and Aydogan Ozcan. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica*, 5(6):704, 6 2018.
- [26] Samuel J. Yang, Marc Berndl, D. Michael Ando, Mariya Barch, Arunachalam Narayanaswamy, Eric Christiansen, Stephan Hoyer, Chris Roat, Jane Hung, Curtis T. Rueden, Asim Shankar, Steven Finkbeiner, and Philip Nelson. Assessing microscope image focus quality with deep learning. *BMC Bioinformatics*, 19(1):28, 2018.
- [27] Siavash Yazdanfar, Kevin B Kenny, Krenar Tasimi, Alex D Corwin, Elizabeth L Dixon, and Robert J Filkins. Simple and robust image-based autofocusing for digital microscopy. *Optics express*, 16(12):8670–8677, 2008.
- [28] Mark D. Zarella, Douglas Bowman, Famke Aeffner, Navid Farahani, Albert Xthona, Syeda Fatima Absar, Anil Parwani, Marilyn Bui, and Douglas J. Hartman. A practical guide to whole slide imaging a white paper from the digital pathology association. *Archives of Pathology and Laboratory Medicine*, 143(2):222–234, 2019.
- [29] Xin Zhang, Fei Zeng, Yi Li, and Yanfeng Qiao. Improvement in focusing accuracy of DNA sequencing microscope with multi-position laser differential confocal autofocus method. *Optics Express*, 26(2):887, 2018.
- [30] Guoan Zheng, Roarke Horstmeyer, Changhui Yang, Guoan Zheng, and Changhui Yang. Wide-field, high-resolution Fourier ptychographic microscopy. *Nature photonics*, 7(9):739–745, 2013.
- [31] Guoan Zheng, Christopher Kolner, and Changhui Yang. Microscopy refocusing and dark-field imaging by using a simple LED array. *Optics Letters*, 36(20):3987, 2011.