Analyzing U-Net Robustness for Single Cell Nucleus Segmentation from Phase Contrast Images

Chenyi Ling
Software and Systems Division
Information Technology Lab, NIST
Gaithersburg, MD 20899
chenyi.ling@nist.gov

Michael Majurski
Software and Systems Division
Information Technology Lab, NIST
Gaithersburg, MD 20899
michael.majurski@nist.gov

Michael Halter
Biosystems and Biomaterials Division
Material Measurement Lab, NIST
Gaithersburg, MD 20899
michael.halter@nist.gov

Jeffrey Stinson
Biosystems and Biomaterials Division
Material Measurement Lab, NIST
Gaithersburg, MD 20899
jeffrey.stinson@nist.gov

Anne Plant
Biosystems and Biomaterials Division
Material Measurement Lab, NIST
Gaithersburg, MD 20899
anne.plant@nist.gov

Joe Chalfoun
Software and Systems Division
Information Technology Lab, NIST
Gaithersburg, MD 20899
joe.chalfoun@nist.gov

Abstract

We quantify the robustness of the semantic segmentation model U-Net, applied to single cell nuclei detection, with respect to the following factors: (1) automated vs manual training annotations, (2) quantity of training data, and (3) microscope image focus. The difficulty of obtaining sufficient volumes of accurate manually annotated training data to create an accurate Convolutional Neural Networks (CNN) model is overcome by the temporary use of fluorescent labels to automate the creation of training datasets using traditional image processing algorithms. The accuracy measurement is computed with respect to manually annotated masks which were also created to evaluate the effectiveness of using automated training set generation via the fluorescent images. The metric to compute the accuracy is the false positive/negative rate of cell nuclei detection. The goal is to maximize the true positive rate while minimizing the false positive rate. We found that automated segmentation of fluorescently labeled nuclei provides viable training data without the need for manual segmentation. A training dataset size of four large stitched images with medium cell density was enough to reach a true positive rate above 88% *and a false positive rate below* 20%.

1. Introduction

Label-free single cell segmentation and tracking of human induced pluripotent stem cells (hiPSC) from widefield (2D) transmitted light microscopy images allows for monitoring dynamic cellular processes such as the growth and division of single cells under minimal light exposure conditions. However, segmentation of single cells in such images is very challenging because iPSCs are densely packed and it is difficult to distinguish cell edges in the resulting images.

Traditional computer vision segmentation techniques fail to meet the required single cell segmentation accuracy in the phase contrast image modality. Single cell segmentation is also complicated by the fact that even in a monolayer, cells in close proximity have a 3D component when their edges overlap. To make the problem more tractable, we chose to segment cell nuclei instead of whole cell bodies. There is commonly only a single nucleus per cell, and they overlap much less frequently than the full cell bodies, simplifying the segmentation task. However, segmenting cell nuclei via traditional computer vision algorithms from phase contrast images is also very difficult, especially since these cells occur in close proximity to one another within colonies.

To obviate the problem of segmenting nuclei from the phase contrast image modality, a nuclear envelope protein,

Lamin B1, modified with green fluorescent protein (GFP), is used to create fluorescent images which highlight the cell nuclei. While this approach is advantageous to nucleus segmentation, it could require fluorescence imaging for all future hiPSC cell dynamics analyses, something we would like to avoid since the excitation light for fluorescent imaging can cause phototoxicity in cells. In this paper we propose using fluorescence imaging to facilitate segmenting cell nuclei directly from phase contrast images of cells within colonies. The fluorescent protein indicates where the nuclei are in the phase contrast images, and then those phase contrast images are used to train the segmentation model. After the model is trained, future segmentations can be performed on phase contrast images, and fluorescence imaging is no longer needed. Manual segmentation of fluorescent nuclei by experts is used as a reference to compare the accuracy of the segmented phase contrast images. Since traditional segmentation methods still display major shortcomings in extracting nuclei masks directly from phase contrast images either via traditional pixel intensity-based or other traditional methods, we used convolutional neural networks (CNNs), specifically a U-Net [11] encoder-decoder architecture with skip connections.

Artificial Intelligence models have already been adopted for nucleus segmentation. CNN models have been widely used to directly segment cell nuclei from fluorescent images: Chidester et al [4], Khoshdeli et al [6], and Kumor et al [8] utilize CNN models for nuclear segmentation from histology slides. Kromp et al [7] demonstrate 3 CNN architectures (U-Net, U-Net with ResNet-34 backbone, and Mask-RCNN) for instance nuclear segmentation directly from fluorescent images. Xu et al [14] performs instance segmentation of nuclei from histopathology slides using a novel architecture. Multiple accuracy metrics have been used by Caicedo J. et al [1] to evaluate two neural networks (U-Net and DeepCell) for nucleus segmentation in fluorescence images. There have also been efforts to perform nuclear segmentation from non-fluorescent channels. Yuan P. et al [15] used RCNN to detect Car-T cell nuclei in brightfield images. Vuola et al [12] ensemble Mask-RCNN and U-Net to segment nuclei from multiple modalities. Piraud et al [10] perform nuclei segmentation from brightfield images. Xing F. et al. [13] uses a U-Net like model to test model robustness, whether a model trained on one microscope can be applied to other datasets acquired on different microscopes. These research endeavors are missing a sensitivity analysis study of the trained U-Net over the quality and quantity of training set with regards to nucleus segmentation.

This work quantifies the change in accuracy of the resulting U-Net model when the automated annotations are used to train the model as opposed to the domain expert annotations. The general workflow with sensitivity analysis is shown in Figure 1. Fluorescence images segmented with our network need to be post-processed to separate touching cell nuclei, a considerably more amenable task than single pixel semantic segmentation of phase contrast cell images. We accomplish this using a FogBank segmentation [2]. All the reference data for developing and testing the quantitative analysis pipeline was acquired with automated microscopy equipment, thus facilitating the collection of CNN training data.

Despite most modern microscopes having an auto-focus module which attempts to take the most in-focus image it can at each location, there are still differing levels of blur in the acquired live image acquisitions. The input images to the U-Net are going to unavoidably be slightly out of focus at times. We will analyze the impact out-of-focus images have on the U-Net accuracy at inference time. This also intersects with the questions about whether the Gaussian blur used in the data augmentation during the training mimics the image blur coming from the microscope.

2. Materials and Methods

This section describes the acquisition protocols and the segmentation workflow of training and inferencing using a U-Net model.

2.1. Image Acquisition Protocol

The human iPSC clonal line in which Lamin B1 has been endogenously tagged with mEGFP (LamB1:mEGFP) using CRISPR/Cas9 technology was generated at the Allen Institute for Cell Sciences (WTC-mEGFP-LMNB1-cl210), and was obtained from Coriell Institute for Medical Research (Catalog # AICS-0013, Camden, NJ). Cells were regularly maintained using complete mTeSR medium supplemented with Pen/Strep in six well tissue culture treated plates (TPP, Product # 92006, Switzerland) coated with Matrigel (hESC certified, from Corning). Generally, cells were passaged using Accutase when 70 to 80% confluent and re-plated at $100k$ to $200k$ cells per well.

Immediately prior to imaging the cell culture media was aspirated, wells were rinsed $1\times$ with DPBS and $2\,\mathrm{mL}$ of DPBS was pipetted into each well. The cell culture plate was placed on the microscope stage (Ludl Electronic Products, Hawthorne, NY) and maintained at $37\,^{\circ}\mathrm{C}$ in a custom-built incubation chamber (Kairos Instruments, Pittsburgh, PA). Image capture was performed on a Zeiss 200M microscope (Carl Zeiss USA, Thornwood, NY) using a Zeiss $10\times$, $0.3NA$ objective (Zeiss part number 420341-9911-000) and a CoolSNAP HQ2 CCD camera (Photometrics, Tucson, Arizona). Stage, filters and shutters were controlled with μManager[1] open source software. The stage was programmed to move from field to field with an over-

[1] https://micro-manager.org/

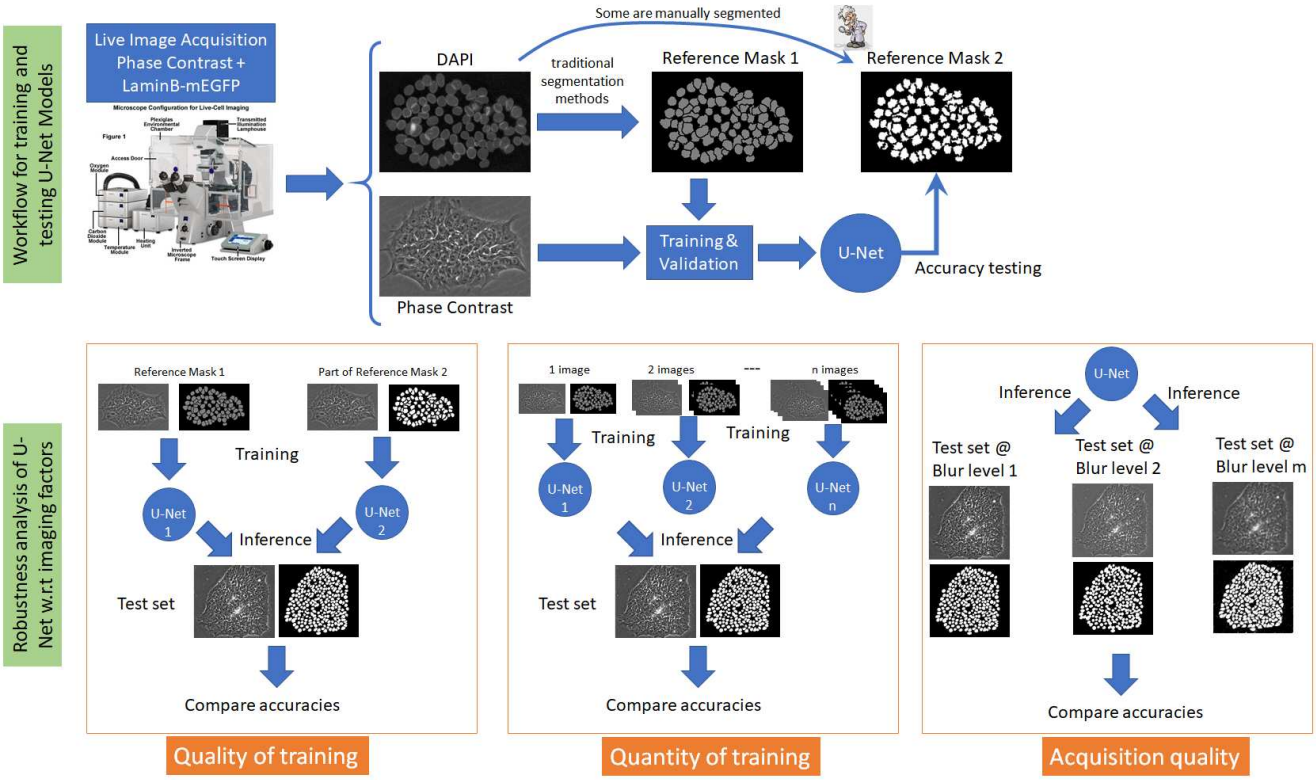

Figure 1. General workflow for training and testing a U-Net model and for the robustness analysis with respect to three imaging factors: Quality and Quantity of training set and the Acquisition quality.

lap of adjacent fields of 10%. At each field, a series of through-focus images were acquired in phase contrast and fluorescence modes. For phase contrast imaging, the sample was exposed to light from a low-power LED (centered at 525 nm, Thorlabs, Newton, NJ) with Kohler aligned Zernike phase contrast optics. For fluorescence imaging of the LamB1:mEGFP, the sample was excited with an LED (470 nm, Thorlabs, Newton, NJ) passed through a filter cube with an excitation filter (470 nm ± 20 nm), emission filter (525 nm ± 20 nm), and a dichroic mirror centered at 495 nm (HE38 GFP filter set, Zeiss, part number 489038-9901-000). The fluorescence excitation power and the phase contrast illumination power were 540 µW and 26 µW, respectively. A spatial calibration target was used to determine that each pixel is equivalent to an area of 0.394 μm^2. The image data sets collected and used for analysis in this study are in Table 1.

Table 1. Four different cell seeding densities were imaged in 4 different locations for a total of 16 image sets. Each image set consisted of a 7 × 8 tiled array of fields of view corresponding to approximately 5.6 mm × 4.8 mm of the sample.

Cell Seeding Density	Days in Culture Prior to Imaging	Z Planes Collected (Focal Planes)	Transmitted Illumination Energy per Z Plane
11 100 cm^2	2	±12, ±9, ±6, ±3, 0 µm	2.6 µJ
22 200 cm^2	2	±12, ±9, ±6, ±3, 0 µm	2.6 µJ
11 100 cm^2	3	±12, ±9, ±6, ±3, 0 µm	2.6 µJ
22 200 cm^2	3	±12, ±9, ±6, ±3, 0 µm	2.6 µJ

2.2. Segmentation Accuracy Metric

Manual segmentation differs between expert scientists at a pixel level [5]. Therefore, we compute the accuracy of segmentation based on the entire nucleus detection instead of on a pixel level [1]. Considering the final segmentation result is a labeled mask instead of binary, the commonly used DICE/F1 metric is not applicable. While mIOU could be used, we decided to report the confusion matrix because the exact borders of the nuclei are less important than whether they were detected at all. We start by computing the overlap matrix between nucleus detected in the reference masks (using manual segmentation) and the ones detected by the U-Net model. The overlap matrix is then

normalized by the size of each manually detected nucleus. We allow a 20% overlap buffer before assigning a manually detected nucleus to one detected by the U-Net model. The confusion matrix composed of the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) is computed on the normalized overlap matrix and the end results are normalized to the number of manually detected nuclei. Figure 2 displays an example of this computation. Our main interest is to maximize the TP rate while minimizing the FP rate. For the example in Figure 2, $TP = 2, n = 5$, hence $TPrate = 2/5 = 0.4$ or 40% and $FPrate = 3/5 = 60\%$.

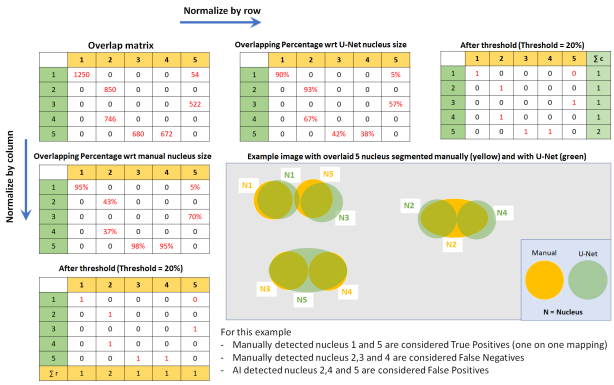

Figure 2. Example of how we compute the segmentation accuracy metrics.

2.3. Training and Testing U-Net Model

The segmentation methods EGT [3] and FogBank [2] failed to reach a correct segmentation of single stem cell nuclei within large colonies in phase contrast image modality, even though these methods have been shown to be very powerful for other applications [1]. Therefore, we decided to use a U-Net model architecture. Since the U-Net requires large training sets, we acquired fluorescent LamB1:mEGFP images of the nuclei in individual cells. Compared to phase contrast images, fluorescence images are easier to segment with traditional segmentation techniques. The training set consists of the phase contrast images (as input to the network) and the segmented nuclei (as output from the fluorescent LamB1:mEGFP images). Images segmented with our network need to be post-processed to separate touching cell nuclei, a considerably more amenable task than single pixel semantic segmentation of phase contrast cell images. We accomplish this using a FogBank segmentation [2] on the inferenced images.

Our input datasets are stitched images with a size of approximately 9000×8000 pixels ($\approx 133\,\mathrm{MB}$). That size exceeds the GPU memory while training the network. Therefore, we cut the large datasets into subsets of images of size 256×256 pixels with 10% overlap.

We acquired a total of 11 datasets. Subsets corresponding to 9 of these datasets were used for training and validation sets, and the remaining 2 datasets were manually segmented for testing the accuracy of the U-Net segmentation. Figure 3 displays some visual results of a U-Net segmentation overlaid on top of the original phase contrast image. The U-Net model used to segment these images was trained on 4 datasets and gave an accuracy of TP rate of 91.1% and a FP rate of 10.5%.

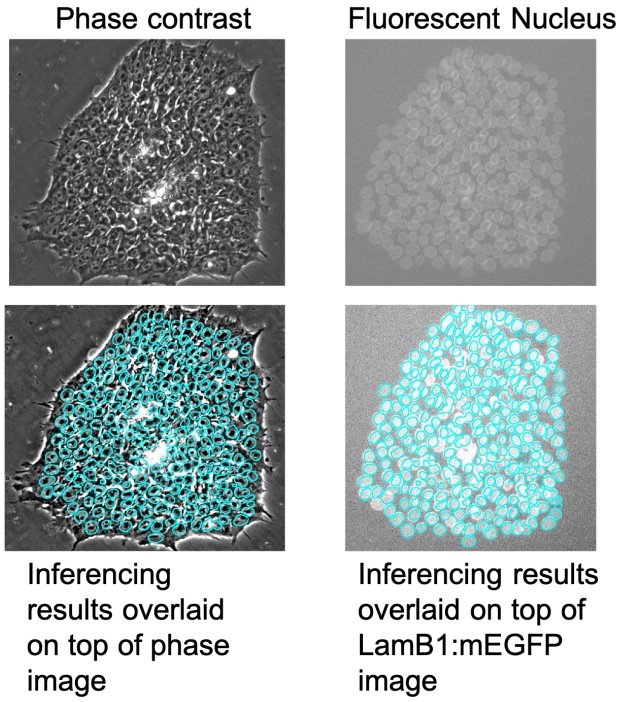

Figure 3. Example of how we compute the segmentation accuracy metrics.

3. Sensitivity Analysis to Quantify U-Net Robustness

We quantify the robustness of the U-Net model with respect to the following three factors: 1) automated vs manual training annotations, 2) quantity of training data, and 3) microscope focus during live acquisition (level of blur in acquired images).

3.1. Impact of Manual vs Automatically Generated Training Data

Given that manual segmentation is resource intensive, we aim to examine whether the training set can be generated using automated segmentation. We manually segmented 2 randomly chosen datasets from the 11 datasets acquired in our collection. We tile the two large stitched datasets into smaller images (blocks of size 256×256 pixels) as shown

in Figure 4. We randomly split the 608 blocks into training/validation and testing, repeating the process 10 times. Two U-Nets were trained using either manual or automated segmentation to create the prediction masks. The performance of each U-Net was evaluated against the test set consisting of manually segmented masks. Figure 5 shows the results of this comparison as an average accuracy with error bar over the 10 runs. The accuracy of the U-Net network trained on manually annotated data has a FN rate of 18% and a false positive rate of 17% while the U-Net trained on automatically generated data has a FN rate of 19% and a false positive rate of 16%. This implies that the overall accuracy of the network trained on manual segmentation is very similar to the one trained on the automated one. We conclude that, for our problem of segmenting nucleus from phase contrast images, the process of generating reference training data requires no need to manually segment single cells.

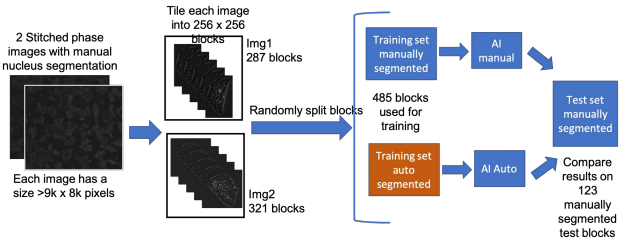

Figure 4. Comparison of manual vs automatic generation of training sets.

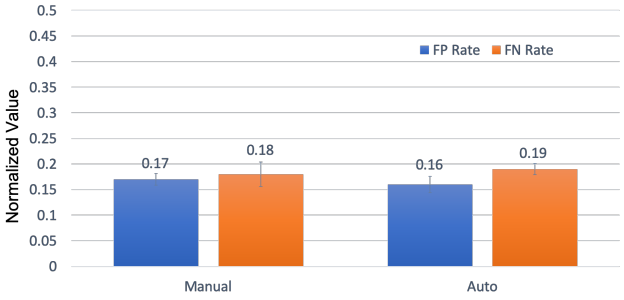

Figure 5. Results of accuracy between manual vs automated training set generation.

3.2. Impact of the Quantity of the Training Set

Results from the previous section suggest that automated segmentation of fluorescent nuclei in single cells from our images is sufficient. This finding reduces the burden on the data acquisition expert. However, acquiring these large image datasets is still time and resource intensive. Thus, we analyze the sensitivity of the overall accuracy of the U-Net model with respect to the number of images used for training, to determine the minimum number of images required to train the U-Net model. We use the two manually segmented datasets as test data to compute the U-Net segmentation accuracy. We perform the sensitivity over a range from 1 to 9 image datasets automatically segmented as training data. The process of choosing the input image is randomized and repeated 9 times when possible. Figure 6 displays the results of this analysis. The conclusion is that after 4 images, the network reaches a stable accuracy of TP rate around 88% and a FP rate around 13% and demonstrates that additional training data beyond four image datasets provides negligible improvement in the inference results. It is noteworthy that each image has a size of roughly 9000×8000 pixels which amounts to 72 million pixels. Since the network is operating on a pixel level, that amounts to roughly 288 million pixels as input data for training the U-Net.

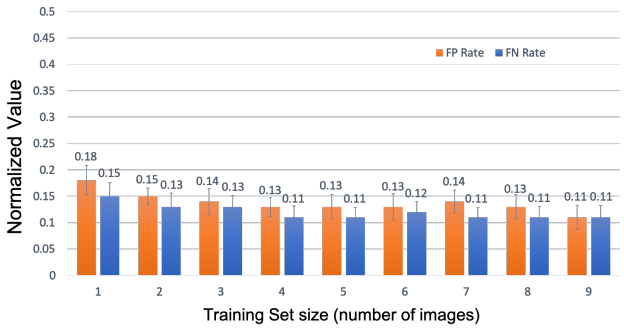

Figure 6. Sensitivity analysis over number of images used in the training set.

3.3. Sensitivity of U-Net Model Inference Results to Image Focus

The robustness of the U-Net model with respect to the amount of blurriness in the images was evaluated. The image blur originated from a z stack of images acquired around the focal plane for each field of view from the CCD camera (Figure 7). Encountering blurred images during acquisition is inevitable. Automated microscopy experiments can utilize different focus strategies, each giving rise to a different final in-focus image. During time lapse acquisitions, these "in-focus" views can also vary. This analysis characterizes the U-Net sensitivity for focus blur and can be used to specify the required focus precision of the microscope system. The range of the z-stack above and below the selected focus plane was selected based on the range of focus blur (the farther from the focal plane the blurrier) that might be encountered during an acquisition. We used the previously trained U-Net model, trained on 4 large image datasets to inference 9 datasets acquired at different z levels. Figure 8 displays the average FP and FN rates and standard deviations over the 9 images for each ΔZ. ΔZ is the

step in µm from the microscope autofocus value (in-focus is considered equal to 0). The network appears to be quite sensitive to out of focus images, with the accuracy dropping by 10% with an approximately 3 µm divergence from the in-focus plane. Despite training the network by augmenting the training set using a Gaussian blur with a sigma up to 10 (a sigma of 10 produces blurrier images in simulation than the ones coming from the z-stack), the network still had trouble with high blur in image acquisition. This issue highlights the interdependence between image acquisition and image analysis, and the importance for image quality metrics like detecting level of blurs in the images by using several blur metrics [9]. We will further examine this problem by using other blurring techniques in the data augmentation step. Another solution might be the use of reference beads during acquisition to help the autofocus of the microscope improve the quality of the acquired image.

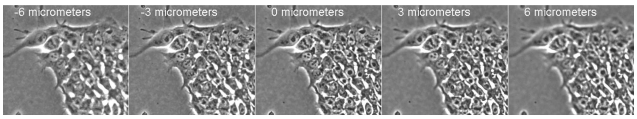

Figure 7. Example of blur level in acquired images at multiple focal planes.

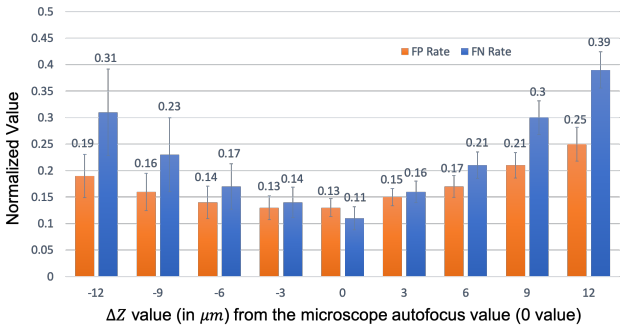

Figure 8. Sensitivity analysis over level blurriness of acquired images.

4. Conclusion

We used a U-Net model to segment single stem cell nuclei within colonies when traditional segmentation techniques did not achieve a desirable accuracy. We quantified the robustness of the U-Net with respect to important factors: the quality and quantity of training annotations and the quality of microscope image focus. We found that, for our experimental system based on fluorescently labeled nuclei, that automated segmentation was sufficient and that a training size of 4 stitched images of size 9000×8000 pixels was adequate to reach an accuracy above 88% in TP rate and below 15% in FP rate. We also found that the networks accuracy decreases with poor focus (high level of blur) and that is a problem that needs to be addressed in the future. We will try using other blurring techniques in the data augmentation step and/or use reference beads during acquisition to help the autofocus of the microscope improve the quality of the acquired image.

5. Disclaimer

Commercial products are identified in this document in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology (NIST), nor is it intended to imply that the products identified are necessarily the best available for the purpose.

References

[1] Juan C Caicedo, Jonathan Roth, Allen Goodman, Tim Becker, Kyle W Karhohs, Matthieu Broisin, Csaba Molnar, Claire McQuin, Shantanu Singh, Fabian J Theis, et al. Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry Part A*, 95(9):952–965, 2019.

[2] Joe Chalfoun, Michael Majurski, Alden Dima, Christina Stuelten, Adele Peskin, and Mary Brady. Fogbank: a single cell segmentation across multiple cell lines and image modalities. *Bmc Bioinformatics*, 15(1):431, 2014.

[3] Joe Chalfoun, M Majurski, A Peskin, Catherine Breen, Peter Bajcsy, and M Brady. Empirical gradient threshold technique for automated segmentation across image modalities and cell lines. *Journal of microscopy*, 260(1):86–99, 2015.

[4] Benjamin Chidester, That-Vinh Ton, Minh-Triet Tran, Jian Ma, and Minh N Do. Enhanced rotation-equivariant u-net for nuclear segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[5] Alden A Dima, John T Elliott, James J Filliben, Michael Halter, Adele Peskin, Javier Bernal, Marcin Kociolek, Mary C Brady, Hai C Tang, and Anne L Plant. Comparison of segmentation algorithms for fluorescence microscopy images of cells. *Cytometry Part A*, 79(7):545–559, 2011.

[6] Mina Khoshdeli, Garrett Winkelmaier, and Bahram Parvin. Fusion of encoder-decoder deep networks improves delineation of multiple nuclear phenotypes. *BMC bioinformatics*, 19(1):294, 2018.

[7] Florian Kromp, Lukas Fischer, Eva Bozsaky, Inge Ambros, Wolfgang Doerr, Sabine Taschner-Mandl, Peter Ambros, and Allan Hanbury. Deep learning architectures for generalized immunofluorescence based nuclear image segmentation. *arXiv preprint arXiv:1907.12975*, 2019.

[8] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36(7):1550–1560, 2017.

[9] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, 2013.

[10] Marie Piraud, Anjany Sekuboyina, and Bjorn H Menze. Multi-level activation for segmentation of hierarchically-nested classes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, may 2015.

[12] Aarno Oskar Vuola, Saad Ullah Akram, and Juho Kannala. Mask-rcnn and u-net ensembled for nuclei segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 208–212. IEEE, 2019.

[13] Fuyong Xing, Yuanpu Xie, Xiaoshuang Shi, Pingjun Chen, Zizhao Zhang, and Lin Yang. Towards pixel-to-pixel deep nucleus detection in microscopy images. *BMC bioinformatics*, 20(1):1–16, 2019.

[14] Zhaoyang Xu, Faranak Sobhani, Carlos Fernández Moro, and Qianni Zhang. Us-net for robust and efficient nuclei instance segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 44–47. IEEE, 2019.

[15] Pengyu Yuan, Ali Rezvan, Xiaoyang Li, Navin Varadarajan, and Hien Van Nguyen. Phasetime: Deep learning approach to detect nuclei in time lapse phase images. *Journal of clinical medicine*, 8(8):1159, 2019.