Research Article | Vol. 12, No. 3 / 1 March 2021 / Biomedical Optics Express 1683
Biomedical Optics EXPRESS

Deep learning for label-free nuclei detection from implicit phase information of mesenchymal stem cells

ZHENGYUN ZHANG,[1,6] KIM WHYE LEONG,[2] KRYSTYN VAN VLIET,[1,3] GEORGE BARBASTATHIS,[1,4] AND ANDREA RAVASIO[5,7]

[1]*BioSyM IRG, Singapore-MIT Alliance for Research and Technology (SMART) Centre, 1 CREATE Way, #04-13/14 Enterprise Wing, Singapore 138602, Singapore*
[2]*Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, Singapore 117558, Singapore*
[3]*Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[4]*Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[5]*Institute of Biological and Medical Engineering, School of Engineering, Medicine and Biological Sciences, Pontificia Universidad Católica de Chile, Vicuña Makenna 4860, Macul, Santiago, Chile*
[6]*zaltor@gmail.com*
[7]*andrea.ravasio@uc.cl*

Abstract: Monitoring of adherent cells in culture is routinely performed in biological and clinical laboratories, and it is crucial for large-scale manufacturing of cells needed in cell-based clinical trials and therapies. However, the lack of reliable and easily implementable label-free techniques makes this task laborious and prone to human subjectivity. We present a deep-learning-based processing pipeline that locates and characterizes mesenchymal stem cell nuclei from a few bright-field images captured at various levels of defocus under collimated illumination. Our approach builds upon phase-from-defocus methods in the optics literature and is easily applicable without the need for special microscopy hardware, for example, phase contrast objectives, or explicit phase reconstruction methods that rely on potentially bias-inducing priors. Experiments show that this label-free method can produce accurate cell counts as well as nuclei shape statistics without the need for invasive staining or ultraviolet radiation. We also provide detailed information on how the deep-learning pipeline was designed, built and validated, making it straightforward to adapt our methodology to different types of cells. Finally, we discuss the limitations of our technique and potential future avenues for exploration.

1. Introduction

Mesenchymal stem cells (MSCs) [1] are undifferentiated stromal cells commonly derived from bone marrow and adipose tissue. Unlike pluripotent embryonic stem cells, MSCs are multipotent and thus are limited in their differentiation capacity—they mainly become osteocytes (bone), chondrocytes (cartilage) and adipocytes (fat). Despite this limitation, MSCs have been shown to hold great potential as a cell therapeutic agent [2], evident in the sheer number of ongoing clinical trials for a variety of diseases and conditions. Routine practice in a clinical trial involves the cultural expansion of MSCs from autologous (derived from the same patient) or allogeneic (derived from a healthy donor) sources, followed by the administration of the MSCs either intravenously or in the extravascular compartment. Since appropriate cell dosage and condition are crucial for successful treatment, the ability to count MSCs accurately as well as track their growth will be instrumental for higher quality clinical research. Although MSCs are nearly transparent in bright-field images, they can be observed noninvasively using phase imaging

#420266 https://doi.org/10.1364/BOE.420266
Journal © 2021 Received 19 Jan 2021; accepted 21 Feb 2021; published 1 Mar 2021

Biomedical Optics EXPRESS

methods, as shown in Fig. 1. However, MSCs grown to high density in a two-dimensional environment, *e.g.*, in culture flasks, can result in their clustering, making it hard to distinguish cells in close proximity. Conventional cell counting methods such as cell segmentation algorithms and manual hemocytometer-assisted counting may not be able to accurately quantify the number of MSCs, whereas commercial cell counters like flow cytometers are severely invasive due to the need to detach cells from the substrate. While fluorescent staining of the nucleus is a rigorous approach to cell counting, as nuclei generally do not touch each other (as shown in Fig. 1), it is also invasive; the teratogenic nature of the fluorescent dye and the ultraviolet radiation needed for common nucleic acid stains may alter the physiology of the MSCs. On the other hand, non-invasive methods such as coherent anti-Stokes Raman spectroscopy (CARS) [3] or holography [4] require complicated optical systems not easily implementable in standard commercial microscopes commonly found in clinical laboratories. Finally, methods based on cell confluency (*e.g.*, [5]) rely on rough estimation of average cell size and therefore are prone to miscounting in heterogeneous, growing or differentiating cell cultures. Hence, in recent years, various approaches based on phase contrast [6] and holographic imaging [7,8] have been implemented to estimate cell numbers grown on flat culturing substrates and on microcarriers [9].

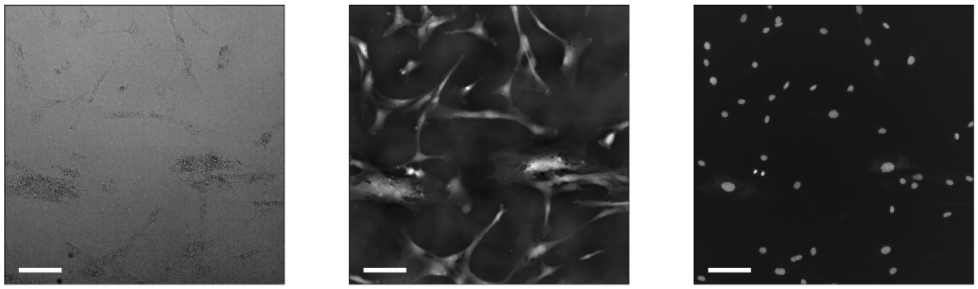

Fig. 1. Bright-field image, quantitative phase image obtained by ±25 µm defocusing, and nuclei-stained fluorescence image of a colony of MSCs. Scale bar is 100 µm.

More recently, deep learning has become a popular tool for biological imaging applications. Deep convolutional neural networks (CNNs) came to the forefront of machine learning after such a network [10] achieved the highest accuracy score (25% error rate) in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Advances in computational power as well as improvements in stochastic gradient descent methods and automatic differentiation helped bring deep neural networks to the forefront, with ImageNet error rates easily reaching single digit numbers for all the contestants in 2017. In bio-imaging, deep neural networks have been used for segmentation [11], detection [12,13] and counting [14] based on stained images. For label-free applications, quantitative phase images from digital holographic microscopy has been used to predict nucleic acid stains [15], autofluorescence images have been used to predict histological staining [16], and focal stacks have been used to predict various fluorescence labels [17].

Here, we propose an original deep-learning approach to MSC counting and nuclei quantification based on contrast induced by optical thickness variations in the sample. Physics of light propagation dictate that these thickness variations in a transparent specimen induces deterministic image patterns at various levels of defocus; a focal stack of images is used as input for many quantitative phase imaging techniques for this reason. Notably, the nucleus is the largest organelle in the cell, and therefore it emerges from the otherwise flat cytosol of MSCs. Basic features of the nucleus, such as size and shape, can also be used as a proxy for cellular states that correlate with therapeutic potency [18]. We hypothesize that, since the focal stack contains the information needed to reconstruct the optical thickness, this information can be directly utilized by a trained

Biomedical Optics EXPRESS

neural network to predict the location and shape of nuclei in the specimen without the need to explicitly reconstruct the quantitative phase image. The latter approach would've been expensive and could introduce bias due to the well-known need for mathematical priors in quantitative phase reconstruction; priors are needed to fill in missing information crucial for a visually pleasing quantitative phase image, but this information is likely not necessary for nuclei detection and quantification.

Our pipeline consists of a neural network trained using sets of images, each capturing a specimen focused at different depths, to predict whether each pixel in an input set of images falls within a nucleus. The output of the neural network is then fed into by an automated detection framework to detect, locate and characterize nuclei. Our approach is label-free, does not require contrast modalities such as phase contrast or differential interference contrast (DIC), and relies only on microscope equipment readily available in commercial microscopes. In this paper, we will explore ways to design such neural networks and detail solutions to the engineering challenges involved in automating the cell counting process. We will give examples for when our technique accurately counts the number of cells and quantifies their shape statistics, and we will also discuss ways to improve results for more challenging cases. Our goal is to provide sufficient details so that our approach can be used as a generalized scheme to develop fast and noninvasive techniques for characterization and counting of the various cell types used in biomedical applications.

2. Design

Our imaging and processing pipeline design originates from the idea that the optical thickness profile, *i.e.*, the phase, of our otherwise transparent specimen provides label-free imaging contrast, without the need for custom hardware. When the specimen is illuminated by a collimated beam, information about its phase profile is embedded in the axial evolution of intensity images (*i.e.*, diffraction patterns) captured at different levels of defocus; this principle is what guides many explicit phase imaging methods, such as those based on the transport of intensity equation (TIE) [19–23] or contrast transfer functions (CTFs) [24,25]. From preliminary experiments with quantitative phase imaging, we found that it was sometimes possible to visually locate nuclei in MSCs due to the presence of nucleoli, but this task was difficult to automate while guaranteeing accurate results. Furthermore, quantitative phase imaging methods are by themselves computationally challenging due to ill-posedness of the measurements. Therefore, we decided to leverage deep learning and skip the explicit reconstruction of the phase profile, feeding the neural network with raw images that would normally be used as input to quantitative phase imaging algorithms so that the network can exploit the implicit phase profile information embedded therein to predict which pixels in the image correspond to nuclei.

As shown in Fig. 2, the analytic pipeline starts with patches of images containing raw diffracted intensities. These patches are then fed into a neural network trained to predict whether the center of each patch is located inside the nucleus of an MSC. The raw output of the neural network is then fed into a detection process that locates blob-like nuclei; an ellipse is fit to each blob, and the average pixel value inside the ellipse is used in a final sieving step to remove false positives. In addition to this design, we also considered several alternative designs:

• *end-to-end regression network to compute cell counts*

We could design a regression network that directly outputs cell counts from a set of input images. It would be much easier to use than our two-step approach, but it would be much harder to analyze the network to determine how it can be improved, due to the lack of meaningful structure in the resulting network. Furthermore, such networks must be trained on larger patches containing multiple nuclei, and thus would possibly require more training data. An end-to-end network might also more easily fall into undesirable local minima, due to the required increased complexity needed.

• *end-to-end regression network to output nuclei locations/statistics*

Biomedical Optics EXPRESS

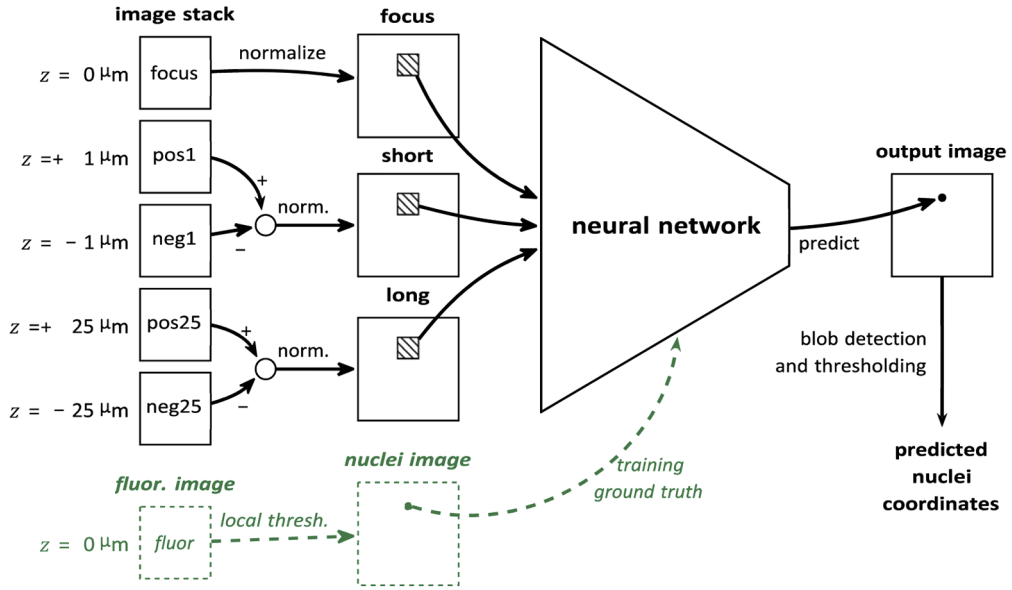

Fig. 2. Overview of the imaging and processing pipeline; blocks used only during training are shown in green with dashed lines and slanted text.

It is not obvious what would be a neural network architecture that can directly output a variable number of nuclei locations from a set of input images. Such a neural network would have either a very large number of outputs, which would be inefficient, or it would have a variable number of outputs, which would not be straight-forward to implement.

• *abstract intermediate representation*

Another alternative would be to use a more abstract intermediate representation, for example filled circles or Gaussian profiles centered on nuclei as opposed to a binary image of nuclei. However, we find that this conflicts with the principle of CNNs: we want the transformation from focal stack to nuclei image to be as spatially invariant in (x, y) as possible to facilitate training and reduce the complexity of the network, but an arbitrary geometric shape would imply a much more complicated relationship between output pixel value and input patch, leading to the network having to learn to recognize entire cells as opposed to small fragments of a cell.

In the rest of this section, we will explain the design of each step in our proposed pipeline.

2.1. Images with implicit phase information

We designed our cell counting pipeline for specimens imaged through a 10 × /0.4 NA (numerical aperture) objective onto a camera with 2048 × 2048 pixels of size 6.5 µm behind a 510 nm to 550 nm bandpass filter, as illustrated in Fig. 3. The exact equipment for capturing these images will be described in Section 3. With this magnification, nucleus-stained fluorescent images of MSCs yielded nuclei with diameters ranging from roughly 20 pixels to 40 pixels (13 µm to 26 µm).

From the literature on contrast transfer functions and multi-distance transport-of-intensity equation imaging [23,24,26,27], images taken at small amounts of defocus yield trustworthy information about high-frequency components of the phase profile but are noise-dominated for low-frequency components. On the other hand, images taken at high amounts of defocus yield better information about low frequency components but are inaccurate for high frequency components, in part due to spatial and temporal incoherence in the light source. Therefore, for a more complete picture, we choose to take two pairs of images, one pair with small defocus on

Biomedical Optics EXPRESS

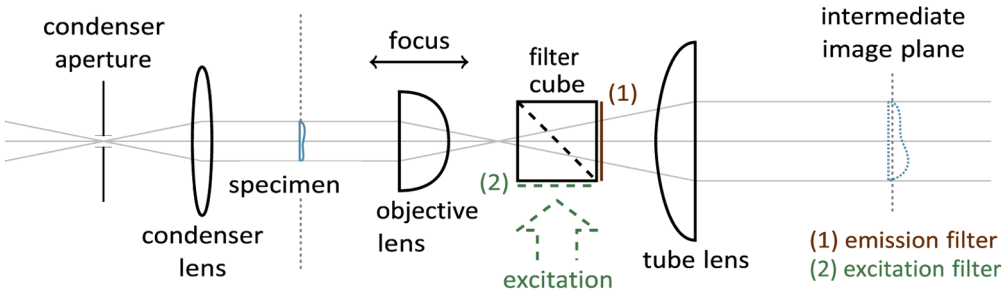

Fig. 3. Simplified optical diagram of a standard light microscope. For phase imaging, the light source is approximately collimated by the condenser aperture and illuminates the specimen; a camera sensor (not shown) located at the intermediate image plane captures an image of the specimen after the light has been filtered by a bandpass emission filter. For fluorescent imaging, the filter cube's dichroic (dashed diagonal) mirror allows excitation light to excite the fluorescent label, which in turn generates longer bandwidth emission light that is unreflected by the dichroic mirror and passes through to the imaging sensor.

either side of the focal plane, and one pair with high defocus. There is no need to capture a dense focal stack with evenly spaced images either; doing so would only increase redundancy at the cost of increased hardware complexity and acquisition time.

Given a nominal wavelength of 530 nm, the depth of field (which is also twice the minimum defocus distance to achieve maximum contrast for the highest frequency phase features) comes to ~3.3 µm. For simplicity, we chose a defocus of ±1 µm for capturing high frequency spatial phase information. Empirically, we found that a defocus of ±25 µm yielded acceptable low frequency information about nucleus structure; nucleoli were still visible in the difference image, and ±25 µm was the minimum defocus distance for maximum contrast on periodic phase features with 6.6 µm periods, which roughly corresponded to twice the observed nucleoli size. Higher defocus resulted in lower contrast due to incoherence and appearance of artifacts due to non-uniformity of the illumination. In addition to the two defocus pairs of images, we also image the sample at the focal plane; this is to both calibrate for the illumination intensity and gain in the camera as well as allow the neural network to learn about and remove unwanted elements such as dust and camera non-uniformities.

The literature also suggests an imaging scenario where the illumination light is *mostly* coherent—both temporal and spatial incoherence reduces contrast of the obtained images [26]. This is achieved in our design by the bandpass filter to increase temporal coherence as well as the stopped-down aperture to increase spatial coherence.

After collecting the images, we scale all the images in each set by a constant scale factor so that the in-focus image has an average pixel value of 1.0; this increases robustness to camera gain and illumination variations as well as eliminate the need for the neural network to also learn that absolute light levels are irrelevant when extracting phase imaging. We also use only the difference of the two pairs of defocus images; this eliminates some non-uniformities and allows us to provide *a priori* information about the imaging system, namely that we are primarily interested in the phase profile. Therefore, each set of images is reduced from five images to three images.

2.2. Neural network classifier

Since our goal is to predict whether the center pixel of a patch is inside the nucleus or not, we employed a binary classifier design template for our neural network, illustrated in Fig. 4.

Using this template, we tested ten different network designs, summarized in Table 1. In all of the designs, the input to the neural network is a 3-channel 61 × 61 pixel image patch, with

Biomedical Optics EXPRESS

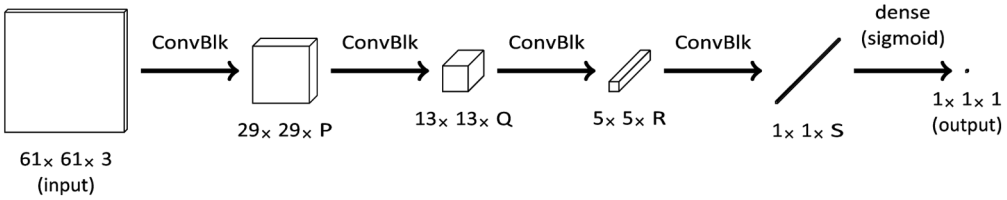

Fig. 4. Diagram of the neural network architecture. Two different designs for ConvBlk and five different sets of depth parameters (P,Q,R,S) are summarized in Table 1.

Table 1. The ten different network designs tested.[a]

name	ConvBlk design	activation	P	Q	R	S	weights
A4	one 5 × 5 convolutional layer	ReLU	4	8	16	32	17193
A6	one 5 × 5 convolutional layer	ReLU	6	12	24	48	38389
A8	one 5 × 5 convolutional layer	ReLU	8	16	32	64	67985
A10	one 5 × 5 convolutional layer	ReLU	10	20	40	80	105981
A12	one 5 × 5 convolutional layer	ReLU	12	24	48	96	153377
B4	two 3 × 3 convolutional layers	ReLU	4	8	16	32	18549
B6	two 3 × 3 convolutional layers	ReLU	6	12	24	48	41539
B8	two 3 × 3 convolutional layers	ReLU	8	16	32	64	73673
B10	two 3 × 3 convolutional layers	ReLU	10	20	40	80	114951
B12	two 3 × 3 convolutional layers	ReLU	12	24	48	96	165373

[a]The last column indicates the total number of trainable weights (*i.e.*, filter coefficients, dense layer weights and bias terms).

the channels corresponding to the scaled in-focus image, the scaled difference between the two ±1 μm defocused images and the scaled difference between the two ±25 μm defocused images. Given that nuclei were roughly 20–40 pixels in diameter, we chose a patch size of 61 pixels so that parts of the surrounding cell body could be used in predicting whether the center of the patch was inside the nucleus. The 61-pixel patch size also works well with the design of the neural network resulting in no need for padding in the convolutional layers. This $61 \times 61 \times 3$ patch is fed into a series of four processing stages—each stage starts with a convolutional block (ConvBlk) consisting of one or more unpadded convolutional layers using rectified linear unit (ReLU) activation functions, with the final convolutional layer in each block having a stride of 2 to implement downsampling. We elected to use striding instead of pooling to simplify the neural network and reduce computation time, following the ideas in [28]. The width and height of the output after each processing stage were designed to be compatible with unpadded convolution, and the depth increases two-fold after subsequent processing stages—this results in an initial expansion in the amount of data passing through the network, followed by gradual reduction for subsequent stages. The output of the last processing stage is fed into a dense layer with a sigmoidal activation function to yield a single value that should be 1 if the center of the patch is inside the nucleus and 0 otherwise.

As is typical of binary classification problems, we used a standard cross-entropy loss function,

$$\text{loss} = \frac{1}{N}\sum_{n=1}^{N} -l_n \log(p_n + 10^{-7}) - (1 - l_n)\log(1 - p_n + 10^{-7}), \tag{1}$$

where l_n and p_n are the ground-truth label and the neural network's predicted value, respectively, for the n^{th} example, and N is the total number of examples presented during training. The 10^{-7} factor to prevent taking the logarithm of zero was the standard value provided by the TensorFlow

Biomedical Optics EXPRESS

framework. We note that the neural network can only predict results for pixels in a region that is shorter by 30 pixels on each side compared to the full image, since we only consider full 61×61 pixel patches. Therefore, an input image set consisting of 2048×2048 pixel images can only predict the center 1988×1988 region.

For prediction and evaluation, *i.e.*, when we wish to output entire images, we make a modification to the neural network to improve computation efficiency. We start with the fact that a convolutional layer with stride s and dilation d feeding into a second convolution layer is the same as a convolutional layer with stride 1 and dilation d feeding into a second convolution layer with dilation sd. By converting to dilated convolutional layers, we can expand the input and outputs to have blocks of adjacent patches processed in parallel. For example, by using a 200×200 input patch, we can simultaneously predict the output in a 140×140 block. Block-wise computation allows us to reuse previously computed convolutions and is one of the reasons why convolutional neural networks are so efficient [29]. We do not use this approach in training as we find that it causes the optimization algorithm to get stuck in a suboptimal local minimum, possibly due to neighboring examples being too highly correlated with each other. In summary, the predicted output image from the neural network is computed from several large non-overlapping output blocks with corresponding large overlapping input patches; input-output pairs of the same size are batched together.

2.3. Blob detection and sieving

Given an output image from the neural network, blob detection followed by sieving of unlikely candidates is used to generate a list of detected nuclei; this process is summarized in Fig. 5. Blob detection starts with a difference-of-Gaussians (DoG) approach, similar in concept to the one used in the popular SIFT descriptor [30] but using a scale-space that is sampled much more finely (seven levels across scale) for detecting blobs of diameters between 21.2 and 42.4 pixels. The coordinates of local minima in the difference-of-Gaussians representation are used to compute an x-y location as well as a blob size. We fit each blob to an ellipse by analyzing a circular patch of pixels centered on the x-y location—we treat the patch as a two-dimensional probability mass function, and we compute the mean and diagonalize the variance of this virtual random two-vector in order to obtain the centroid, major and minor axes of an ellipse. Overlapping ellipses (detected using bounding boxes and then an image-based overlap method, grouped using an adjacency matrix) are merged by fitting an ellipse to the union of the overlapping ellipses. Ellipses whose areas are too small (less than 50 square pixels, or roughly a circle with 8 pixel diameter) or too large (more than 2000 square pixels, or roughly a circle with 50 pixel diameter) are removed. A score for each ellipse is computed by summing the pixels inside the ellipse and dividing by the ellipse area. Finally, only ellipses with a score higher than a user-set threshold are

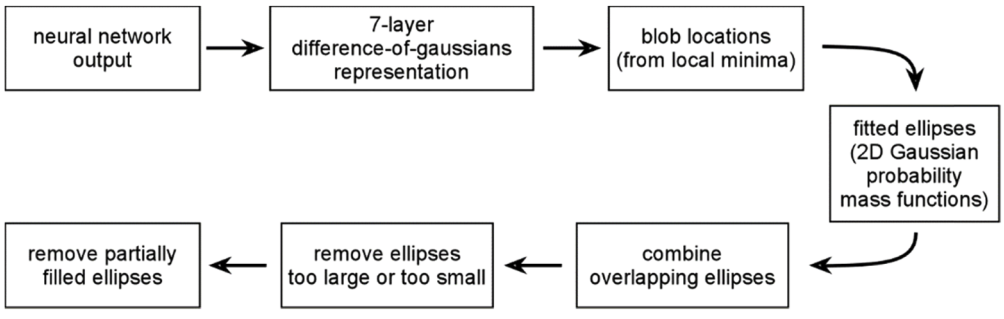

Fig. 5. Overview of the blob detection and sieving process used to locate nuclei using the output of the neural network.

Biomedical Optics EXPRESS

kept. This threshold is usually determined empirically from validation data and has an impact on the accuracy of our approach.

3. Methods

3.1. Sample preparation

Human MSCs (hMSCs) obtained commercially (Lonza) were cultured in Dulbecco's modified Eagle medium (DMEM) with low glucose and pyruvate (Thermo Fisher 11885076) supplemented with 10% (v/v) of fetal bovine serum (FBS) (Thermo Fisher 26140079) and 1% (v/v) of Penicillin-Streptomycin (10,000 U/mL) (Thermo Fisher 15140122). The hMSCs were seeded in six-well glass bottom plates (Cellvis P06-1.5H-N) prior to imaging. Culture media was changed every 2–3 days and cells were passaged at 80% confluency. All the cells were grown in a humidified atmosphere with 5% carbon dioxide and a constant temperature of 37°C.

A paraformaldehyde solution, 4% in phosphate-buffered saline (PBS), (Santa Cruz sc-281692) was used to fix the hMSCs. The cells were stained using Hoechst 33342 (Sigma Aldrich B2261-25MG), via a 1:750 dilution in basal DMEM of a 10mgmL^{-1} stock solution.

3.2. Imaging

Imaging was performed using a motorized inverted fluorescence microscope (Olympus IX83) fitted with an sCMOS camera (Hamamatsu ORCA-Flash 4.0 V3 digital) with 2048 × 2048 pixels of size 6.5 µm. MetaMorph (Intelligent Imaging Innovation Inc.) was used to automate the image capture process, and all images were captured using an Olympus UPLSAPO 10X2 objective (10×/0.4 NA). Each captured *image set* consists of a stack of five bright-field images under collimated illumination and a corresponding fluorescent image.

The bright-field images were taken at focus as well as at positions ±1 µm and ±25 µm away from focus. Illumination was with a U-LH100-3 100W halogen lamp, and the microscope was set up in Köhler illumination, but with the aperture stopped closed to the minimum possible to produce a roughly collimated illumination beam. A U-FBNA mirror cube, typically used for imaging green fluorescent protein (GFP), was inserted in the imaging path so that its emission filter centered at 530 nm can keep the image roughly monochromatic. The combination of a stopped-down aperture and emission filter was used to ensure high contrast, as was mentioned in Section 2.1. Fluorescence images were captured under excitation by the X-Cite XLED1 UVA module (360 nm to 380 nm) and a U-FUW mirror cube, appropriate for the Hoechst dye, was used to filter both excitation and emission light.

With ease of application in mind, we kept the imaging apparatus free from technicalities difficult to reproduce in non-specialized laboratories. We wrote a MetaMorph journal (a program which consists of a sequence of instructions for the microscope) to automate the capture of each image set. The journal starts with capturing the fluorescence image at the focal plane (adjusted according to the bright-field imaging set-up) with an exposure time of 100 ms. Next, the illumination, aperture and filter cubes are swapped for collimated bright-field imaging. The stack of collimated bright-field images was taken in the order of: 25 µm, 1 µm, −1 µm, −25 µm, and finally back to the in-focus plane, each at an exposure time of 150 ms. A separate journal was utilized to control the position of the stage and automate its movement. The following datasets were captured, with each dataset consisting of multiple image sets.

• *MSC image datasets (MID) 1*

MSCs were seeded at a density of 5000cm^{-2} into four out of six wells in a six-well glass bottom plate. The growth media was changed after one day to remove unattached cells, and the cells were allowed to culture for an additional two days before being fixed and stained for imaging.

Biomedical Optics EXPRESS

A total of 78 image sets were taken for the first two wells, referred to as MID1a and MID1b, respectively, and 80 image sets were taken for the third well, referred to as MID1c. The fourth well was not imaged. The stage was translated to different positions with non-overlapping fields-of-view for each of the image sets.

• *MSC image datasets (MID) 2*

MSCs were seeded at a density of 1500cm^{-2} into all six wells of a six-well glass bottom plate. The imaging procedure was slightly more complicated, as we wished to also image cells before being fixed or being stained while trying to minimize movement of the cells in the interim; we found that 10 to 15 image sets per well worked best. For this dataset, three wells (resulting in MID2a, MID2b and MID2c) were imaged using the following procedure for each well, with all imaging for one well finished before proceeding to subsequent wells.

For each well, we first captured image sets of the live cells at 10 different locations (datasets labeled "live"). The well plate was then removed from the microscope to fix the cells (and new growth medium changed) before placing back inside the microscope. We then captured image sets of the fixed cells at the same 10 locations (datasets labeled "fixed"). The well plate was removed again to add fluorescent dye before placing back inside the microscope. Finally, we captured image sets of the stained cells at the same 10 locations again (datasets labeled "stained"). MetaMorph and a Prior XY translation stage were used to memorize the various positions within the wells so that the positions were repeatable. A manual realignment was performed before imaging the fixed and/or stained images so that small shifts in the well plates were roughly adjusted for. Any remaining shifts were registered using software in the processing pipeline.

• *MSC image dataset (MID) 3*

The remaining three unimaged wells in the six-well glass bottom plate for MID2 were allowed to culture for two days before being imaged, resulting in MID3a, MID3b and MID3c. The imaging procedure was roughly the same, also generating live, fixed and stained datasets. Cells in MID3 showed lower cell density (number of cells per unit of area). Hence, 15 positions per well were imaged to compensate. After imaging, it was found that image sets for one of the 15 positions in MID3a was erroneous, and therefore all the images for that position were omitted.

The collected datasets are summarized in Table 2. Cell counts were verified manually. They are not available for MID1a, since it was only used for training the neural network. However, cell density in all MID1 samples was comparable.

3.3. Training

The same 78 image sets from MID1a were used to train ten different neural networks, each corresponding to one of the ten designs given in Section 2.2. From the fluorescence image in each image set, we generated a "ground truth" image specifying which pixels belonged to nuclei according to the method described in Section 3.3.1. Augmentation, described in Section 3.3.2 below, was also used to increase the training set and robustness, resulting in a total of roughly 2.5 billion binary classification examples being presented to the neural network during training. All neural networks were trained for a total of one epoch, with a snapshot of each neural network saved after every subepoch.

We shuffled the input examples by starting at a randomly picked image index for the first example to be fed into the network, and each example we subsequently pick is from an index that is the previous index plus a number M modulus N, where N is the total number of examples and $M \geq N / 2$ is a number coprime with N. This simple linear congruential generator ensures that any contiguous sequence of N indexes has only unique indexes. We used this shuffling approach to make sure adjacent examples were sufficiently uncorrelated with each other (unshuffled neighbors were highly correlated due to the sliding window approach used to generate the examples) while not needing to keep a large array of randomly shuffled indexes. The shuffled examples were then assembled into batches before feeding into the neural network during the training process.

Biomedical Optics EXPRESS

Table 2. Summary of collected image datasets.

Sample name	Cell count		Dataset label	Number of images	Fixed	Stained
	Total	Min - Max				
MID1a	Not counted	-	-	78	Yes	Yes
MID1b	12948	112 - 238	-	78	Yes	Yes
MID1c	13149	105 - 240	-	80	Yes	Yes
MID2a	870	59 -103	live	10	No	No
			fixed	10	Yes	No
			stained	10	Yes	Yes
MID2b	930	73–127	live	10	No	No
			fixed	10	Yes	No
			stained	10	Yes	Yes
MID2c	822	58 - 117	live	10	No	No
			fixed	10	Yes	No
			stained	10	Yes	Yes
MID3a	574	25 - 53	live	14	No	No
			fixed	14	Yes	No
			stained	14	Yes	Yes
MID3b	599	26 - 57	live	15	No	No
			fixed	15	Yes	No
			stained	15	Yes	Yes
MID3c	593	32 - 65	live	15	No	No
			fixed	15	Yes	No
			stained	15	Yes	Yes

To iteratively reduce the loss function, we used the Adam optimization algorithm [31] with a learning rate of 8×10^{-3}, $\epsilon = 0.1$, and standard parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The higher ϵ value was used to prevent the optimization algorithm from catastrophic failure when the estimated second moment of the gradient could sometimes approach zero due to an unlucky sequence of inputs. Instead of decaying the learning rate, we opted to periodically increase the batch size by a factor of two every two subepochs (starting with a size of 256 and ending with 2048) in order to achieve the benefits of simulated annealing while improving parallelism for training efficiency, as was demonstrated in [32].

All neural networks were implemented using the TensorFlow framework in Python, and code was executed on a workstation equipped with a 6-core Intel Xeon processor running at 2.6 GHz, 64 gigabytes of memory, and an NVIDIA GeForce GTX-1060 graphics processor that included 6 gigabytes of onboard memory. Custom code was used to generate the relevant training statistics as well as subepoch snapshots.

3.3.1. Nuclei images

For training, we generated "ground-truth" images that marked which pixels in an image set corresponded to nuclei. Since the Hoechst-stained images yielded good contrast, we employed a simple local thresholding approach [33,34] to binarize the image, with bright pixels belonging to nuclei. The thresholding algorithm looks at a 39×39 pixel window, finds the minimum and maximum values, and marks the center pixel as 1 if and only if (a) its value exceeds the average of the minimum and maximum, and (b) the difference between maximum and minimum is greater than a threshold, which we set to 10% of the image's maximum pixel value in order to

Biomedical Optics EXPRESS

ignore noise in regions of the image with no nuclei. The window size of 39 pixels was set to be approximately the maximum size of a nucleus so that for each pixel inside the nucleus, its window contained pixels both inside and outside the nucleus. Setting the window size too large runs the risk of having brightly labeled nuclei suppress nearby nuclei that are not as brightly labeled, and setting the window size too small can result in nuclei in the binarized image having holes in the middle. The automatic generation of nuclei images is not perfect, but it is scalable and the effects of erroneous binarization can be ameliorated when the neural network is fed enough examples.

3.3.2. Augmentation

Sometimes it may be difficult to produce that many images for training, so we also employ data augmentation techniques to increase the training set. Namely, we expand the data set eight-fold by allowing for 90-degree rotations and mirroring for each patch. Hence, a single image can yield 32 million training examples instead of just 4 million, and this augmentation helps prevent the neural network from overfitting and being rotation-sensitive. To clarify terminology, we say that we have trained for one *epoch* when all of the examples, including the augmented examples, have been presented to the neural network, and we have trained for one *subepoch* when the number of examples presented to the network is equal to the number of original (unaugmented) examples.

3.4. Verification

To verify the output of our pipeline, we need to compare the detected nuclei with nuclei manually labeled according to Section 3.4.1 from raw fluorescent images. As nuclei staining occurred after cell fixation, images from live cells from MID2 and MID3 were paired with their corresponding fluorescent image from the dataset "stained" to generate the manual labels. Pairing required automatic offset alignment implemented according to Section 3.4.2. The two sets of nuclei positions are then matched using an algorithm described in Section 3.4.3. The resulting counts of true positives, false positives and false negatives were used to compute various statistics such as precision (fraction of predicted nuclei that are true nuclei), recall (fraction of true nuclei that were predicted), F_1 score (harmonic mean of precision and recall), and relative count (ratio of the number of predicted nuclei to the number of true nuclei).

3.4.1. Manually-labeled nuclei positions

We used a separate semi-automated procedure to generate ground truth nuclei positions from the Hoechst-stained images. The procedure started with first binarizing the Hoechst-stained image using moment-preserving thresholding [35], including ignoring pixel values below a certain value (a pixel value of 50 in an 8-bit image in our case) for noise rejection. Binarization was made intentionally different from the binarization used for training to avoid the introduction of bias. Binarization was followed by morphological erosion of 8 pixels and then centroid computation of the resulting contiguous blobs, performed automatically by ImageJ's "Find Maxima. . . " routine. Rough automatic nuclei position extraction was followed by manual adjustment of the positions including adding or deleting nuclei. As a graphical user interface was needed, both parts of the procedure were implemented in ImageJ.

3.4.2. Alignment

Since the specimen had to be physically removed from the microscope during the imaging for the MID2a,b,c and MID3a,b,c data sets, the resulting image sets had to be registered—we needed to know which pixels in the live cell and fixed cell images corresponded to a pixel in the stained cell image. Assuming rotations to be small, we only attempted to calibrate for translations. The procedure consisted of using block matching across the 25 µm defocused images corresponding to the two datasets we wished to register. We considered nine 256×256 pixel blocks evenly distributed in a 3×3 matrix in one image; for each block, we attempted to find a 256×256 block

Biomedical Optics EXPRESS

in the other image that maximized the Pearson correlation coefficient between them, examining all blocks shifted by at most 30 pixels horizontally or vertically. We then kept the median of the nine resulting (x, y) shift vectors. This shift vector was used when preparing the inputs to the neural network by shifting the 1988×1988 window of valid pixels in the output (nuclei) image by an amount determined by the shift vector.

3.4.3. Matching algorithm

Given two sets of nuclei positions, we employed the following technique for generating matches and determining how many true positives (predicted nuclei match ground truth), false positives (predicted nuclei with nuclei in the ground truth) and false negatives (nuclei in the ground truth but not predicted) were present. We first removed nuclei whose centers were within 60 pixels of the edge of the valid region in the predicted image, since blob detection could not be accurate with missing information along the margins. We then found the closest nuclei position in the second set for each nuclei position in the first set, and vice versa, retaining only mutual matches whose Cartesian distance was less than 30 pixels, which allowed for human error in the placement of nuclei centers as well as some movement of the cells between the live and fixed imaging phases in the MID2 and MID3 series. The resulting unmatched nuclei in the predicted set were considered to be false positives and the unmatched nuclei in the ground truth set were considered to be false negatives.

3.5. Network and threshold selection

To enable an educated choice on which network design and which training snapshot to use, a measure of fidelity for each snapshot was calculated as follows. For each snapshot of each network, output images were generated using the 78 image sets from MID1b. Blob detection and sieving were performed for 101 candidate threshold values evenly distributed between 0.00 and 1.00 in order to determine an optimal candidate.

To select a threshold, we could choose the candidate threshold value that maximizes the mean F_1 score across all 78 image sets. However, in practice, the mean F_1 score as a function of threshold tends to be a function that is rather flat yet noisy near the maximum, so we instead choose the threshold that brought the mean relative count closest to 1.0. The F_1 score at this threshold is what we will call the *fidelity* for that particular training snapshot and network design.

This threshold selection criterion is much more robust to noise and fluctuations than choosing the maximum F_1 score, simply because the mean relative count is always monotonic with respect to the threshold value (since the nuclei we keep are the ones that exceed the threshold), resulting in a clear indication of where the optimal candidate occurs. Furthermore, such a selection generally results in an F_1 score close to the maximum F_1 score in practice. Finally, choosing a threshold this way helps reduce systemic bias for counting and also roughly balances precision and recall, making sure neither one is too low compared to the other.

For each network design, we then chose the snapshot that had the highest fidelity, and we retained the individual precision, recall, F_1 scores as well as relative counts for that threshold and particular snapshot for further statistical analysis. A network design (and corresponding snapshot) was then chosen based on both fidelity (which indicates how well the network performs) as well as the number of trainable weights; we ideally want fewer weights and higher scores.

4. Results

4.1. Training

The total training time as well as the lowest loss function values for the augmented training data set (MID1a) and the validation data set (MID1b) are given in Table 3. Also listed are the highest fidelity attained across all possible snapshots. The pattern of total training times suggest that

Biomedical Optics EXPRESS

the preprocessing on the CPU for feeding the input data to the neural networks dominated the training time for simpler networks with fewer layers (the A series as well as B4 and B6), whereas actual training on the GPU started playing a larger role with the bigger networks, as seen by the appreciable increase in training time starting with B8.

Table 3. Training statistics for the ten network design candidates, including the lowest training loss and validation loss values across all checkpoints for each candidate as well as maximum attainable fidelity. The eventually chosen candidate, B8, is highlighted in bold.

network name	training time	min. loss (training)	min. loss (validation)	max. fidelity
A4	43.2h	0.0158	0.0160	0.962
A6	43.1h	0.0130	0.0133	0.972
A8	43.1h	0.0109	0.0115	0.978
A10	43.4h	0.0101	0.0104	0.981
A12	43.8h	0.00940	0.0103	0.980
B4	44.4h	0.0114	0.0131	0.968
B6	43.1h	0.00965	0.0105	0.977
B8	**54.5h**	**0.00857**	**0.00948**	**0.980**
B10	62.7h	0.00799	0.00881	0.982
B12	69.8h	0.00769	0.00867	0.983

4.2. Network and threshold selection

Ideally, we want a network with high fidelity with as few weights as possible (*i.e.*, more efficient and less prone to overfitting). As can be seen in Fig. 6, since both the A series and B series curves flatten after A8 and B8, respectively, it looks like an optimal choice may be A8 or B8.

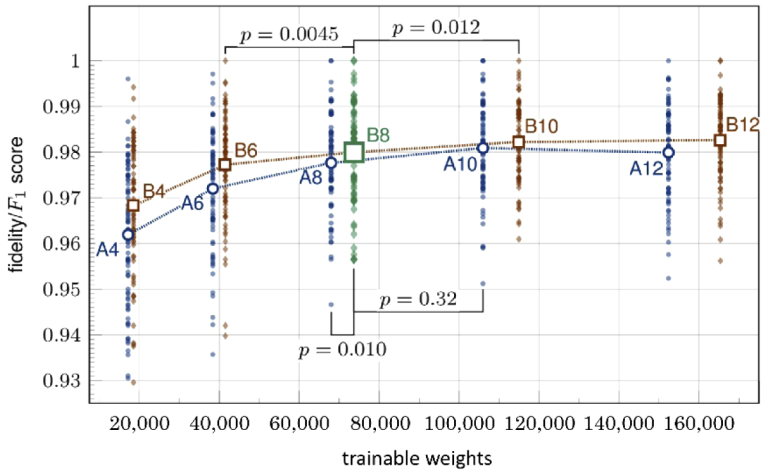

Fig. 6. Highest mean fidelity achieved for each network design shown as hollow markers: blue circles for the A series, red squares for the B series and a green square for the chosen B8 network. For each network, scatter plots denote fidelity for each image set in MID1b when using the snapshot and threshold that maximizes mean fidelity. Probability that two networks perform the same are given by p-scores computed using a paired t-test.

We chose B8 over A8 because B8 provided a statistically significant advantage (probability $p = 0.010$ of no advantage from a paired t-test) with only a small (8%) increase in number of

Biomedical Optics EXPRESS

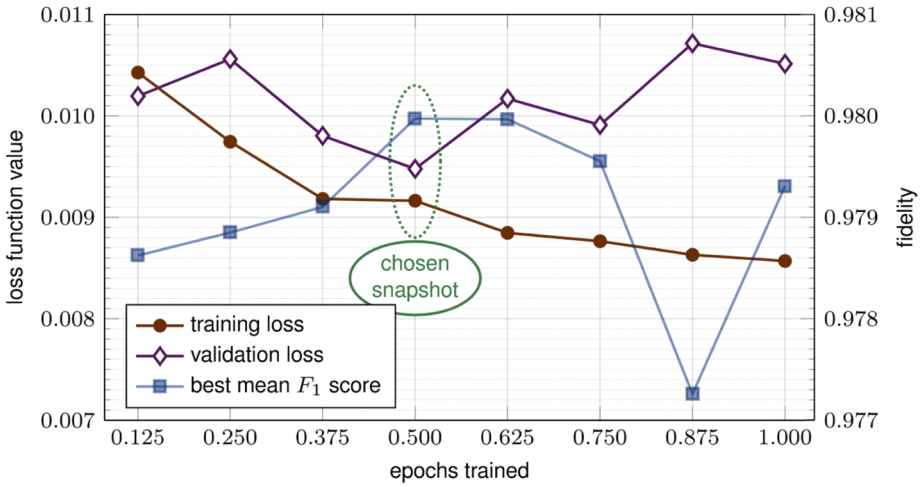

Fig. 7. Training loss (red filled circles), MID1b validation loss (violet open diamonds) and fidelity (blue filled squares) for each of network B8's eight checkpoints during training. The chosen checkpoint, circled in green dashes, has the largest attainable mean $F1$ score.

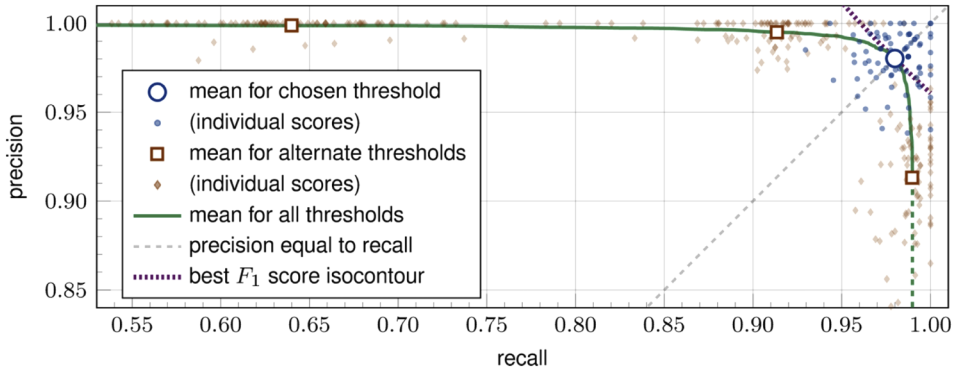

Fig. 8. Precision-recall curve for detected nuclei from validation dataset MID1b. The solid green curve gives the mean across all 78 images of the precision and recall for various possible thresholds in the blob-detection pipeline. For a subset of thresholds (from left, 0.90, 0.75, 0.00), hollow red squares indicate the mean precision and recall values, and corresponding small red diamonds indicate actual values across 78 images. The chosen threshold's (0.33) mean precision and recall values are illustrated using a blue hollow circle, with corresponding blue scatter points. A dashed light gray diagonal line indicates precision-recall values that result in no error in the overall cell count, and a hatched violet line on the top right indicates points where the $F1$ score is equal to the fidelity.

weights. Further paired t-tests show a significant improvement of B8 over B6 as well (p= 0.0045 chance of no advantage), while there is not a statistically significant difference between B8 and A10 ($p = 0.51$). Between B8 and B10 is a statistically significant improvement ($p = 0.10$), but we decided to use B8 because B10 was much more computationally expensive. We chose to use paired t-tests since all pipelines were using the same inputs to generate the validation data, and therefore any pair of results from the same input are correlated, making standard t-tests invalid.

In Fig. 7 we show for completeness the per-subepoch training set and validation set loss function values concurrent with the fidelity. For selection of snapshot, we used fidelity as opposed to the validation loss value, since the former quantifies the potential performance of the entire

Biomedical Optics EXPRESS

pipeline while the latter only quantifies the neural network part of the pipeline, which we are not directly interested in. However, in this particular case, the snapshot with the highest fidelity also corresponded to the snapshot with the lowest validation loss.

The resulting threshold corresponding to the chosen network design and snapshot was 0.33. For the chosen network and snapshot, a graph of mean precision versus mean recall (means taken over all 78 image sets in MID1b) is shown in Fig. 8 along with scatter plots of precision versus recall for individual image sets at various threshold values. The mean precision, recall and F1 score for the validation data at the chosen threshold were all equal to 0.980 within three significant digits.

4.3. B8 network for fixed and stained data from the same preparation

We first applied the trained network and processing pipeline to dataset MID1c to test the accuracy of our approach in ideal conditions (*i.e.*, cells prepared at the same time, with fixing and staining for accurate ground truth extraction). Using the chosen neural network snapshot and blob detection threshold, we computed precision, recall, F_1 score, relative count and predicted counts for the 80 image sets in dataset MID1c, with the results shown in Fig. 9 and Table 4.

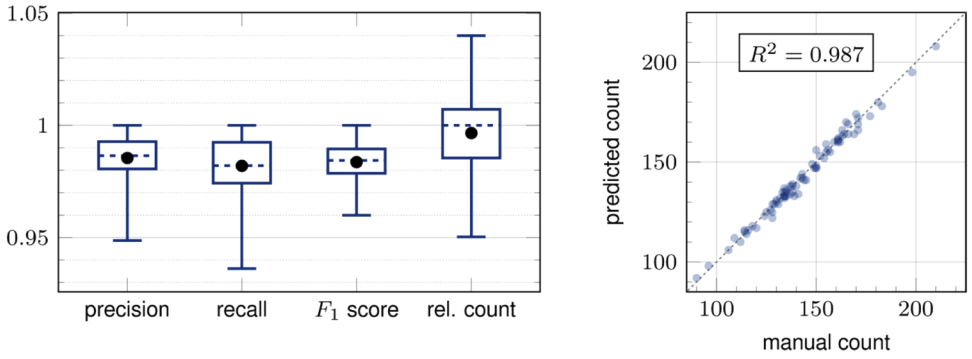

Fig. 9. On left, box plots for the precision, recall, $F1$ score and relative count. Top and bottom whiskers correspond to maximum and minimum values, respectively, boundaries of boxes give first and third quartiles, dashed line indicates median, and solid dot indicates mean. On right, scatter plot of predicted cell count per image versus actual cell count per image, with dashed line indicating an ideal perfect predictor.

Table 4. Minimum, first quartile, median, third quartile, maximum, mean and standard deviation for computed statistics from dataset MID1c.

	min.	Q1	median	Q3	max.	mean	std. dev.
precision	0.949	0.981	0.986	0.993	1.00	0.985	0.0104
recall	0.936	0.974	0.982	0.992	1.00	0.982	0.0132
$F1$ score	0.960	0.979	0.984	0.990	1.00	0.984	0.00810
rel. count	0.950	0.985	1.00	1.01	1.04	0.997	0.0178

From the results, it is apparent that performance on the test dataset MID1c is in agreement with the predicted performance from the validation dataset MID1b; MID1c's mean precision, recall and F_1 scores of 0.985, 0.982 and 0.984, respectively, match closely with the corresponding scores from MID1b. The mean relative count for MID1c is also very close to 1.00, which means that on average the cell count reported by our pipeline for MID1c should be unbiased. Although in extreme cases the counts differed by up to 5%, at least half of the predicted cell counts were within 1.5% of the actual cell counts.

Biomedical Optics EXPRESS

In addition to cell counting, we tested the reliability of our method to quantify morphological parameters such as area and eccentricity of the nuclei. The high correlation (Pearson Coeff. 0.8933; $p < 0.0001$) shown in the left plot of Fig. 10 demonstrates the quality of the match between our pipeline's prediction of the area of cell nuclei to estimates obtained using thresholded fluorescent images. The right graph shows an example of shape statistics our pipeline can generate to discriminate subpopulations of cells at different biological states [18].

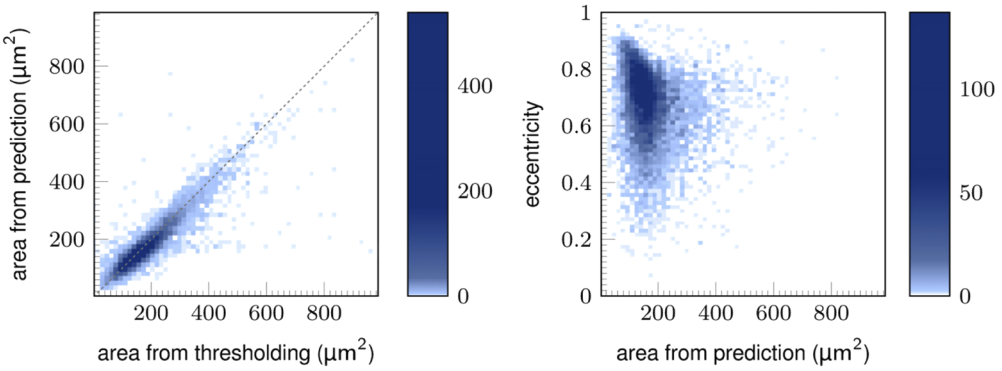

Fig. 10. On left, a two-dimensional histogram of predicted nuclei area from our method versus nuclei area estimated from fluorescence images; dashed line indicates perfect fit (Pearson Coeff. 0.8933; $p < 0.0001$). On right, a two-dimensional histogram illustrating the statistical relationship between eccentricity (how slim a nucleus is) versus area based on data gathered via our method.

An example of a challenging case where our method is able to correctly identify nuclei and count cells is shown in Fig. 11; in this 500×500 pixel crop of one of the image sets, there are optically thick regions that look very similar to cell nuclei and also very dense regions as well. Two circular blobs pointed to by orange arrows are correctly identified as *not* being nuclei whereas a similarly smooth circular region immediately below the two blobs *is* identified as a nucleus. The optically thick region at the center top has a non-uniform texture, but it is correctly identified as *not* a nucleus whereas another large and textured optically thick region on the right edge of the image *is* correctly identified. Furthermore, it is very difficult to visually examine and count the number of nuclei in the region inside the dashed green ellipse, but our approach correctly identifies all the nuclei and do not distort their relative shape or sizes. Another challenging example is shown in Fig. 12; in this case our method proved to be able to correctly recognize and distinguish the nuclei of partially overlapping cells (blue arrows), whose fluorescence images overlaps at low thresholding values. At present, we are unable to predict the performance of our method in case of nuclei with a high degree of overlap, which may occur at exceptionally high cell densities, as this scenario was not observed in any of our images.

While the overall fidelity was quite high at over 0.98, we show an example of an extreme failure case in Fig. 13. The false negatives result from the neural network not generating a bright enough blob to be recognized by the blob detection scheme, likely due to inability of the network to recognize that particular cell shape. The neural network also misidentified thick parts of a large cell body as other nuclei. Despite these problems, the neural network is able to identify tightly spaced nuclei for cells of various shapes, including nuclei that would be difficult to discern and/or separate with the human eye. The neural network also manages to ignore detritus, and for many cells it provides a good approximation to nuclei area and shape as well.

Biomedical Optics EXPRESS

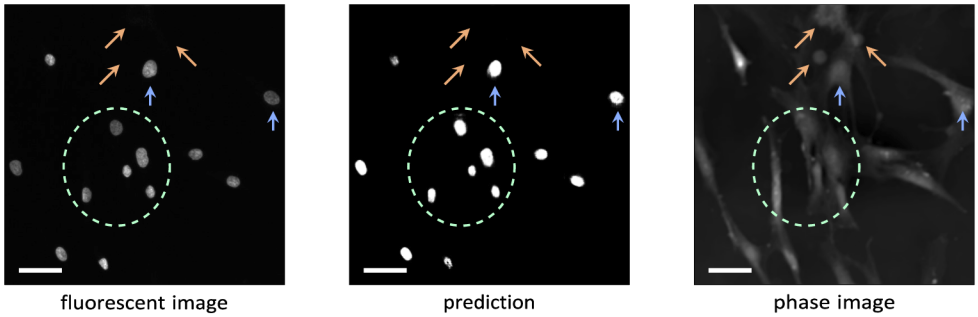

Fig. 11. True negatives (orange arrows pointing diagonally) and true positives (blue arrows pointing up) shown superimposed over the Hoechst-stained image (left), the neural network prediction output (middle), and a quantitative phase image (right) generated from the ±25 µm images using a standard TIE algorithm provided by [23]. A green dashed ellipse encloses a region containing a high concentration of cells. Scale bar is 50 µm.

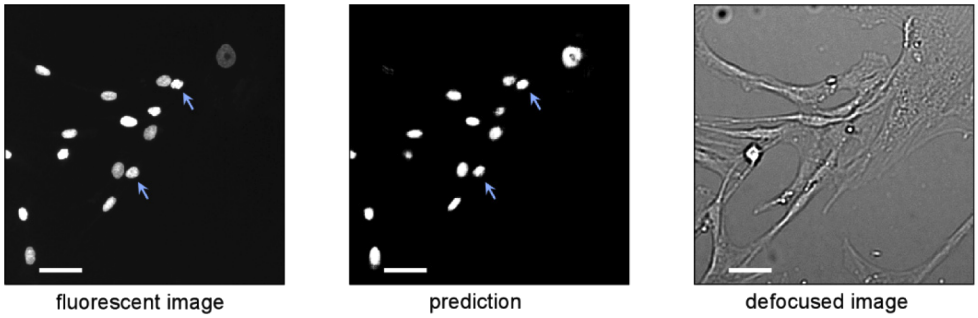

Fig. 12. Examples of nuclei in intimate contact (blue arrows) shown superimposed over the Hoechst-stained image (left), the neural network prediction output (middle), and a brightfield image (right) obtained at -25 µm defocus. Scale bar is 50 µm.

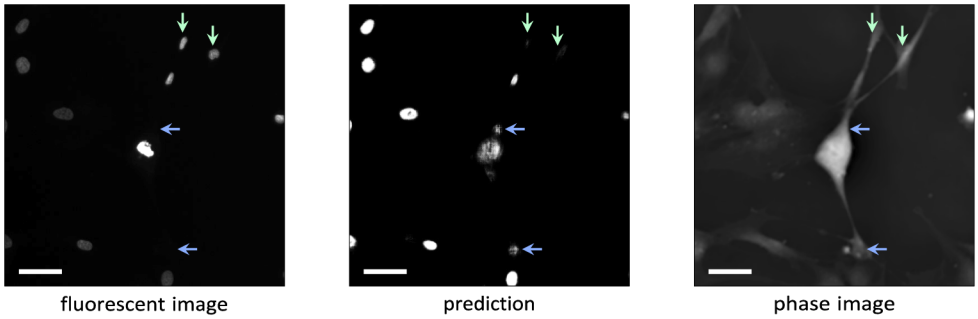

Fig. 13. An extreme failure case with false positives (blue arrows pointing left) and false negatives (green arrow pointing down) superimposed over the Hoechst-stained image (left), the neural network prediction output (middle), and a quantitative phase image (right) generated from the ±25 µm images using a standard TIE algorithm provided by [23]. Scale bar is 50 µm.

Biomedical Optics EXPRESS

4.4. B8 network for live or unstained data from a different preparation

For a more challenging scenario, we used our currently configured pipeline to count cells from the MID2a datasets, which imaged cells from a totally different preparation. The results are shown in Fig. 14. Overall, there is little difference between the stained and unstained (but fixed) results, and they are remarkably similar to the results obtained for cells from the same preparation as the training data, demonstrating the robustness of our approach when applied to independent datasets of similar nature. There is, however, a significant drop in precision for the live cell case, with approximately 15% overcounting. We also applied our currently configured pipeline to the MID2b dataset, showing qualitatively similar results, as in Fig. 15. This drop in precision is likely caused by changes in index of refraction or aspect ratios of the cells and nuclei due to chemical processes (protein cross-linking) used in the fixation process.

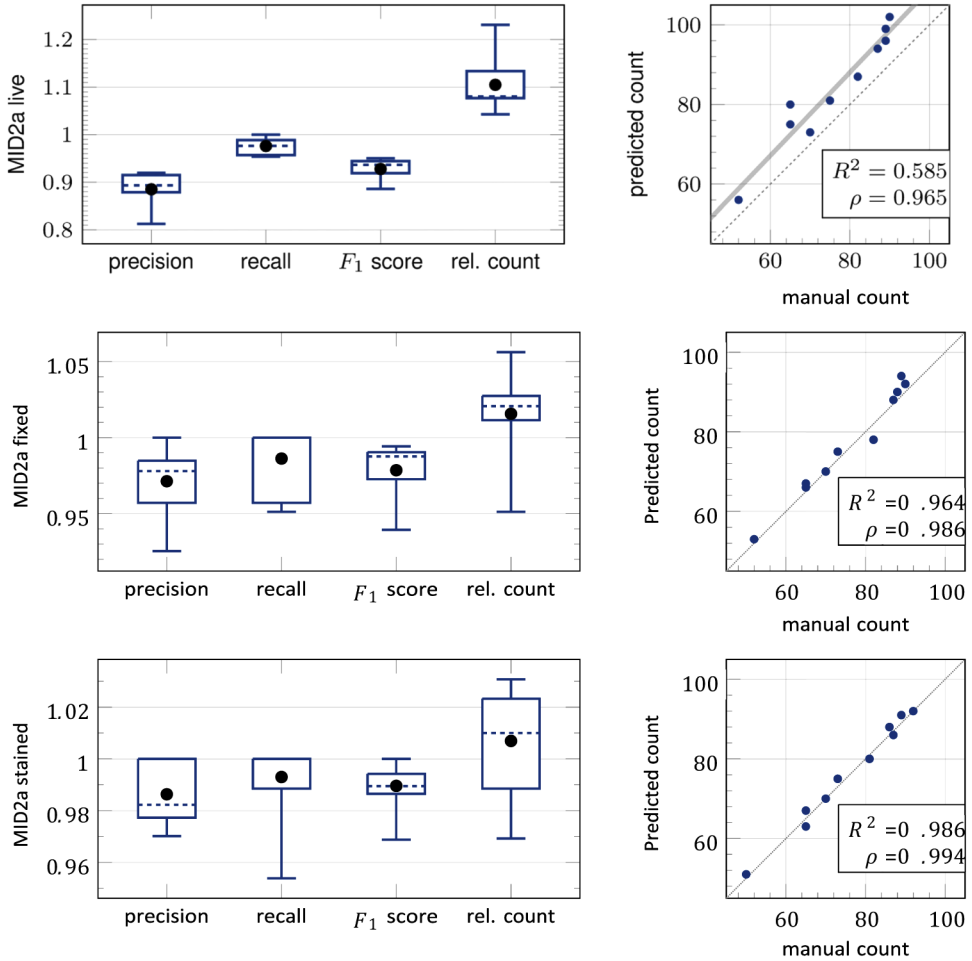

Fig. 14. On left, box plots of cell counting statistics generated from images of stained, fixed and live cells from MID2a; top and bottom whiskers indicate minimums and maximums, whereas boxes indicate quartiles. On right, a scatter plot of predicted cell count per image set from our pipeline versus manual counts from fluorescent images. The coefficient of determination for a direct mapping is given by R^2, while the correlation coefficient ρ gives the goodness of fit for a linear fit.

Biomedical Optics EXPRESS

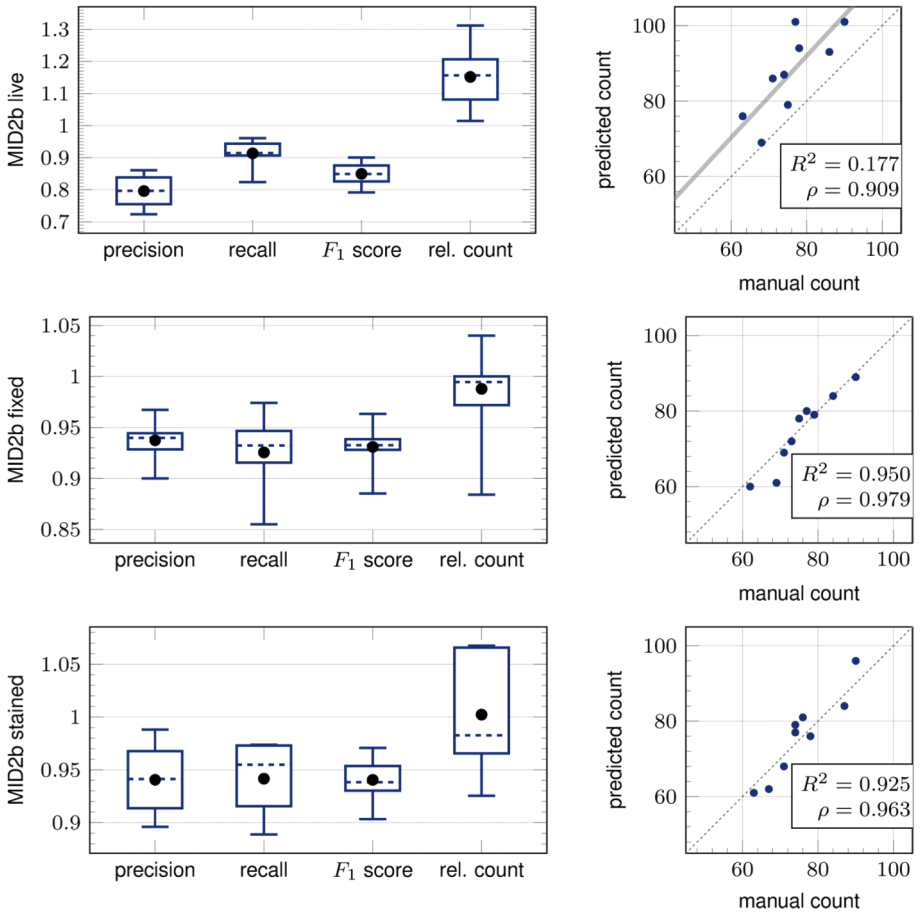

Fig. 15. On left, box plots of cell counting statistics generated from images of stained, fixed and live cells from MID2b; top and bottom whiskers indicate minimums and maximums, whereas boxes indicate quartiles. On right, a scatter plot of predicted cell count per image set from our pipeline versus manual counts from fluorescent images. The coefficient of determination for a direct mapping is given by R^2, while the correlation coefficient ρ gives the goodness of fit for a linear fit.

Biomedical Optics EXPRESS

4.5. Alternate B8 network

To overcome the above-mentioned problem as well as minimize the effects of sample-to-sample variability, we trained another neural network with the same B8 architecture using image data from MID2a and MID3a. A total of 72 image sets was used as training input (10 MID2a live images, 10 MID2a fixed, 10 MID2a stained, 14 MID3a live, 14 MID3a fixed and 14 MID3a stained). Each of these datasets was paired with the corresponding stained dataset (MID2a and MID3a stained) as desired output. Further, calibration was performed using the MID2b and MID3b data sets, with different thresholds for the live, fixed and stained stages. This alternate pipeline was then applied to the MID2c and MID3c data sets.

The resulting performance of this alternate pipeline is compared with the original pipeline in Fig. 16 as well in Fig. 17 versus Fig. 18. There is noticeable improvement using the alternate pipeline for MID2c, but performance is relatively unchanged for MID3c.

One possible factor that could lead to poor performance for MID3c is that the similarly-imaged MID3a training data set contains relatively fewer cells per field of view when compared to the MID2a training data set; in Fig. 18, the MID3c and MID2c datasets contain in each image set (field of view) on average approximately 30 and 70 cells, respectively. A smaller representation of cells in the MID3a training set would mean that there may be certain variations in cellular shape and structure that appear in MID3c but not in MID3a, leading to poorer performance.

Given these results, we hypothesize that performance would be best in cases where (1) training is performed using samples as optically similar to the samples to be measured, and (2) there are enough cells per training set to properly characterize all the possible variations in cellular shape, structure and orientation.

Biomedical Optics EXPRESS

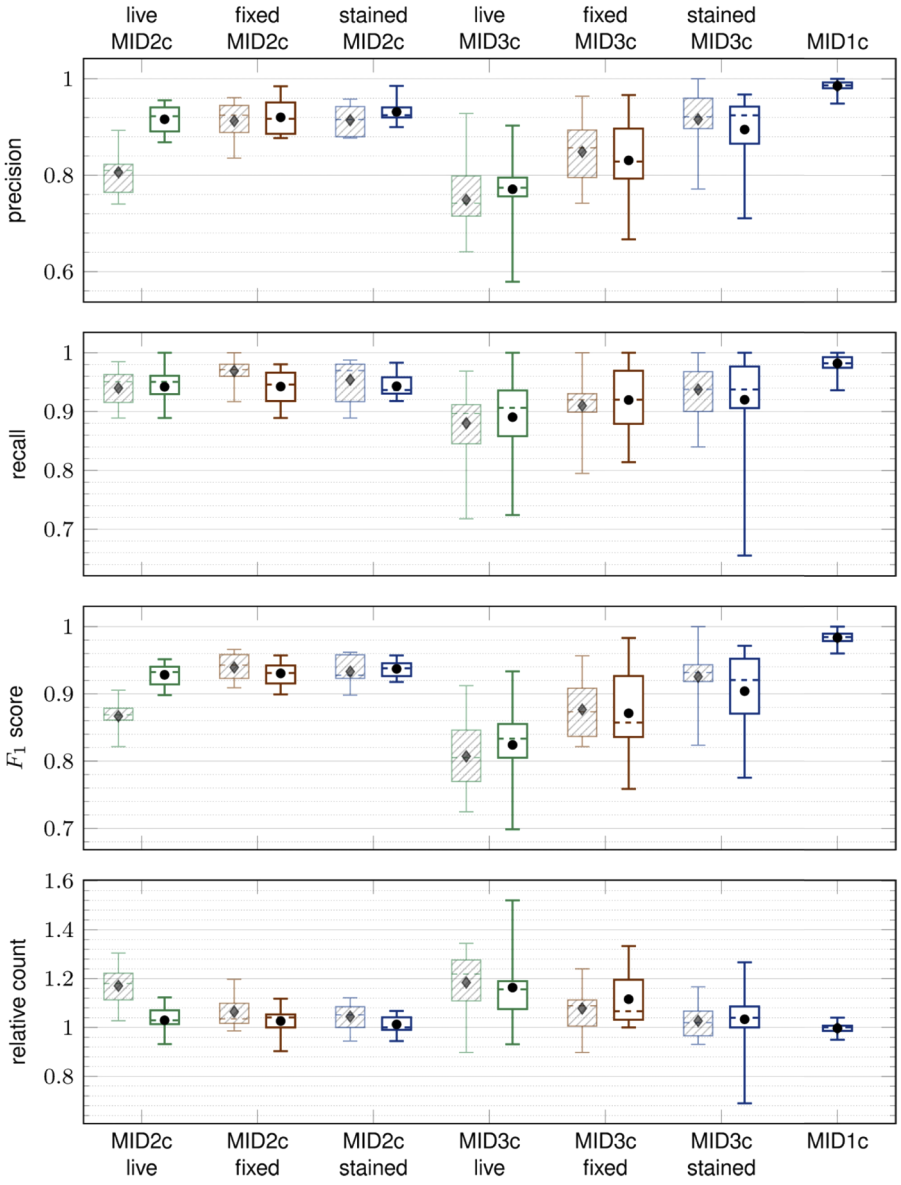

Fig. 16. Box plots of the resulting precision, recall, $F1$ score and relative counts for the live, fixed and stained images from the MID2c and MID3c datasets, along with results from MID1c for reference. For the MID2c and MID3c datasets, the faint and hatched box plot (with diamond marker indicating the mean) represents the result from the original B8 network, whereas the clear box plot (with circular marker indicating the mean) represents the result from the alternate B8 network trained with MID2a and MID3a data.

Biomedical Optics EXPRESS

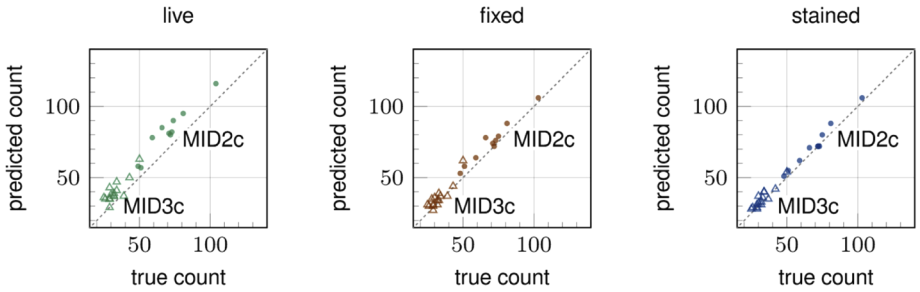

Fig. 17. Predicted versus actual cell counts shown as a scatter plot for all the image sets in MID2c (filled circles) and MID3c (hollow triangles). The B8 network was used.

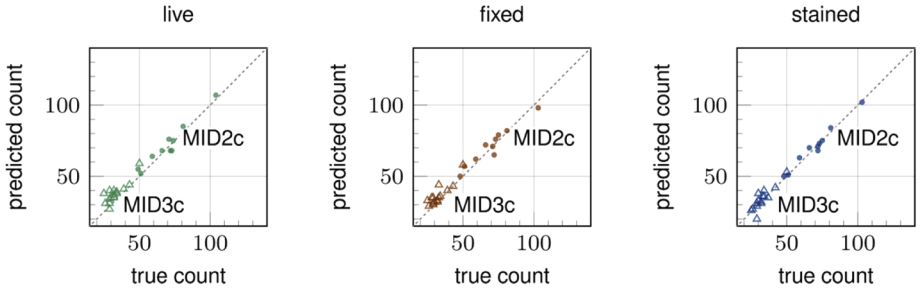

Fig. 18. Predicted versus actual cell counts shown as a scatter plot for all the image sets in MID2c (filled circles) and MID3c (hollow triangles). The alternate B8 network was used.

5. Discussion

Results with the MID1c dataset demonstrate that our learning-driven processing pipeline is capable of providing accurate nuclei localization and cell counts when applied to specimens prepared contemporaneously with the specimens used to train the neural network. Nuclei that would be difficult to identify with the human eye from quantitative phase images are readily discovered by the deep neural network. In an ideal setting (MID1c), our approach outperforms cell-counting methods based on holography visualization [4] (w.r.t. coefficient of determination) and cell confluency determination [5] (w.r.t. percent error in cell count).

Our neural network behaves very similarly in function to one of the trained networks from [17], although we optimized the defocus distances and reduced the number of input images due to *a priori* knowledge that information is present in the optical thickness profile of the specimen. Furthermore, since our goal is not to produce synthesized fluorescence images for human viewing, but rather as input to an automated framework, we are able to simplify the design of our neural network and preprocessing of training data, leading to more efficient computation.

Since our method generates statistics about the shape of the nuclei, we can apply our approach to biological studies in order to investigate how these statistics can be used to predict pluripotency, so that cells in culture can be noninvasively surveyed for viability. For example, it was shown in [18] that nucleus size can be used to predict pluripotency.

Results for the fixed but not stained cells in MID2a show that the neural network is indeed extracting implicit phase information from the images and not somehow picking up out-of-band information from the nuclei stain. A comparison of results between fixed cells and live cells show that there does exist systematic prediction error for live cell specimens, most likely due to the fixing process changing the phase profile of the specimen; it was observed that chemical fixation using paraformaldehyde induces shrinkage of the cells (due to protein crosslinking) and may sometimes

Biomedical Optics EXPRESS

cause the formation of bubble-like structures in the cells. While recalibration/retraining can be used to offset systematic prediction error for live cells, faster automated imaging and live staining can be a more practical solution to this problem, *i.e.*, for a cell culture under study, a subset of the cells is sampled for the training process. We think that this is a good potential avenue for future study based on the fundamentals we established in this paper.

The current design of the pipeline can be used for cell-counting applications when monitoring cell culture growth, if a specimen can be prepared contemporaneously and then fixed and stained for training after the experiment. Training times can be improved with higher end hardware, as we used the mid-end GeForce GTX 1060 GPU, whereas newer and higher-powered GPUs can provide much higher memory and computation throughput. Furthermore, we used a straight-forward blob-detection algorithm for the final stage in the pipeline, whereas a more sophisticated algorithm (*e.g.*, a neural network) may be able to achieve better accuracy. Finally, a transfer learning approach [36] where we keep the convolutional layers of our network but retrain the weights for the fully-connected layer can significantly decrease the training time required.

Funding. National Research Foundation Singapore (SMART-BioSyM IRG); National Agency for Research and Development ((ANID)/PIA/ACT192015); Pontificia Universidad Católica de Chile (VRI/Puente 004/2019 and 012/2020).

Acknowledgments. We would like to thank Vipra Guneta for providing the cell line and Ee Xien Ng for assistance with using the microscope and advice on cell culturing.

This research was supported by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its CREATE programme, Singapore-MIT Alliance for Research and Technology (SMART) BioSystems and Micromechanics (BioSyM) IRG. A. Ravasio acknowledge financial support by the National Agency for Research and Development (ANID)/PIA/ACT192015 and seed funds of the Pontificia Universidad Católica de Chile (VRI/Puente 004/2019 and 012/2020).

Disclosures. The authors do not have any conflict of interest to disclose.

References

1. A. I. Caplan and S. P. Bruder, "Mesenchymal stem cells: Building blocks for molecular medicine in the 21st century," Trends Mol. Med. **7**(6), 259–264 (2001).
2. J. Galipeau and L. Sensébé, "Mesenchymal Stromal Cells: Clinical Challenges and Therapeutic Opportunities," Cell Stem Cell **22**(6), 824–833 (2018).
3. Y. J. Lee, S. L. Vega, P. J. Patel, K. A. Aamer, P. V. Moghe, and M. T. Cicerone, "Quantitative, label-free characterization of stem cell differentiation at the single-cell level by broadband coherent anti-stokes raman scattering microscopy," Tissue Eng., Part C **20**(7), 562–569 (2014).
4. A. Mölder, M. Sebesta, M. Gustafsson, L. Gisselson, A. G. Wingren, and K. Alm, "Non-invasive, label-free cell counting and quantitative analysis of adherent cells using digital holography," J. Microsc. **232**(2), 240–247 (2008).
5. N. Jaccard, L. D. Griffin, A. Keser, R. J. Macown, A. Super, F. S. Veraitch, and N. Szita, "Automated method for the rapid and precise estimation of adherent cell culture characteristics from phase contrast microscopy images," Biotechnol. Bioeng. **111**(3), 504–517 (2014).
6. R. Flight, G. Landini, I. B. Styles, R. M. Shelton, M. R. Milward, and P. R. Cooper, "Automated noninvasive epithelial cell counting in phase contrast microscopy images with automated parameter selection," J. Microsc. **271**(3), 345–354 (2018).
7. B. Janicke, A. Kårsnäs, P. Egelberg, and K. Alm, "Label-free high temporal resolution assessment of cell proliferation using digital holographic microscopy," Cytometry **91**(5), 460–469 (2017).
8. C. Allier, T. Bordy, O. Cioni, L. Hervé, G. Esteban, M. Pisaneschi, and S. Morales, "Label-free cell viability assay using lens-free microscopy," in *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XVI*, D. L. Farkas, D. V. Nicolau, and R. C. Leif, eds. (SPIE, 2018), 10497, p. 10.
9. A. O. O. Odeleye, S. Castillo-Avila, M. Boon, H. Martin, and K. Coopman, "Development of an optical system for the non-invasive tracking of stem cell growth on microcarriers," Biotechnol. Bioeng. **114**(9), 2032–2042 (2017).
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM **60**(6), 84–90 (2017).
11. Y. Cui, G. Zhang, Z. Liu, Z. Xiong, and J. Hu, "A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images," Med. Biol. Eng. Comput. **57**(9), 2027–2043 (2019).
12. W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting and detection with fully convolutional regression networks," Comput. Methods Biomech. Biomed. Eng. Imaging Vis. **6**(3), 283–292 (2018).
13. I. Suleymanova, T. Balassa, S. Tripathi, C. Molnar, M. Saarma, Y. Sidorova, and P. Horvath, "A deep convolutional neural network approach for astrocyte detection," Sci. Rep. **8**(1), 12878 (2018).
14. C. X. Hernández, M. M. Sultan, and V. S. Pande, "Using deep learning for segmentation and counting within microscopy data," arXiv (2018).

Biomedical Optics EXPRESS

15. T. Vicar, J. Gumulec, J. Balvan, M. Hracho, and R. Kolar, "Label-free nuclear staining reconstruction in quantitative phase images using deep learning," in *IFMBE Proceedings* (Springer Verlag, 2019), 68(1), pp. 239–242.
16. Y. Rivenson, H. Wang, Z. Wei, Y. Zhang, H. Gunaydin, and A. Ozcan, "Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue," Nat. Biomed. Eng. **3**(6), 466–477 (2019).
17. E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl, L. L. Rubin, P. Nelson, and S. Finkbeiner, "In silico labeling: predicting fluorescent labels in unlabeled images," Cell **173**(3), 792–803.e19 (2018).
18. W. C. Lee, H. Shi, Z. Poon, L. M. Nyan, T. Kaushik, G. V. Shivashankar, J. K. Y. Chan, C. T. Lim, J. Han, and K. J. Van Vliet, "Multivariate biophysical markers predictive of mesenchymal stromal cell multipotency," Proc. Natl. Acad. Sci. U. S. A. **111**(42), E4409–E4418 (2014).
19. M. R. Teague, "Deterministic phase retrieval: a Green's function solution," J. Opt. Soc. Am. **73**(11), 1434–1441 (1983).
20. N. Streibl, "Phase imaging by the transport equation of intensity," Opt. Commun. **49**(1), 6–10 (1984).
21. A. Barty, K. A. Nugent, D. Paganin, and A. Roberts, "Quantitative optical phase microscopy," Opt. Lett. **23**(11), 817 (1998).
22. L. Waller, L. Tian, and G. Barbastathis, "Transport of intensity phase-amplitude imaging with higher order derivatives," Opt. Express **18**(12), 12552 (2010).
23. Z. Jingshan, R. A. Claus, J. Dauwels, L. Tian, and L. Waller, "Transport of Intensity phase imaging by intensity spectrum fitting of exponentially spaced defocus planes," Opt. Express **22**(9), 10661 (2014).
24. J. P. Guigay, "Fourier transform analysis of Fresnel diffraction patterns and in-line holograms," Optik (Stuttg) **49**(1), 121–125 (1977).
25. Y. I. Nesterets and T. E. Gureyev, "Partially coherent contrast-transfer-function approximation," J. Opt. Soc. Am. A **33**(4), 464 (2016).
26. J. C. Petruccelli, L. Tian, and G. Barbastathis, "The transport of intensity equation for optical path length recovery using partially coherent illumination," Opt. Express **21**(12), 14430 (2013).
27. Z. Zhang, W.-N. Li, A. Asundi, and G. Barbastathis, "Simultaneous measurement and reconstruction tailoring for quantitative phase imaging," Opt. Express **26**(25), 32532 (2018).
28. J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: the all convolutional net," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Work. Track Proc.* (2014).
29. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: integrated recognition, localization and detection using convolutional networks," *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.* (2013).
30. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis. **60**(2), 91–110 (2004).
31. D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (International Conference on Learning Representations, ICLR, 2015).
32. S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," arXiv (2017).
33. J. Bernsen, "Dynamic thresholding of gray level image," *Proc. 8th Int. Conf. Pattern Recognit.* 1251–1255 (1986).
34. M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," J. Electron. Imaging **13**(1), 146 (2004).
35. W. H. Tsai, "Moment-preserving thresholding: a new approach," Comput. Vision, Graph Image Process. **29**(3), 377–393 (1985).
36. S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010).