

**Pune Institute of Computer Technology
Dhankawadi, Pune**

**A SEMINAR REPORT
ON**

SEGMENTATION IN MICROSCOPY IMAGES

SUBMITTED BY

PRANAV J. KULKARNI

Roll No. 31432

Class TE

**Under the guidance of
Prof. G. V. Kale**



**DEPARTMENT OF COMPUTER ENGINEERING
Academic Year 2019-20**



DEPARTMENT OF COMPUTER ENGINEERING
Pune Institute of Computer Technology
Dhankawadi, Pune-43

CERTIFICATE

This is to certify that the Seminar report entitled
“SEGMENTATION IN MICROSCOPY IMAGES”

Submitted by
PRANAV J. KULKARNI Roll No. 31432

has satisfactorily completed a seminar report under the guidance of
Prof. G. V. Kale towards the partial fulfillment of third year
Computer Engineering Semester II, Academic Year 2019-20 of
Savitribai Phule Pune University.

Prof. G. V. Kale
Internal Guide

Prof. M.S.Takalikar
Head
Department of Computer Engineering

Place:
Date:

ACKNOWLEDGEMENT

I sincerely thank our Seminar Coordinator Prof. B.D.Zope and Head of Department Prof. M.S.Takalikar for their support.

I also sincerely convey my gratitude to my guide Prof. G. V. Kale, Department of Computer Engineering for her constant support, providing all the help, motivation and encouragement from beginning till end to make this seminar a grand success.

It was fun working on this short Seminar and I am glad that I got to learn a variety of things in the field of Deep Learning.

Contents

1	INTRODUCTION	1
2	MOTIVATION	3
3	A SURVEY ON PAPERS	4
3.1	Convolutional neural networks: an overview and application in radiology [4]	4
3.2	U-Net: Convolutional Networks for Biomedical Image Segmentation [5]	4
3.3	Mask R-CNN [3]	4
3.4	Nuclear Segmentation in Microscope Cell Images: A hand segmented dataset and comparison of algorithms [2]	5
3.5	Mask-RCNN and U-Net Ensembled for Nuclei Segmentation [1]	5
4	PROBLEM DEFINITION AND SCOPE	6
4.1	Problem Definition	6
4.2	Scope	6
5	METHODOLOGY	7
5.1	U-Net	7
5.1.1	Architecture	7
5.1.2	Implementation	7
5.2	Mask-RCNN	8
5.3	Ensemble model	9
6	EXPERIMENTAL DETAILS	10
6.1	Data	10
6.2	Training	10
6.3	Post Processing	10
6.4	Evaluation	11
7	RESULTS	12
7.1	Metrics	12
7.1.1	IoU	12
7.1.2	Dice score	12
7.1.3	Precision	12
7.1.4	maP	12
7.1.5	Recall	12
7.2	Results	13
7.3	Result Analysis	14
7.4	Results of implemented module	16
8	CONCLUSION	17
	References	18

List of Tables

- | | | |
|---|----------------------------------------------------------------------|----|
| 1 | Ref.[1] Overview of the performance of the models at IoU of 0.7. . . | 13 |
|---|----------------------------------------------------------------------|----|

List of Figures

- | | | |
|---|-------------------------------------------------------------------------------------------------|----|
| 1 | Ref.[5] General U-net architecture (example for 32x32 pixels in the lowest resolution). | 7 |
| 2 | Ref.[1] Inputs to U-Net and training data. | 8 |
| 3 | Ref.[3] Mask-RCNN Head Architecture | 9 |
| 4 | Ref.[1] Image modalities | 10 |
| 5 | Ref.[1] Average Precision curve | 13 |
| 6 | Ref.[1] Recall in different scenarios | 14 |
| 7 | Ref.[1] Models' prediction on an image | 15 |
| 8 | Predictions from implemented module. | 16 |

Abstract

Deep Learning involves using Machine Learning concepts for solving many problems. Traditionally, deep learning is widely used in computer vision for various tasks like classification, detection, segmentation, etc.

Automatic segmentation of microscopy images is a very important task in medical image processing and analysis. Cell nuclei segmentation is the most crucial step toward the implementation of a computer-aided diagnosis system for cancer cells. Nuclei segmentation is very important task which can be considered very ideal for modern CV (Computer Vision) methods. The process of segmentation in microscopy images can be easily simplified and automated with the help of computer vision. Although the recent developments in technology have made the implementation easy, proper expertise is still required for training setup.

As per [1], two popular segmentation frameworks, U-net and Mask-RCNN, are implemented in the cell nuclei segmentation task and compared to study their strengths and failures. To further improve their performance, an ensemble model which combines predictions from both of them is also studied.

Keywords

Image Segmentation, Nuclei Segmentation, Faster-RCNN, MaskRCNN, U-Net, CNN, Ensemble.

1 INTRODUCTION

Today, artificial intelligence (AI) is a thriving field with many practical applications and active research topics. The computers are apparently very smart in accomplishing formal tasks, may be solving math problem and returning the solution much more efficiently than a human being. But the greatest challenges for AI is to make a computer acquire intelligence and do tasks using informal knowledge - tasks such as recognizing spoken words, faces, images etc. and, to retrieve meaningful information from the same.

A machine learning algorithm is an algorithm that is able to learn from data. Mitchell has given the most famous definition of Machine Learning : “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ”.

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance.

Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do. In computer vision, image segmentation is the process of partitioning a image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Advances in computer vision algorithms can often be applied to a wide variety of fields, including biomedical imaging. Image segmentation is currently being widely used in biomedical analysis by processing microscopy images.

Medical image segmentation [2] has an essential role in computer-aided diagnosis systems in different applications. The vast investment and development of medical imaging modalities such as X-ray, microscopy, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI) etc. facilitated the creation of vast resources for implementing new medical image-processing algorithms. Image segmentation is considered the most essential medical imaging process as it extracts the regions of interest through a semiautomatic or automatic process. As segmentation divides the image into variety of coherent regions, variety of clustering methods can be applied for segmentation to separate the regions of interest from the background.

Neural networks with sufficient training data perform significantly better than traditional systems and with the constantly increasing amount of computational power and, free and open source software libraries, their day-to-day use in laboratories have become viable.

This study [1] throws light on two popular segmentation frameworks, U-net and Mask-RCNN, comparing their strengths and failures. To further improve their performance, an ensemble model which combines predictions from both of them is also studied.

2 MOTIVATION

Traditional rule based systems are hand designed system by experts in a particular field, and are crafted for solving a very specific problem. Machine Learning provides a general solution to many problems. Instead of hard coding a solution, machine learning allows computer to learn from data and use the experience to solve similar problems.

Fifty percent of cancer patients have a high possibility of developing malignant pleural effusion. For the assessment of malignancy, a cytological examination is performed by pathologists because it is cheap, less invasive, simple, and highly effective.

The cytological exam is a manual process wherein cytologists visually observe every single cell on the cytology glass slides using a microscopic camera, identify any abnormality in a cell, and finally make a decision. However, it is a very time-consuming process and requires good amount of skill. In addition, the diagnosis accuracy strongly depends on the attention and expertise of cytologists.

To implement a CAD (Computer Aided Diagnosis) system for cancer cells, cell nuclei segmentation is a prerequisite because cancer cells are largely observed on the basis of the morphological changes in the cell nuclei. Therefore, it is very crucial to select an accurate and effective cell nuclei segmentation method.

3 A SURVEY ON PAPERS

3.1 Convolutional neural networks: an overview and application in radiology [4]

Presently, Neural Networks is one of the most popular machine learning algorithms. It is already proven over time that neural networks outperform other algorithms in accuracy and speed. **Convolutional Neural Networks (CNN)** is one of the variants of neural networks used heavily in the field of Image processing. The hidden layers of a CNN typically consist of convolutional layers, fully connected layers, normalization layers and pooling layers.

CNN can learn various features by using multiple building blocks (pooling layers, convolution layers etc) through a process called back propagation. This article discusses the basics of convolutional neural networks and its various radiology applications, and its obstacles and future scope in the field of radiology. The challenges such as small dataset and overfitting, are also explored in this article, as well as their minimization techniques. The article also discusses application of CNN in radiology such as classification, segmentation, detection etc.

3.2 U-Net: Convolutional Networks for Biomedical Image Segmentation [5]

In the last years, deep convolutional networks have outperformed the state-of-the-art in many visual recognition tasks. But, their success was limited due to the size of the available training data and the size of the considered networks. Also, in many visual tasks, especially in biomedical image processing, the desired output should include localization, i.e., each pixel is supposed to be assigned a specific class label. Moreover, thousands of training images are usually beyond reach in biomedical tasks.

In this paper, a network and training strategy is presented, which strongly depends upon the data augmentations, to make efficient and effective use of available annotated samples. The architecture includes a contracting path for capturing the context and a symmetric expanding path that helps precise localization. It is shown that training of such a network can be done using a very few images and still outperformed the then-prior best method on the ISBI challenge for neuronal structure segmentation in electron microscopic stacks.

3.3 Mask R-CNN [3]

Instance segmentation includes correctly detecting all objects in an image and also segmenting each instance at the same time. So, it is a very challenging task. It therefore combines object detection, where the aim is classification of individual objects and localization using bounding box, and semantic segmentation, where the aim is pixel classification without instance differentiation.

This paper presented a object instance segmentation framework which was conceptually simple and a lot flexible. It efficiently detected objects in an image and generated a quality segmentation mask for instances at the same time. The method (Mask R-CNN) extends Faster R-CNN by adding a branch for object mask

prediction in parallel with the existing branch for recognizing bounding boxes. It is a lot simpler to train and only a small overhead is added to Faster R-CNN.

3.4 Nuclear Segmentation in Microscope Cell Images: A hand segmented dataset and comparison of algorithms [2]

Nuclei segmentation is a very crucial step in the process of many cytometric analyses. It acts as base for various operations (cell-cycle assignment, cell counting, etc.) and is often considered as the initial step in cell segmentation. A dataset of hand-segmented fluorescence microscopy images was prepared for objectively evaluating nuclei segmentation algorithms. Some published algorithms were evaluated for this problem on the prepared dataset.

This paper describes the dataset required for nuclear segmentation. It also describes numerous traditional approaches for biomedical image segmentation such as thresholding, active masks, seeded watershed, etc.

3.5 Mask-RCNN and U-Net Ensembled for Nuclei Segmentation [1]

Neural networks with sufficient training data perform significantly better than traditional systems and with the constantly increasing amount of computational power and, free and open source software libraries, their day-to-day use in laboratories have become viable.

This paper compared two well known segmentation frameworks, U-Net and Mask-RCNN. These frameworks are evaluated so as to find where they excel and fail. Each of them has their own pros and cons. To further improve their performance, an ensemble model which combines predictions from both of them is also studied.

4 PROBLEM DEFINITION AND SCOPE

4.1 Problem Definition

To study various segmentation frameworks and approaches for nuclei segmentation in microscopy images.

4.2 Scope

Nuclei segmentation is a very crucial and essential step in various cytometric analyses. Segmenting nuclei is also the initial step in cell segmentation. Segmenting overlapping nuclei is a very challenging task in nuclei segmentation in microscopy images.

We look at approach of [1] for nuclei segmentation in microscopy images. The authors of this paper have tried to use various segmentation frameworks for nuclei segmentation and compared them using various evaluation metrics. The comparison of frameworks and identifying their pros and cons is specifically interesting and worth studying. We attempt to do the same.

5 METHODOLOGY

5.1 U-Net

5.1.1 Architecture

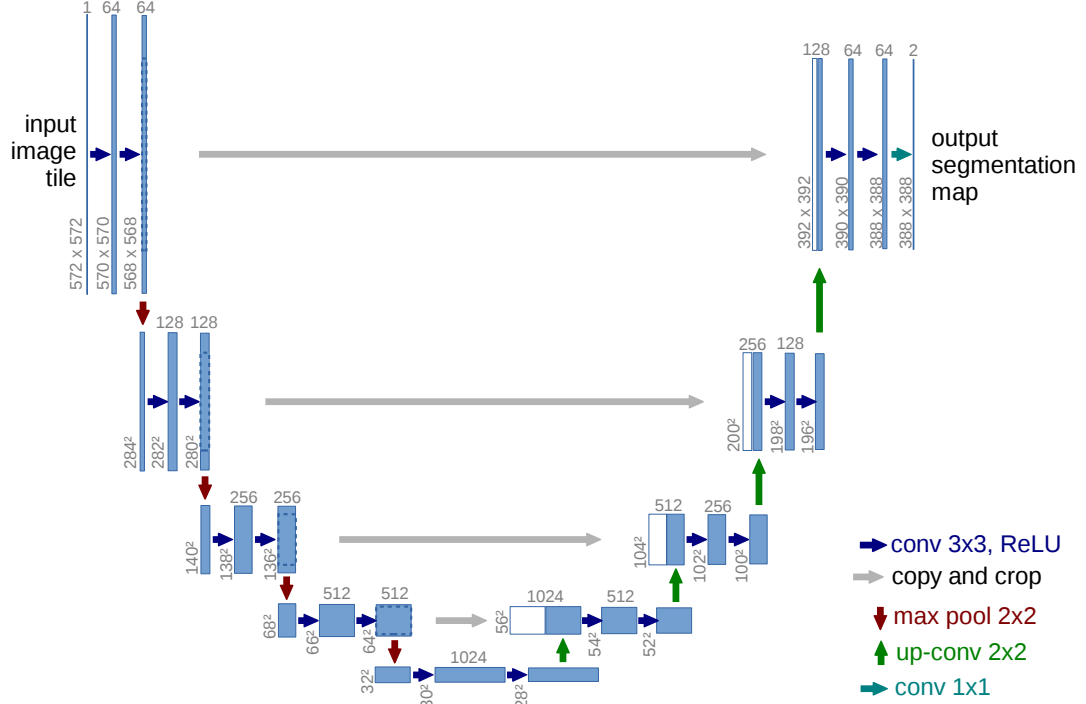


Figure 1: Ref.[5] General U-net architecture (example for 32x32 pixels in the lowest resolution).

Each blue box denotes a multi-channel feature map. The number of channels is given at the top of the box. The x-y-size is indicated at the lower left edge of the box. Copied feature maps are denoted by white boxes. The different operations are represented by arrows.

5.1.2 Implementation

U-Net [5] is a convolutional network that uses skip-connections to preserve features at a range of different resolutions. It is called U-net due to U-shape network architecture. The basic model uses a simple downsampling path, which can also be replaced with a deeper network such as ResNet. This enables the model to learn more complex features as the network depth can be increased using residual blocks. The interesting thing is U-net can also be used when relatively less amount of data is available for training by initializing the network with pretrained ResNet networks. With this the network's learning capacity is effectively increased. In this study [1], the performance accomplished by using ResNet101 initialized with weights from a pretrained ImageNet network was the best.

Instance segmentation is relatively difficult with U-Net, as the output is a binary segmentation mask for the whole input image. There are many solutions that can be used for the same, such as weighting border pixels heavily in the loss

function. The original paper [5] used the same. On top of this, [1] an additional output channel is added that predicts the borders between nuclei, which can later be subtracted from the segmentation output channel. This additional channel also helps the model to learn more regarding the nuclei shapes.

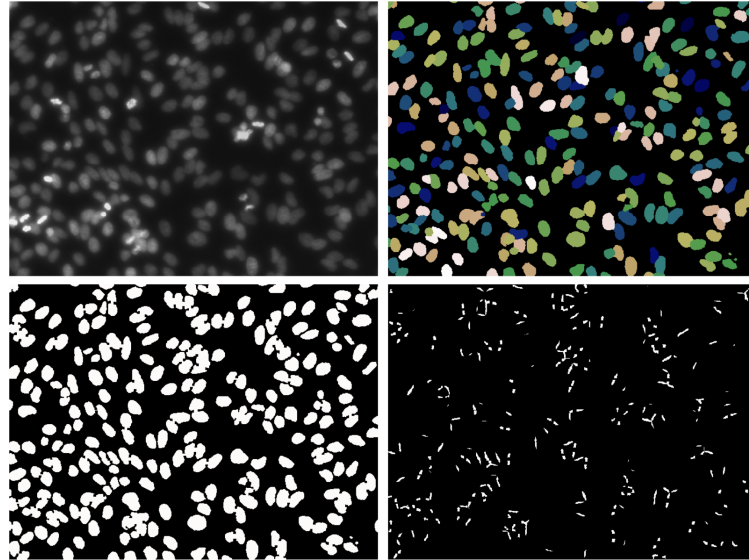


Figure 2: Ref.[1] Inputs to U-Net and training data.

Top left: Input image. Top right: GT segmentation mask. Each of the nuclei is a separate instance. Lower left: Segmentation mask target for the U-net framework. Lower right: Overlapping border mask target for U-Net.

5.2 Mask-RCNN

Another well known method for object segmentation is to make use of Mask-RCNN [3]. It is primarily designed to directly resolve the problem of instance segmentation and the efforts can then be channeled to tweaking the numerous hyperparameters of the network.

Initially, bounding boxes are predicted by the model for the nuclei. Then, segmentation is carried out inside each bounding box. While the model is able to predict the bounding boxes accurately, its performance on segmentation seems lower than that of U-Net. This can be seen in the results [1], it is better at detecting nuclei but segmentation seems poorer.

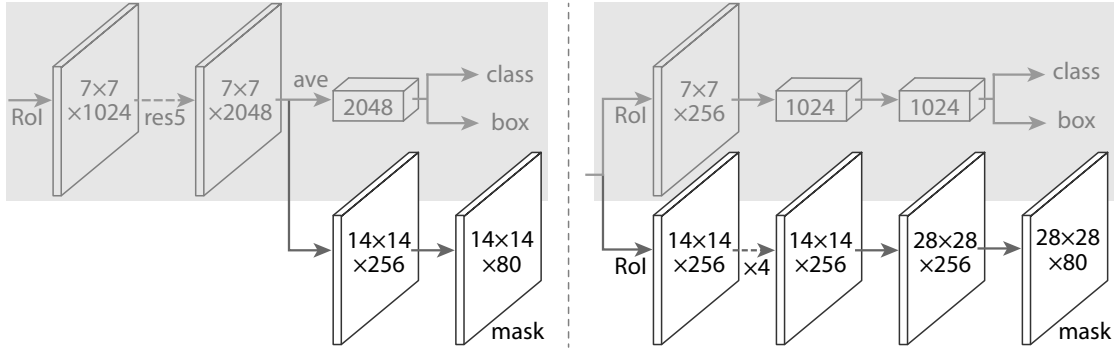


Figure 3: Ref.[3] Mask-RCNN Head Architecture

Two existing Faster R-CNN heads are extended. Left/Right panels show the heads for the ResNet C4 and FPN backbones respectively, to which a mask branch is added. Numbers denote spatial resolution and channels.

5.3 Ensemble model

As Mask-RCNN and U-Net frameworks work very differently, they have their own pros and cons. So, an ensemble approach was developed to combine their predictions. Features used were the structural properties of the prediction mask and the target was the intersection over union (IoU) with ground truth, and using this a gradient boosting model was trained on out of fold predictions of training data. The features used by it consisted of structural properties such as perimeter, convex area, eccentricity, solidity etc.

At the time of testing, nucleus mask predictions from both the models were provided as input to the ensemble model, which resulted in IoU with GT estimates for predictions of both the models. [1].

As a result, in case of overlapping masks, the mask which had highest IoU prediction was considered, while in case of non-overlapping masks, they were considered in output segmentation only if their predicted IoU was above the threshold.

6 EXPERIMENTAL DETAILS

6.1 Data

In [1], training data was gathered from a variety of sources, including a dataset from 2018 Kaggle Data Science Bowl and referencing numerous research papers, which consisted of various types of staining and microscopy techniques. This dataset consisting of both fluorescence and histology images, to the total of around 800 images and masks. Some images had a very low signal to noise ratio while some were out of focus and contained various background structures. Different image modalities are displayed in Fig. 4.

To increase the variance in training data, different augmentations were used. In this task, augmentations such as CLAHE, shearing, added noise and elastic deformations were unable to work as effectively. Consequently, only simpler augmentations such as rotations, random flips, scaling and shifts were used. Image contrast transfer was used with histology images. Images had dimensions of 256x256 px.

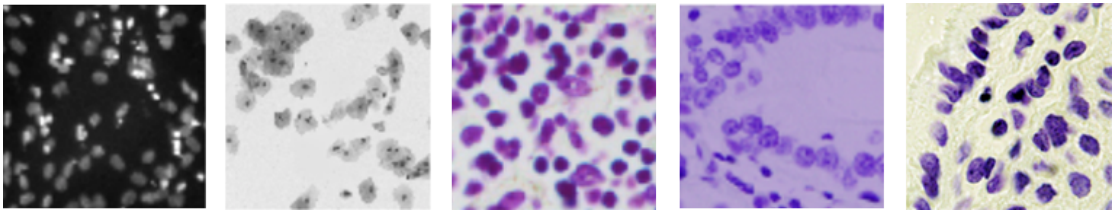


Figure 4: Ref.[1] Image modalities

From left: Fluorescence, brightfield and three different histology images.

6.2 Training

Four models were trained using 4-fold cross validation and the results were averaged while evaluating. Balancing of different image classes was done in each fold. [1] Model initialization was done with ResNet101 backbone, which was pretrained with ImageNet network.

6.3 Post Processing

Relatively small masks (less than around 10 pixels in area) were removed from the predictions and the holes in the segmentations were filled by using various morphological operations. To improve U-Net's predictions, final segmentation masks were extracted using a watershed based post processing method. The watershed was restricted to regions inside the predicted mask, using the border subtracted prediction mask as a marker and then the mask prediction was given as input to the algorithm. The segmentation masks from Mask-RCNN may overlap. So, in case of instances where this happened, using the distance from each instance the common region was splitted.

6.4 Evaluation

The kaggle competition (2018 datascience bowl) used the mean average precision as the metric. mAP at different thresholds of the IoU between GT and prediction was used. Here, the same metric along with Dice coefficient, precision, recall, under and over segmentation were used. If the model clumps or clusters multiple nuclei it is called under segmentation. And if the model segments a single nucleus into multiple nuclei is called over segmentation.

7 RESULTS

7.1 Metrics

We first try to understand various metrics used.

7.1.1 IoU

IoU (Intersection over Union) is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. Ranges from 0 to 1.

7.1.2 Dice score

Dice Coefficient is $2 \times$ the Area of Overlap divided by the total number of pixels in both images. Like the IoU, they both range from 0 to 1, with 1 signifying the greatest similarity between predicted and truth.

$$Dice = \frac{2 \times IoU}{1 + IoU}$$

7.1.3 Precision

Precision measures how accurate is your prediction. i.e. the percentage of predictions that are correct.

$$\text{Precision} = \frac{\text{True Positives}}{\text{Total Predicted Positives}}$$

7.1.4 maP

maP is the mean average Precision.

7.1.5 Recall

Recall refers to the percentage of total relevant results correctly predicted by the algorithm.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

7.2 Results

Overall	mAP	Dice	Precision	Recall	oseg	useg
U-Net	0.515	0.660	0.680	0.577	52	383
MRCNN	0.519	0.617	0.812	0.596	14	164
Ensemble	0.523	0.659	0.725	0.607	27	328
Fluorescence	mAP	Dice	Precision	Recall	oseg	useg
U-Net	0.564	0.708	0.733	0.643	38	301
MRCNN	0.569	0.684	0.841	0.663	13	132
Ensemble	0.570	0.703	0.767	0.664	23	266
Histology	mAP	Dice	Precision	Recall	oseg	useg
U-Net	0.298	0.285	0.502	0.385	14	82
MRCNN	0.300	0.236	0.698	0.405	1	32
Ensemble	0.316	0.289	0.586	0.442	4	62

Table 1: Ref.[1] Overview of the performance of the models at IoU of 0.7.
oseg indicates over segmentation and useg denotes under segmentation.

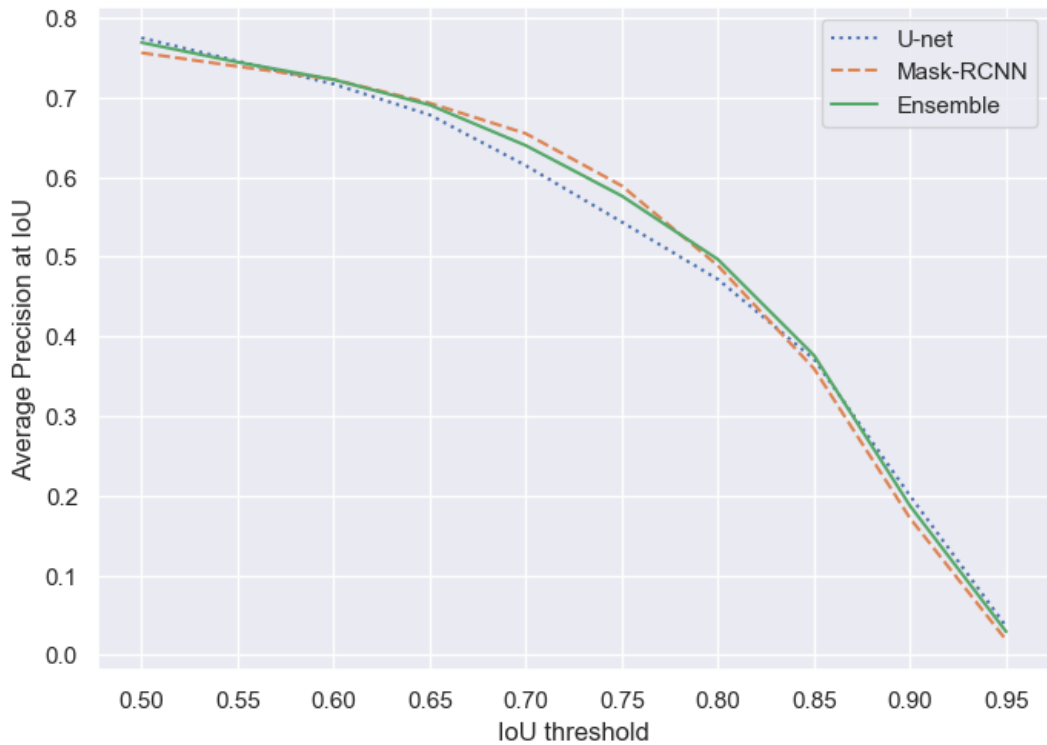


Figure 5: Ref.[1] Average Precision curve
Average precision is visualized at different IoU thresholds for the tested models.

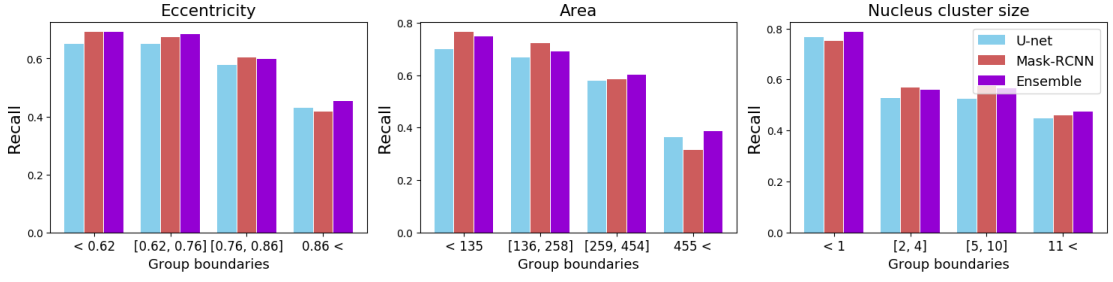


Figure 6: Ref.[1] Recall in different scenarios

Denotes the recall at 0.7 IoU in different scenarios for different models. Cluster size, area and eccentricity are compared. Each groups contain the same number of nuclei due to boundary setting.

Table 1 shows overall performance for the different models. Precision, recall and under and over segmentation were calculated at IoU threshold of 0.7 (which requires precise masks while errors limiting to only a few pixels are neglected).

7.3 Result Analysis

These results [1] highlight the differences in the Mask-RCNN and U-net’s differences. Even though the mAP values of both the frameworks seem to be similar, the other metrics show their pros and cons. U-Net’s better performance on the Dice coefficient score infers that it is able to predict precise segmentation masks, but at the same time number of detection errors increased. Mask-RCNN had comparatively worse Dice score, but better recall and precision, which indicates that nuclei are more accurately detected but the model finds it difficult to predict a good segmentation mask. Another interesting observation is the extent of under and over segmentation where Mask-RCNN has performed better than U-net. It is observed that Mask-RCNN is better at individual nuclei detection from a clump or cluster whereas U-Net has a tendency to cluster or clump them into one big nucleus.

The ensemble model aced the recall and mAP measure in all the scenarios. Fig. 5 plots the average precision at different IoU thresholds, from which it is observed that the ensemble model follows the upper bound of both models’ results very closely. U-Net faced some difficulty at the mid-range IoU thresholds, which can also be seen from the slightly decreased performance of the ensemble model. Mask-RCNN performed better at the higher and lower IoU thresholds where U-Net showed poorer performance. Good results were shown across all thresholds by the ensemble model which shows it does well at combining predictions by picking best ones.

Even though the ensemble model performed slightly worse on the precision metric, it fared well in terms of recall against both U-Net and Mask-RCNN. The ensemble model’s increase in recall score shows the ability of it to pick best predictions from both models leading to increased true positives.

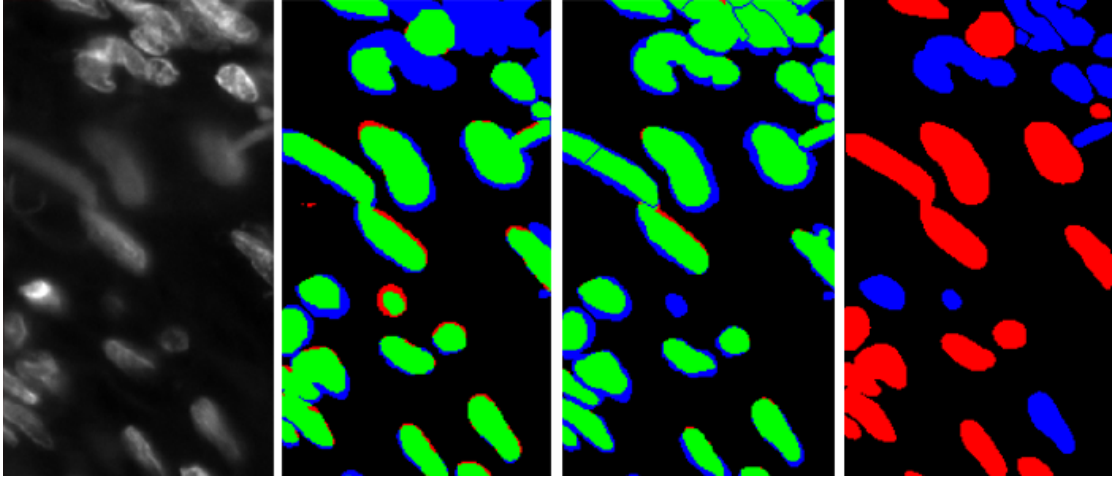


Figure 7: Ref.[1] Models' prediction on an image

First image is the network input, second image is the prediction of Mask-RCNN and third image is U-Net's prediction. Green pixels overlap with the ground truth, blue pixels are ground truth without any matching prediction and red pixels a prediction without overlapping ground truth. The fourth image shows the ensemble prediction (blue - U-Nets's predictions, red- Mask-RCNN's prediction).

Fig. 7 shows predictions on an input image case, where the predictions from both the models is combined using the ensemble model and the accuracy is increased. Combined result prediction is better than individual prediction of the models.

The results [1] infer that a single nucleus is easy to segment, but in case of multiple clustered nuclei, the network should learn to separate different nuclei instances. This can be easily seen in Fig.6 : Prediction quality deteriorates immediately when the nuclei are clustered. U-Net seemed to perform better when predicting a single lone nucleus, while Mask-RCNN's performance was slightly better on grouped or clumped nuclei. U-Net's poor performance on clumped nuclei is due to the output channel where even the smallest errors can result in merged masks. The performance of ensemble model was relatively better on all sizes of nuclei clusters, inferring that even when the nuclei are closely clumped together, it can effectively and accurately pick the best masks from both models.

7.4 Results of implemented module

For the scope of the seminar, a primitive model of U-Net implementation is implemented. Below is the sample prediction done with the help of the implemented model.

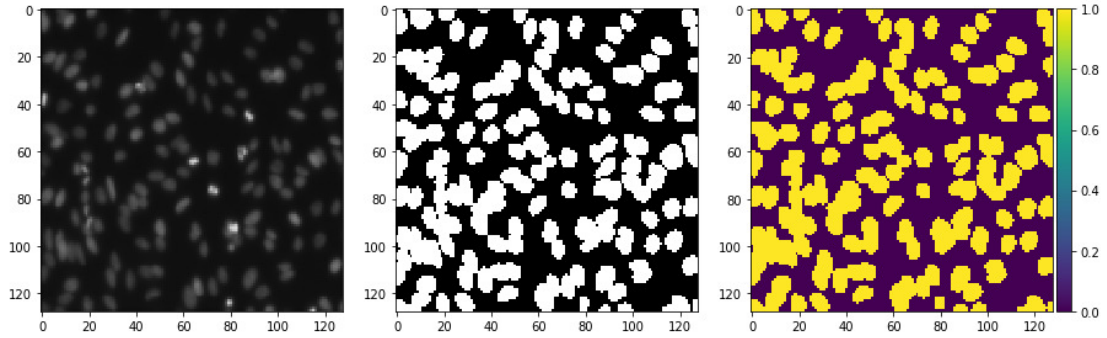


Figure 8: Predictions from implemented module.

The leftmost image is a microscopy image consisting of multiple cell nuclei. The figure in the middle is the ground truth. The rightmost image is the predicted segmentation mask for the image.

Here, we can observe that all the nuclei are very accurately segmented. In this case, performance deteriorates in the segmentation of clumped or overlapping nuclei. The model cannot precisely segment the clumped nuclei into separate segments. This drawback of U-Net framework is already discussed above.

8 CONCLUSION

We have looked at the basics of image segmentation in microscopy images with reference to its application in nuclei segmentation. We have studied various segmentation frameworks, like U-net and Mask-RCNN, and the ensemble technique.

It is observed that the frameworks have their own pros and cons in predicting the cell nuclei. U-net is able to create accurate segmentation masks but detection errors are increased. Whereas, mask-RCNN is more accurate in detecting nuclei but finds it quite difficult to predict a good segmentation mask. Many such pros and cons are studied.

Ensemble technique seemed to be a particularly interesting approach. It picks the best predictions from both the models (U-net and mask-RCNN) and performs better than them.

References

- [1] A. O. Vuola, S. U. Akram and J. Kannala, "Mask-RCNN and U-Net Ensembled for Nuclei Segmentation," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 2019, pp. 208-212.
- [2] L. P. Coelho, A. Shariff and R. F. Murphy, "Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms," 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, 2009, pp. 518-521.
- [3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask r-cnn," in The IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.
- [4] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9, 611–629 (2018). <https://doi.org/10.1007/s13244-018-0639-9>
- [5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention (MICCAI), pp. 234–241, 2015.

.

.