Deep Learning Methods in Automatic Nuclei Detection in Microscopy

Authors:
Daniel Wu (5214001) (wuxx1495@umn.edu)

Professor: Nikos Papanikolopoulos
Course: CSCI 5561

Abstract

Creating an algorithm to automate nucleus detection in medical imaging could help accelerate research in being able to identify drugs and treatment solutions to diseases that lack any solution currently. [4] This would free resources in medical science labs to focus on developing medical treatment, while also improving the efficiency of data analysis in determining the effectiveness of medical or drug treatments. The goal of this project is to attempt to develop computer vision algorithms in combination with machine learning algorithms to attempt to automatically identify and segment nuclei in microscope images from various conditions, stains, and image sources. Traditional connected region extraction with Otsu thresholds was compared with current convolutional networks U-Net and Mask R-CNN. [6][9] Using an Intersection over Union average metric to compare the methods, Mask R-CNN outperformed U-Net and Otsu threshold region extraction.

Introduction

Drug research efficiency has been on the decline in recent years, with the number of new drugs approved per billion US dollars halved roughly every 9 years since 1950 [3]. While there are many conditions to this, one of the main bottlenecks is the underlying process of how drug research methodologies rely on software to analyze the effectiveness of drugs. One of the primary ways in which drug research is developed is through assays that evaluate cell response to a drug candidate.

Current traditional methods are able to measure various properties of cells and are able to identify benign, uniform cell nuclei with ease. [1] This allows scientists to measure nuclei size with ease, but when tasked with irregular nuclei with irregular shape, or entire tissue samples, these algorithms tend to fail and biologists have to manually count and identify nuclei. If a model can identify nuclei with images across different experimental systems, such as resolution, stains, and microscopic backgrounds and environment, this would save much of the research time.

A tedious task that remains in medical research is the ability to identify nucleus in microscopy images from various imaging sources, including light, electron microscopy, and fluoroscopy. While there have been advances in automating the task of capturing microscopy images of cells quickly, a major bottleneck is in properly identifying nuclei in the images. Quick and correct identification nuclei in pathology helps enhance drug development, advance medical research that could lead to novel medical treatment methods.

In this paper, novel approaches using a combination of deep learning and computer vision techniques will attempt to segment nuclei and the results will be compared using a calculated metric.

Data

The ideal model would work across all imaging modalities in biology, no matter the size of the nuclei, color, or resolution of the nuclei. This would eliminate additional steps that biologists would need to take to factor in different types of image acquisitions in hypothetical software that biologists would use. To attempt to develop a model, the training set provided by Kaggle in the Data Science Bowl 2018 Competition consists of 670 training images from a variety of sources. [2] Masks were provided by scientists that manually segmented the nuclei from all images. After reviewing images, 3 images were taken out of the training set given that they were leaks from the test image or masks were improperly segmented. Furthermore, two main types of images were incorporated in the training data set, fluorescent microscopy and brightfield microscopy with some images stained. The test dataset consisted of 65 images and associated annotated masks that were unseen when used with training data.

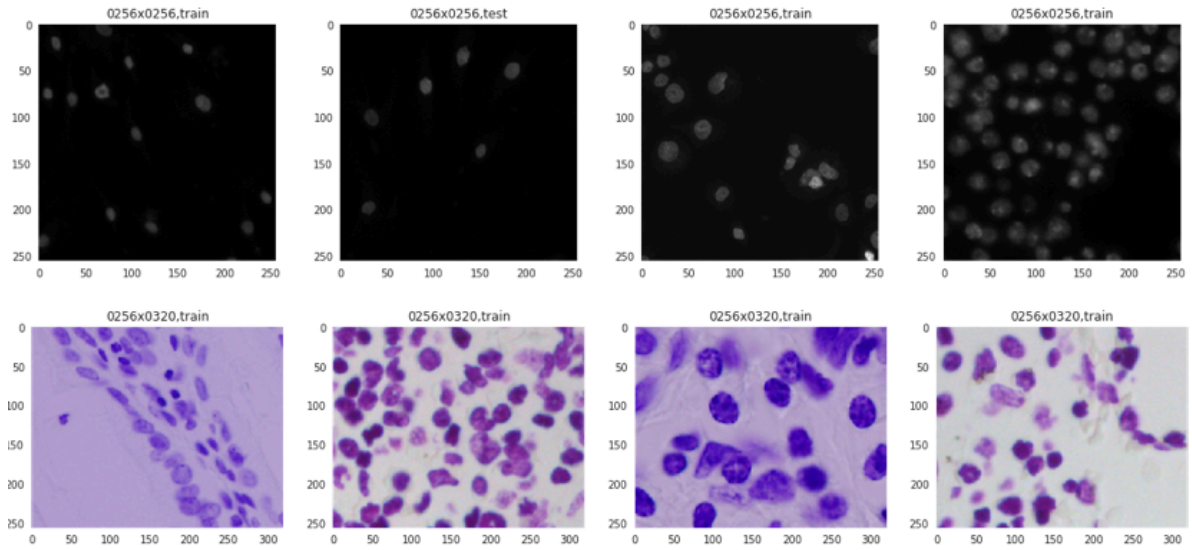

Figure 1: Sample of different imaging modalities. The top row is greyscaled images from a fluorescent microscope. The bottom row is from a light microscope stained with Hematoxylin and eosin stain (HE) stain.

Methods

Evaluation Metric

In order to compare different methods of nuclei segmentation, the mean average precision of different intersection over union (IoU) thresholds was used as an evaluation metric. The IoU of a predicted region, A, and the provided true region, B, of the nuclei pixels is calculated with the equation: [2]

$$IoU(A,B) = \frac{A \cap B}{A \cup B}$$

The IoU is then threshold with values ranged from 0.5 to 0.95 in .05 increments. [2] An object is considered detected if its IoU is above the threshold, meaning that the predicted region intersects with a ground truth region greater than the set threshold. For each threshold, an evaluation metric is proposed that looks at the proportion of true positive (TP), compared to the sum of TPs, false negatives (FN), and false positives (FP) resulting in that comparison. This precision metric is calculated for each image for each threshold t:

$$Precision(t) = \frac{TP(t)}{TP(t) + FP(t) + FN(t)}$$

A TP means that a single predicted region matches a ground truth region with the IoU above the threshold. A FP means that a predicted region had no associated region from the ground truth. Finally, a FN means that a ground truth region was not associated with a predicted region. For each region in each image, an average precision metric is calculated across each IoU threshold using the equation [2]:

$$Avg.Precision = \frac{1}{n_{thresholds}} \sum_{t} \frac{TP(t)}{TP(t) + FP(t) + FN(T)}$$

This is essentially an average of the precision values across the threshold range to get the mean precision for each image. This is averaged over each image to compute our final precision metric over the entire dataset.

Basic Computer Vision Segmentation

In attempting basic segmentation techniques using traditional region processing computer vision techniques of Otsu thresholding. [5]

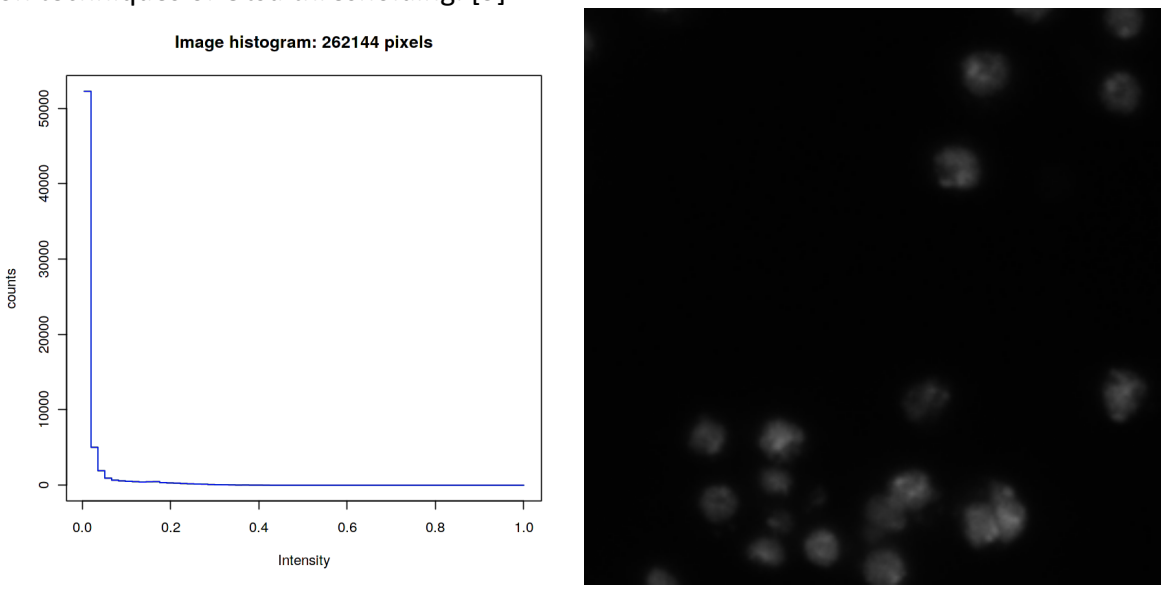

Figure 2: Sample image and histogram of image from the training data. Based on this histogram, Otsu thresholding seemed to be ideal given that we are attempting to identify only one class (the nuclei) from the background.

Most of the images were RGB with 3 channels. In order to standardize our algorithm, the images were converted to grayscale. Before thresholding for connected component detection, morphological opening was used to attempt to remove noise in the grayscale images. Opening was used for its ability to remove potential small noisy foreground regions while preserving the larger foreground regions from being eroded. The structuring element size was determined by observing the given region mask shapes in the training dataset. An elliptical structuring element was used given that the region of interest of nuclei will often be rounded, with the majority being approximated by an ellipse. The test dataset was assumed to be similar in distribution of image sizes. Because there is no training required, the method was not changed for the test images. Optimization was performed using basic grid search to find optimal parameters for the structuring element and other parameters. Then dilation and erosion are used in post-processing the masks of the nuclei before generating the final predicted masks in images.

U-NET

The U-NET model was trained using Google Colaboratory, a Jupyter notebook environment that utilizes a NVIDIA Tesla K80 GPU. The implementation was written using Keras with a TensorFlow backend.

The first convolutional network that was implemented is an implementation of U-Net from Ronneberger et al. [6] The overall goal is to segment images into different classes while addressing the limited access to data in the biomedical setting and the need to separate regions that are touching and in the same class. [6]The network consists of two complementary paths, a contracting path and expansive path. The input image is fed through the network at the beginning of the contracting path, and is propagated through the network. The first operations are convolutional layers followed by a nonlinear activation function, using ReLU activation functions. [6] The contraction path creates a feature map, in which the U-net uses on the "expansion" path to create a high-resolution segmentation map. [6]

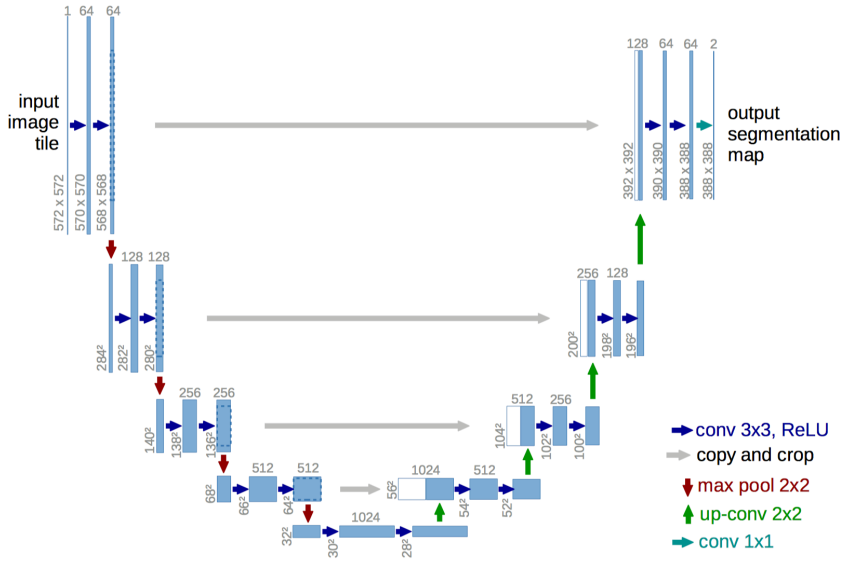

Figure 3: U-Net Convolutional network architecture [6]

From Ronneberger et al., there are several advantages of the U-net architecture. One of the primary benefits is that this architecture can handle various input image sizes since it does not contain fully connected layers. The original intent of U-net was specifically for biomedical image segmentation, which fits the desired use case of this paper. [6] They also specifically attempt to address the problem of touching connected regions that belong to the same class, by assigning loss functions for each individual pixel, with higher weights used for the border pixels created from the ground truth masks.[6] The original paper used this method to segment cells, and will be attempted to be used to segment nuclei.

Mask Regional Convolutional Neural Network (R-CNN)

The Mask R-CNN model was trained using Google Colaboratory, a Jupyter notebook environment that utilizes a NVIDIA Tesla K80 GPU. The implementation was written using Keras with a TensorFlow backend.

Mask R-CNN is based on the Region-based Convolutional Neural Network (R-CNN), but decouples the classification portion and the mask prediction instance segmentation. [9] To briefly summarize, R-CNN attempts to learn to identify bounding-box object region candidates and attempts to classify each region through a Convolutional Neural Network (CNN) feature from each region independently for classification. [10] The initial region proposal candidates are created from a process called Selective Search, which attempts to find multiple potential crops of different sizes to look for clustering of pixels to identify objects. [10] However, initially, the computational requirements are heavy, given that R-CNN requires three separate models to train image features, a classifier, and a regression model for bounding boxes. [7]

There have been numerous improvements to the original R-CNN, with Fast R-CNN combining each region proposal and the image into one forward pass into the network architecture to decrease computation requirements. [7] Fast R-CNN adjusts the last max pooling layer with a RoI pooling layer, which uses the feature map of the CNN on each image to be used for all the ROI proposals rather than running each ROI proposal through the CNN network separately. [7] An additional improvement to this architecture was to train the CNN, classifier, and bounding box regression in a single model. By replacing the SVM classifier with two output layers, a linear regression layer and a softmax layer, all outputs come from one single network. Further improvements came with Faster R-CNN, which incorporates the region proposal network into the existing CNN model from Fast R-CNN instead of using Selective Search. [8] By using the feature map results of the CNN for each image, the region proposals can be generated from those results rather than running a separate selective search algorithm.

Mask R-CNN further extends Faster R-CNN by incorporating a third output for instance segmentation of the masks, in addition to classification and ROI localization. [9] The segmentation portion of the model is a fully-connected network done to each ROI that learns a prediction for segmentation into a binary mask, then using the classification output of the Faster R-CNN network, we can associate that mask with a specific class.

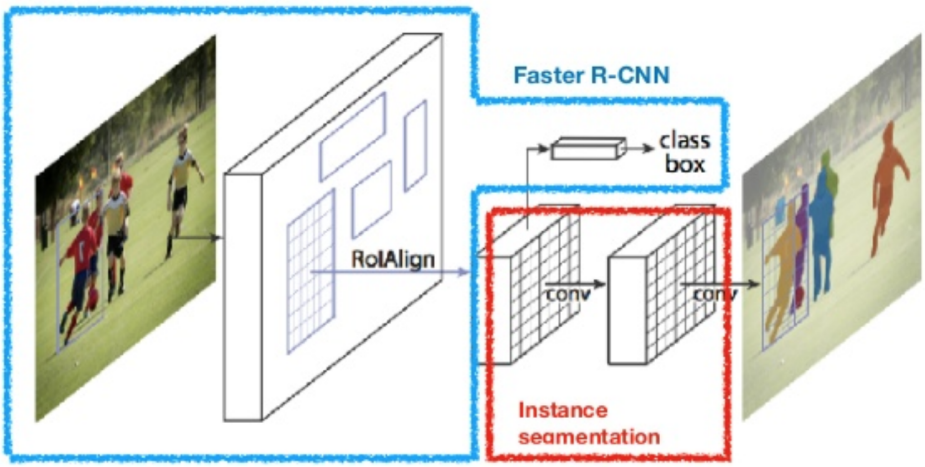

Figure 4: Mask R-CNN is able to segment as well as classify the objects in an image. [9]

Because instance segmentation requires a finely tuned region that does not crop out areas of interest for segmentation, a new RoI pooling layer was developed by He et al. called "RoIAlign layer". [9] This prevents the rounding of the feature maps when regions are contracted in the network in the original Fast R-CNN implementation, and instead uses interpolation to estimate the floating-point values of the region. By avoiding the rounding, this prevents misalignment that is unwanted when attempting segmentation.

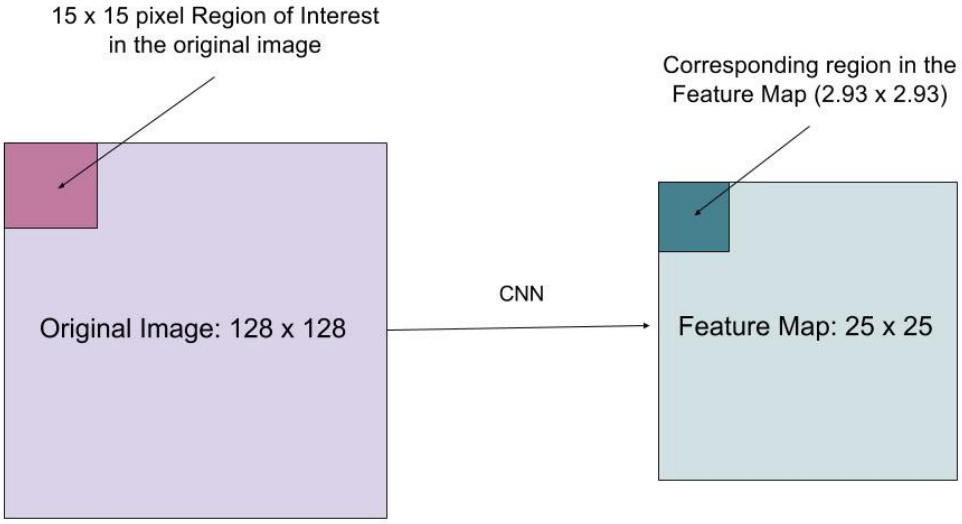

Figure 5: Image from [9] detailing the RoI Align adjustment to Faster R-CNN in Mask R-CNN

Image Augmentation

For U-Net, during training, random image distortions were done: flipping images horizontally and vertically, rotations in 90 degree increments of 90, 180, 270, including random zoom and shifts on images. Gaussian Blur and Additive Gaussian Noise were not used for image augmentation in U-Net

For Mask R-CNN, the implementation used required square inputs of 512x512. Because images were varied in resolution and dimensions, random cropping of images of 512x512 were used with zero-padding when necessary. All images were also converted to grayscale to maintain simplicity without having to consider RGB images and were normalized based on a mean pixel value determined from the entire training data set. During training, random image distortions were done, which included flipped images, rotations in 90 degree increments of 90, 180, 270, and Gaussian blurring and additive Gaussian noise filters. These image augmentations served to prevent overfitting and to help generalize the model to learn beyond the training set.

Results

	Mean Precision
CV	0.229
Unet	0.311
Mask RCNN	0.440

Figure 6: Comparison of Results using Mean Precision Metric

The results of the three methods are displayed in Figure 6. For the Otsu thresholding with connected region extraction using traditional methods. For simple fluorescent microscopic images, this is a relatively straightforward process for connected region extraction. All three methods handled simple images with just nuclei and background with few exceptions.

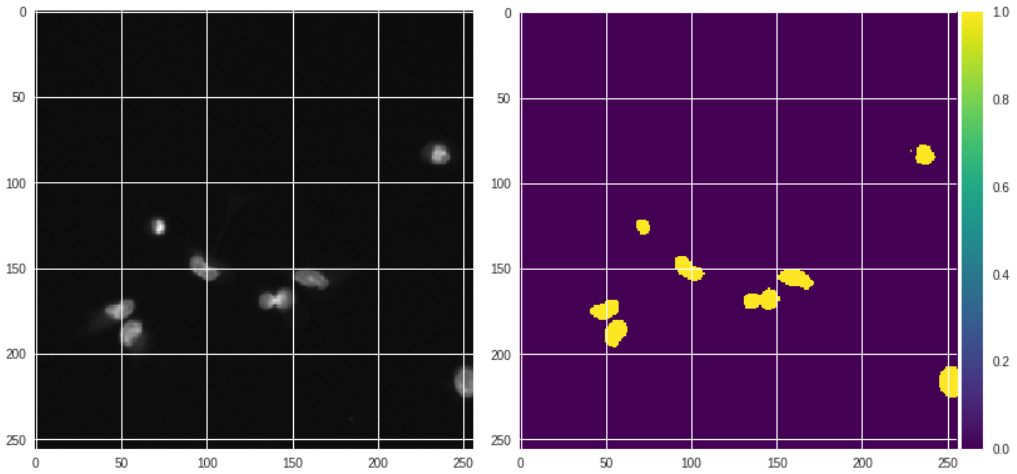

Figure 7: Results using simple Otsu Threshold on a fluorescent microscope image. Similar results were obtained visually using U-Net and Mask R-CNN.

More complicated images are displayed in Figure 8. All methods struggled in being able to find nuclei in the image, when the nuclei vary in shape, color, intensity, and size.

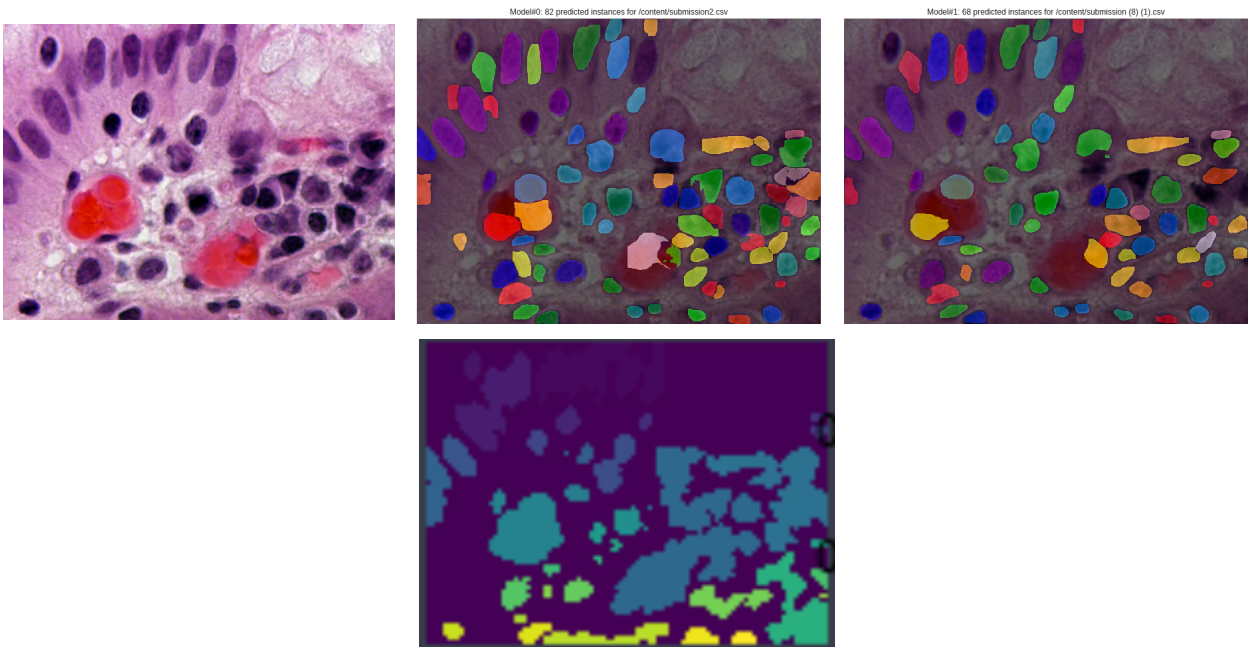

Figure 8: Comparison of Mask R-CNN (middle) and U-Net (right) with the original image (left). Otsu threshold (bottom) is shown for comparison.

Discussion

In reviewing the results, it is clear that CNNs for object segmentation and classification are superior than traditional connected region extraction with Otsu thresholding, even with optimization of parameters. While traditional computer vision techniques were capable of extracting nuclei given a very simple background and high contrast between regions of interest and background, when given more complicated images, such as the brightfield microscopy images with additional staining of not just nuclei but also other cell components, the method fails to separate those regions, and gave a lot of false positive, which hurt the overall evaluation metric.

The performance difference between the U-Net and Mask R-CNN was also a surprising result. While the Mask R-CNN model outperformed the U-Net model in our evaluation, the overall ceiling of both methods remains to be determined. The U-Net model seemed to have trouble differentiating between connected or overlapping regions, and given our evaluation metric this penalized its performance. After initial evaluation of both models, it became clear that preprocessing of the data boosted performance of both models. The type of image augmentations may need to be optimized to improve performance of the model.

Despite the difference in performance of U-Net and Mask R-CNN, additional preprocessing, post-processing, and tuning of model parameters need to be investigated and evaluated before definitive conclusions can be made about the performance each of the models. For U-Net, different loss functions can be evaluated, as in our implementation we used the IoU as the metric, but different losses exist.

Conclusion

Overall, none of the models were able to segment nuclei with very high accuracy, as all methods failed to perform well in complicated images that had multiple cell components and features with abnormal nuclei. The ability to detect nuclei in these conditions is crucial considering that many medical conditions and diseases are detected based on the

Furthermore, the upper-limit of these methods will be determined by the variability in the human manual segmentation between different annotations since we are assuming that those annotations are the ground-truth masks that the models are being trained on. Further steps can be taken to investigate variability between different human segmentation, and whether our models are close to the achievable IoU in considering human variability.

Overall, Mask R-CNN remains the most promising method in automatic nuclei segmentation based on this project. U-Net also outperformed traditional computer vision connected region extraction and more methods may be helpful in improving the performance of these models.

Citations

[1] Gustafsdottir, S., Ljosa, V., Sokolnicki, K., Anthony Wilson, J., Walpita, D., Kemp, M., Petri Seiler, K., Carrel, H., Golub, T., Schreiber, S., Clemons, P., Carpenter, A. and Shamji, A. (2013). Multiplex Cytological Profiling Assay to Measure Diverse Cellular States. *PLoS ONE*, 8(12), p.e80999.

[2] Kaggle.com. (2018). *2018 Data Science Bowl | Kaggle*. [online] Available at: https://www.kaggle.com/c/data-science-bowl-2018 [Accessed 27 Feb. 2018].

[3] Scannell, J., Blanckley, A., Boldon, H. and Warrington, B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nature Reviews Drug Discovery*, 11(3), pp.191-200.
SIOPE - the European Society for Paediatric Oncology. (2018). *Rare Diseases - SIOPE - the European Society for Paediatric Oncology*. [online] Available at: https://www.siope.eu/activities/european-advocacy/rare-diseases/ [Accessed 27 Feb. 2018].

[4] World Health Organization. (2018). *The top 10 causes of death*. [online] Available at: http://www.who.int/mediacentre/factsheets/fs310/en/ [Accessed 27 Feb. 2018].

[5] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, Jan. 1979. doi: 10.1109/TSMC.1979.4310076
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4310076&isnumber=4310064

[6] O. Ronneberger, P. Fischer, and T. Brox, "*U-net: Convolutional networks for biomedical image segmentation*," in MICCAI, pp. 234– 241, Springer, 2015.

[7] Ross Girshick. "Fast R-CNN." In Proc. IEEE Intl. Conf. on computer vision, pp. 1440-1448. 2015.

[8] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. "Faster R-CNN: Towards real-time object detection with region proposal networks." In Advances in neural information processing systems (NIPS), pp. 91-99. 2015.

[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask R-CNN." arXiv preprint arXiv:1703.06870, 2017.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region- based convolutional networks for accurate object detection and segmentation. *TPAMI*, 2015.