Predicting subcellular organelle localization from label-free brightfield microscopy images

Anjali Bisaria
abisaria@stanford.edu

Yilin Fan
yilinfan@stanford.edu

Nalin Ratnayeke
nalinr@stanford.edu

Abstract

Fluorescence microscopy is often limited by the number of distinct channels that can be simultaneously resolved. This in turn restricts the number of inputs that can measured, for example, in the same biological system. Here we aim to train a U-Net based convolutional neural network to predict the location of cellular organelle from bright-field images, thereby freeing up a "color" that would normally used for cell segmentation or tracking. We provided labeled masks for nuclear and cytoplasmic areas automatically generated from fluorescent labeling probes and trained the model on 1500 images. Our model achieved an F1 score of 0.852 on pixel-level semantic segmentation tasks. Furthermore, when focused on only being able to predict nuclei, our bright field based segmentation achieved an F1 score of 0.894, and achieved good accuracy when trained on smaller subsets of full training data. This high performing segmentation is on par with other similar implementations, and is more than adequate for use in research applications (e.g. segmentation, tracking, signal quantification).

1 Introduction

Molecular biology and medicine increasingly rely on high-throughput microscopy to understand the dynamics of complex biological processes. A large repertoire of fluorescent markers have been developed which can directly report cellular signaling and regulation in single cells over time. Given the advancement of automated microscopy in which hundreds of thousands of images can be acquired in a single experiment, computer-based automated image analysis has become increasingly necessary for biology. A traditional imaging-based pipeline has cells stained with multiple fluorescent markers, one of which is a fiduciary marker to segment and identify single cells for tracking and further signal quantification. Researchers are often limited by the number of different fluorescent colors that can be simultaneously captured, which is further hindered by the need for said segmentation marker. Some groups have started to overcome this challenge by combining fluorescence imaging with orthogonal imaging modalities such as brightfield imaging, which can give information on cell shape and subcellular structure without using molecular probes or taking up a fluorescence channel. In addition to augmenting fluorescence microscopy experiments, brightfield imaging can be used alone to track cell movement and proliferation in cells lacking fluorescent markers on relatively low-cost imaging equipment. Despite these advantages, brightfield imaging and similar techniques produce low-contrast and complex images, necessitating the use of deep-learning to segment cells in these experiments (Moen et al.). Here we implement a deep learning-based segmentation pipeline which uses brightfield images to predict sub-cellular organelle location (nuclear and cytoplasm). We use a U-net CNN architecture for semantic segmentation to predict if each pixel is in one of

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

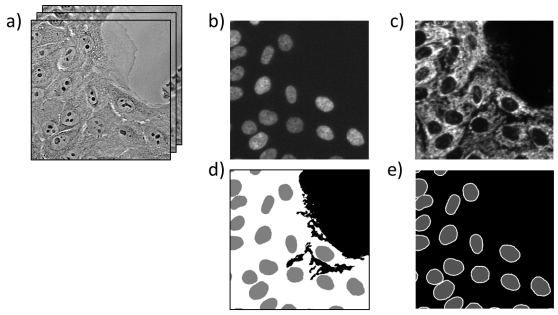

Figure 1: Sample images and matching masks. a) Brightfield images. b) Nuclear stain. c) Cytoplasmic stain. d) Background-nucleus-cytoplasm mask. e) Background-nucleus-boundary-split mask.

three classes: background, nuclear, or cytoplasmic. The algorithm was trained on masks that were generated using automated pixel-wise ground-truth annotation using gold-standard chemical probes.

2 Related work

Currently, most high through-put single-cell analytical pipelines use automatic thresholding or edge detection of high-contrast fluorescent markers for segmentation, as brightfield imaging produces low-contrast and complex images which are not suitable for traditional segmentation methods. Previous work using U-Net (Ronneberger, Fischer, and Brox) architectures have been very successful at reconstructing various sub-cellular structure from brightfield imaging. For our particular task of nuclear segmentation using brightfield, a previous U-Net alone approach has yielded a very high accuracy (0.89-0.97, F1 0.76-0.86) (Fishman et al.). Using a Deep Neural Net another group was able to outperform manual ground truth annotation(88% vs 85% of cells were segmented with an F-score0.6) (Sadanandan et al.). Some groups have moved beyond multi-class segmentation and have begun using CNN and U-Net architectures to directly predict fluorescent signals based on non-fluorescence microscopy. In these cases, the output of training is an inferred fluorescent signal intensity from bright field images (Guo et al.,Ounkomol et al.,Christiansen et al.). While these methods require less data preprocessing for generating the training set, as no segmentation is required, for a high prediction-implementation of intensity prediction, more advanced transmitted-light measurements such as quantitative phase-contrastGuo et al. or differential-interference microscopy are required. Thus, rather than predicting fluorescence intensity and then segmenting based on predicted fluorescence values, we favor an end-to-end approach which would take as an input simple brightfield images and output segmentation maps directly. Furthermore, there has been previous work done on instance segmentation for label-free cell tracking to identify single cells and their masks (e.g. Mask R-CNN). However these methods are more intensive computationally, and given the regularity of cell shapes it is relatively straightforward to identify individual objects using morphological analyses and transformations from a semantic segmentation map.

3 Dataset and features

We acquired images on an automated ImageXpress Micro microscope, which can acquire brightfield images together with multi-color fluorescence images in quick succession on 96-well cell culturing plates. To generate our training set, we simultaneously imaged brightfield images taken at an in-focus imaging plane and 3 μm above and below (Figure 1a), and fluorescence images of nuclear stain Hoechst 33342 and mitochondrial stain MitoTracker (for identifying cytoplasm) (Figure 1b, 1c). In total, we acquired 1,920 sites using a 20x (0.75 NA) objective on a digital camera capturing 16-bit 2160x2160 images. The 1920 images were randomly split into training-dev-test sets with a 80-10-10 split (1536-192-192 images). For each set of images, we generated ground truth masks of the nucleus and cytoplasm though traditional segmentation methods of the fluorescent stains. The two masks were then flattened with the nuclear mask on top to generate a 3-class pixel-wise classification mask (background-nucleus-cytoplasm). The resulting masks contained fused cells, where neighboring cells are not split by boundary pixels. Typically, cell nuclei can be split using concavity-based

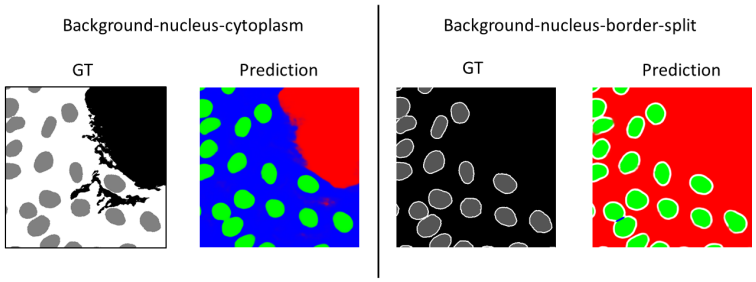

Figure 2: Example images of ground truth (GT) and predictions made by models trained on single-plane brightfield images

morphological splitting, while cell cytoplasm can be split using a watershed algorithm. However the cytoplasmic stain is not cleanly separated between neighboring cells and thus we concentrate on separating cell nuclei, which is sufficient for many experiments. We propose two schemes for splitting cell nuclei: 1) predicting a background-nucleus-cytoplasm masks with the CNN and then morphologically splitting masks of touching or nearby nuclei, or 2) morphologically splitting nuclei in the training mask, creating a new label for the split region (3 pixels wide) as well as a new label for nuclear boundaries (3 pixels wide). In this alternate labeling scheme, we combine cytoplasm and background classes, creating a 4-class background-nucleus-boundary-split mask. Thus, our two segmentation schemes involve either a rough segmentation (background-nucleus-cytoplasm, Figure 1d) which is later cleaned up by object splitting, and a single-step analysis wherein the network explicitly learns split regions (background-nucleus-boundary-split, Figure 1e).

4 Methods

To generate semantic segmentation maps we utilize an existing implementation of the U-Net architecture (https://github.com/qubvel/segmentation_models). The U-Net is a fully convolutional network which is made up of a series of encoding convolutional layers followed by decoding layers which upsamples the output to the same size as the input. Skip connections between the encoding and decoding layers allow small-scale information to pass to the output of the network. In this way, a segmentation map comprised of pixel-wise classifications can be output. To speed up training, we utilized the convolutional layers of the VGG16 network as the encoder, with weights pretrained on ImageNet. During training, we randomly crop images to 640x640 (mini-batch size 4) to accommodate memory limitations, and randomly apply image rotation and reflection and use a hybrid loss function (Dice loss + focal loss). Efficientnetb4 was trained using input images of 320x320 (minibatch size 8). Dev and test set evaluation was performed on whole images.

5 Results and discussion

We first trained models on ground truth masks from the 3-class labeling scheme (background-nucleus-cytoplasm). To evaluate our segmentation models, we used pixel-level intersection-over-union (IoU) and F1 scores (harmonic mean of precision and recall) as primary metrics. We performed optimization for at least 20 epochs, and monitored loss on the training set, which converged after approximately 10 epochs (data not shown). The best model from each training session was selected based on the lowest loss evaluated on the training set.

Next we evaluated model performance on the dev set. The model qualitatively captured cellular features including the nucleus and cytoplasm (Figure 2, left panel). By examining quantitative performance metrics (Table 1), we found that a model trained on single-plane brightfield images performed similarly to a model trained on three-plane brightfield images (compare rows 1 and 2). Using single-plane images as model inputs has additional advantages, more specifically shorter acquisition time during microscopy imaging (at least three-fold) and faster training. Thus, we focused on single-plane images as we explored hyperparameters and alternative labeling schemes.

Table 1: Model metrics evaluated on the dev set

Scheme	Loss	IoU (average over classes)	F1 (average over classes)	IoU (nucleus)	F1 (nucleus)
Background-nucleus-cytoplasm (three-plane image)	0.08477	0.77114	0.84612	0.78615	0.87966
Background-nucleus-cytoplasm (single-plane image)	0.082248	0.77612	0.84943	0.79262	0.88383
Background-nucleus-boundary-split (single-plane image)	0.4144	0.52031	0.62138	0.75169	0.85767

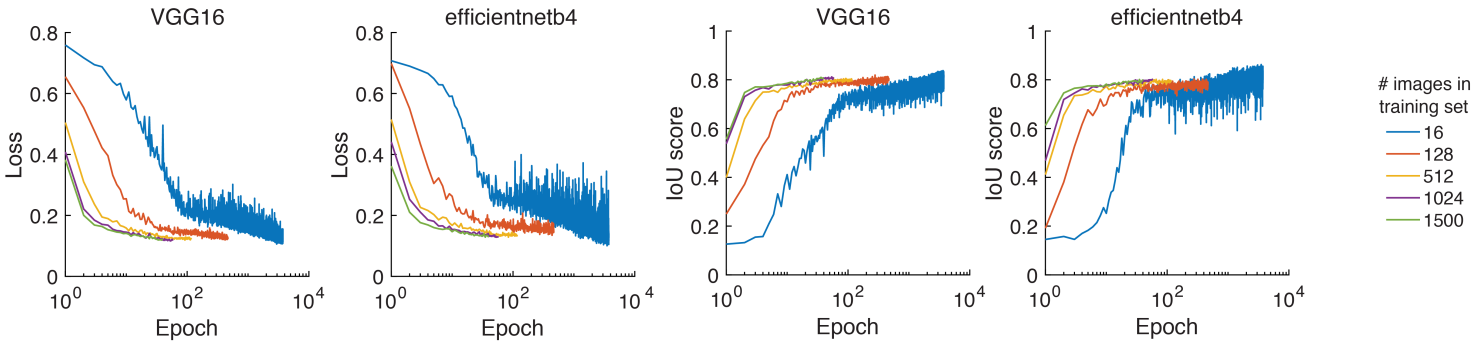

Figure 3: Loss and IoU during training based on number of training images and model architecture

To split fused and sometimes overlapping nuclei, we explicitly trained the model on nuclear outline ("boundary") and border region between two neighboring nuclei ("split"). Qualitatively, the model recovered both boundary and split regions (Figure 2, right panel, note the blue split region between two nuclei). However, we observed that predicted labels are often thicker than ground truth labels, resulting in higher loss and lower IoU/F1 scores (Table 1). We currently do not have a quantitative measure for the accuracy and efficiency of splitting fused nuclei. We propose to manually split nuclei based on the nucleus-specific Hoechst stain, against which different models will be evaluated.

Next, we explored the effect of training set size (number of images), loss function, and model architecture on model performance, focusing on predicting 3-class labels (background-nucleus-cytoplasm) from single-plane images. We were able to achieve similar loss and IoU score during training when using a smaller training set (Figure 3), suggesting that smaller training set is not causing a significant bias problem. When evaluating on the test set, we found that 500 images were sufficient to train a model with a high IoU score comparable to larger training sets (Figure 4). Additionally, we found that a model using VGG16 as the encoder performed better than efficientnetb4 (Figure 4, Table 2). And we found that hybrid loss (sum of Dice loss and focal loss) performed better than either Dice loss or focal loss alone (Table 2).

While examining ground truth and predicted labels, we noticed that the model performs poorly on mitotic cells, which comprise a small fraction of total cells (~5%) and display a distinct morphology. We propose to manually annotate mitotic cells based on brightfield and Hoechst stain, and include them as an additional label class.

6 Contributions

A.B., Y.F., and N.R. conceived the project, performed the experiments, and wrote the report.

7 Code availability

Codes were uploaded to Github: https://github.com/nalinratnayeke/CS230-cell-segmentation. Codes were adapted from https://github.com/qubvel/segmentation_models.

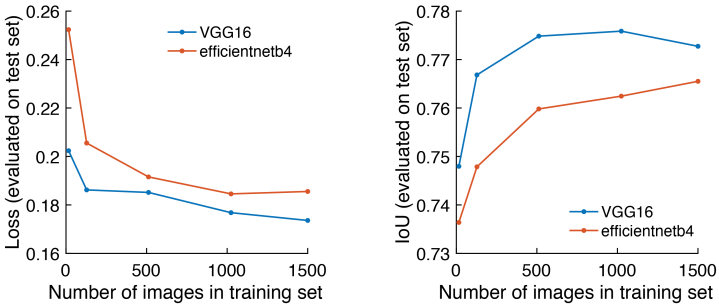

Figure 4: Loss and IoU on test set based on number of training images and model architecture

Table 2: Model metrics evaluated on the test set

Schemes	Loss	IoU (average over classes)	F1 (average over classes)	IoU (nucleus)	F1 (nucleus)
Different loss functions					
Hybrid loss	0.085647	0.78086	0.85228	0.80887	0.89389
Dice loss	0.080889	0.76105	0.83958	0.78483	0.87895
Focal loss	0.001512	0.77823	0.85296	0.78522	0.87921
Different model architectures					
VGG16	0.085647	0.78086	0.85228	0.80887	0.89389
efficientnetb4	0.092493	0.76835	0.84596	0.77549	0.87266

References

Works Cited

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015. 234–241. Print.

Sadanandan, Sajith Kecheril, et al. "Automated training of deep convolutional neural networks for cell segmentation". *Scientific reports* 7.1 (2017): 1–7. Print.

Christiansen, Eric M, et al. "In silico labeling: predicting fluorescent labels in unlabeled images". *Cell* 173.3 (2018): 792–803. Print.

Ounkomol, Chawin, et al. "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy". *Nature methods* 15.11 (2018): 917–920. Print.

Zhang, Zhengxin, Qingjie Liu, and Yunhong Wang. "Road extraction by deep residual u-net". *IEEE Geoscience and Remote Sensing Letters* 15.5 (2018): 749–753. Print.

Fishman, Dmytro, et al. "Segmenting nuclei in brightfield images with neural networks". *bioRxiv* (2019): 764894. Print.

Guo, Syuan-Ming, et al. "Revealing architectural order with quantitative label-free imaging and deep neural networks". *BioRxiv* (2019): 631101. Print.

Kromp, Florian, et al. "Deep Learning architectures for generalized immunofluorescence based nuclear image segmentation". *arXiv preprint arXiv:1907.12975* (2019). Print.

Moen, Erick, et al. "Deep learning for cellular image analysis". *Nature methods* (2019): 1–14. Print.

Schormann, Wiebke, Santosh Hariharan, and David W Andrews. "A reference library for assigning protein subcellular localizations by image-based machine learning". *The Journal of Cell Biology* 219.3 (2020). Print.