U-Net and Mask R-CNN for GI Tract Image Segmentation

Alina Chou, Wenqi Li, Edgar Roman

Department of Computer Science, Stanford University

Introduction

The problem that we are investigating is how to effectively track healthy organs in medical scans in order to improve cancer treatment. Cancer takes enough of a toll, and solving the problem will enable radiation oncologists to safely deliver higher doses of radiation to tumors while avoiding the stomach and intestines. This will make cancer patients' daily treatments faster and allow them to get more effective treatment with less side effects and better long-term cancer control.

Problem Statement

For this project, we are building a model to accurately segment the stomach and intestines on MRI scans. The model takes in inputs of MRI scan images from cancer patients by the UW-Madison Carbone Cancer Center. We then use U-Net and Mask R-CNN methods to obtain predicted segmented areas of patients' MRI scans with labels for "stomach", "large bowel", and "small bowel".

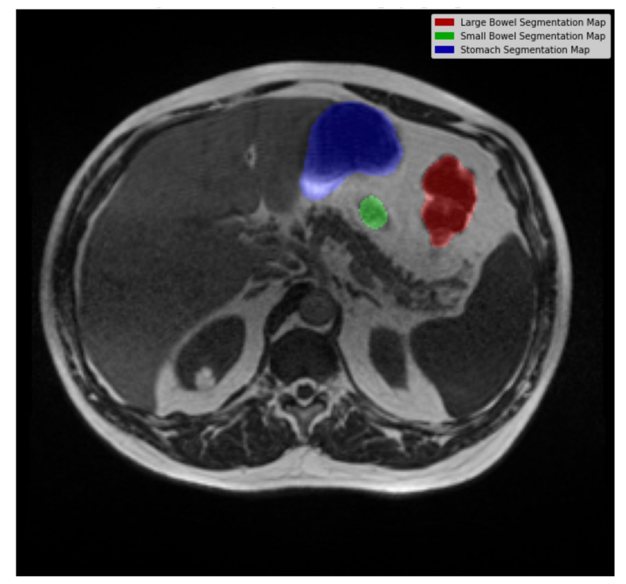
Datasets & Metrics

The data we use is anonymized MRIs of patients treated with MRI guided radiotherapy provided by the UW-Madison Carbone Cancer Center. The dataset contains 85 cases with 38496 scan slices of organs represented in 16-bit grayscale PNG format, and the annotations are provided in a csv format with the segmented areas represented as RLE-encoded masks.

We preprocess the data by:

- Obtaining images with sufficient visual information
- Standardizing image size & rescaling masks
- Decoding RLE-encoded masks and visualizing
- Processing & storing image, mask, and target as tensors for more efficient dataset loading
- Splitting data into train, validation, and test sets

Here is an example of the preprocessed MRI scan with mask overlay visualization:

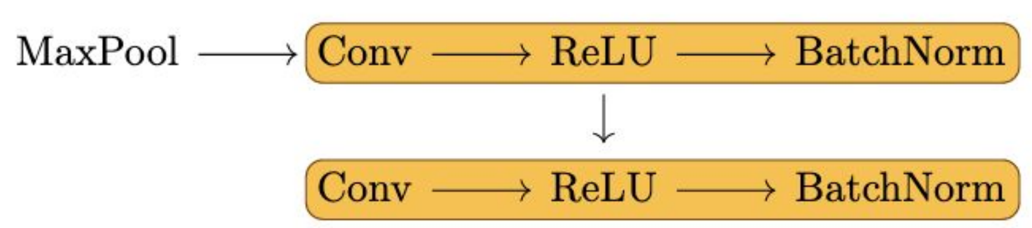


We use the Dice score as our evaluation metric, computed as follows:

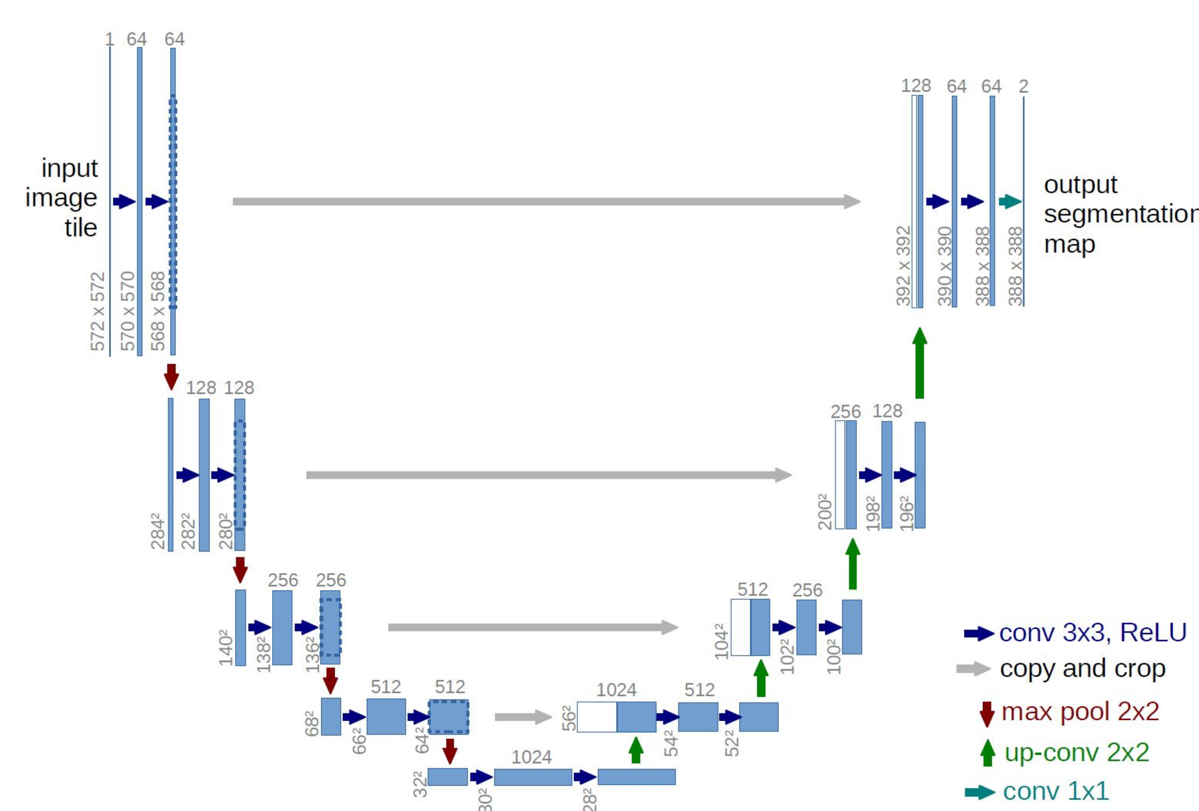
$$\frac{2|M_{\text{pred}} \cap M_{\text{true}}|}{|M_{\text{pred}}| + |M_{\text{true}}|}$$

Model I: U-Net

The UNet model consists of two main subnetworks: down and up. The down subnetwork consists of 4 growing step down layers:

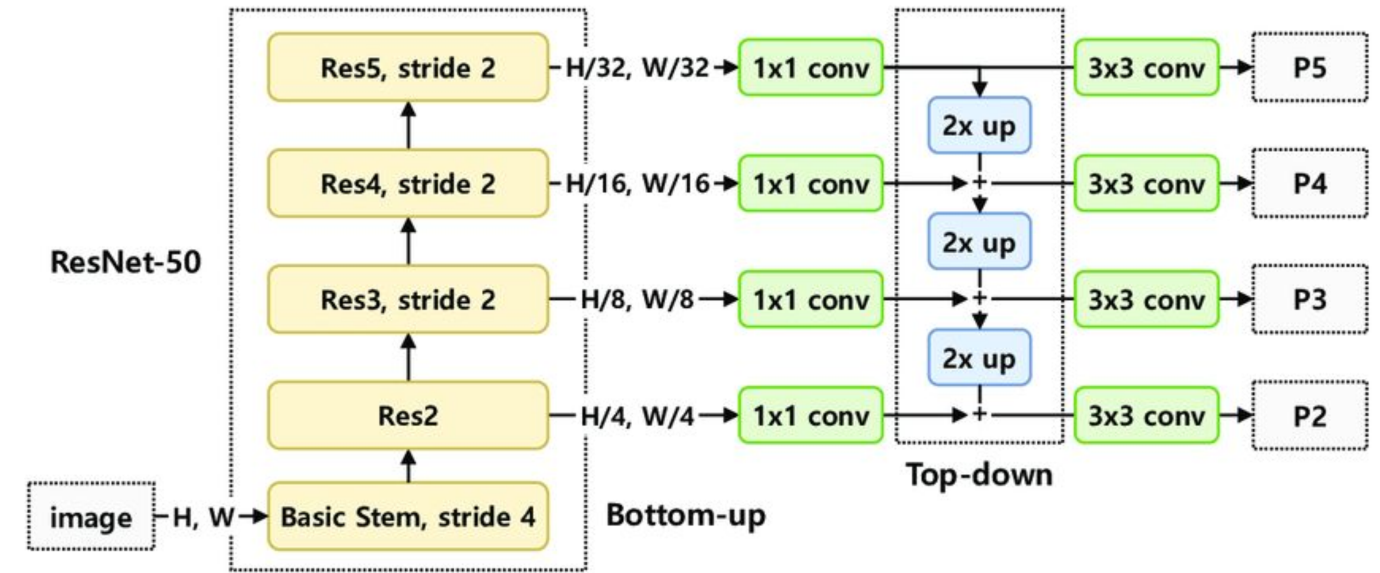


The up subnetwork consists of 4 corresponding step up layers, which are similar to the step down layers except MaxPool layer is replaced by a UpSample layer, and the convolutional layer is getting smaller. We added dropout layers between convolutional layers.

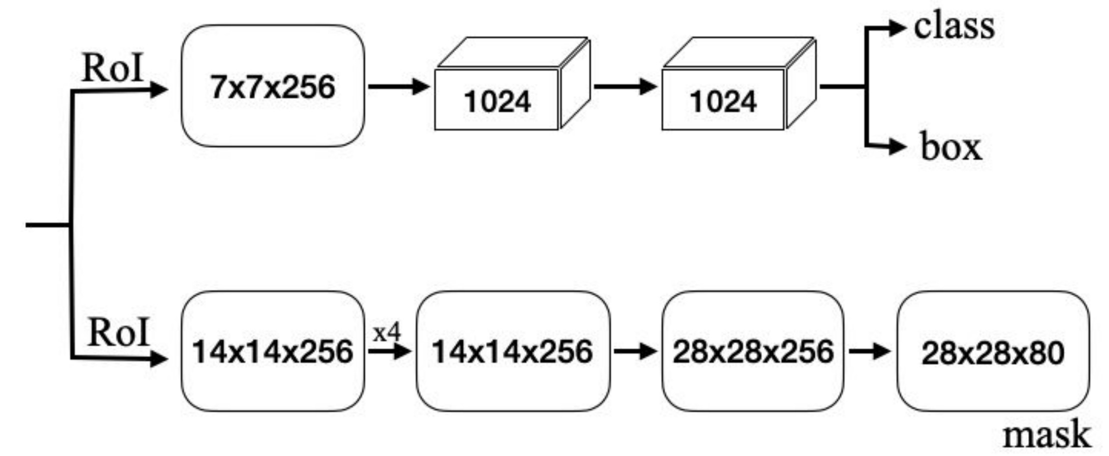


Model II: Mask R-CNN

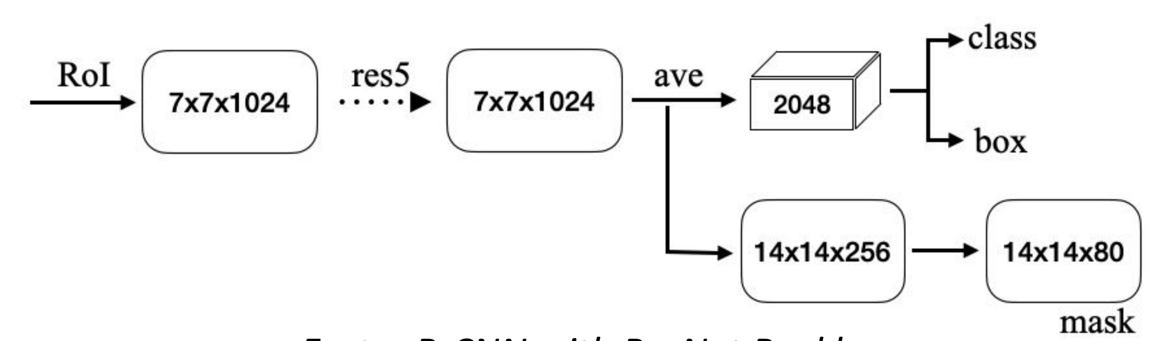
We use a ResNet-50-FPN as our backbone in the Mask R-CNN:



The head structures of the Mask R-CNN are:



Faster R-CNN with FPN Backbone



Faster R-CNN with ResNet Backbone

Experiments & Analysis

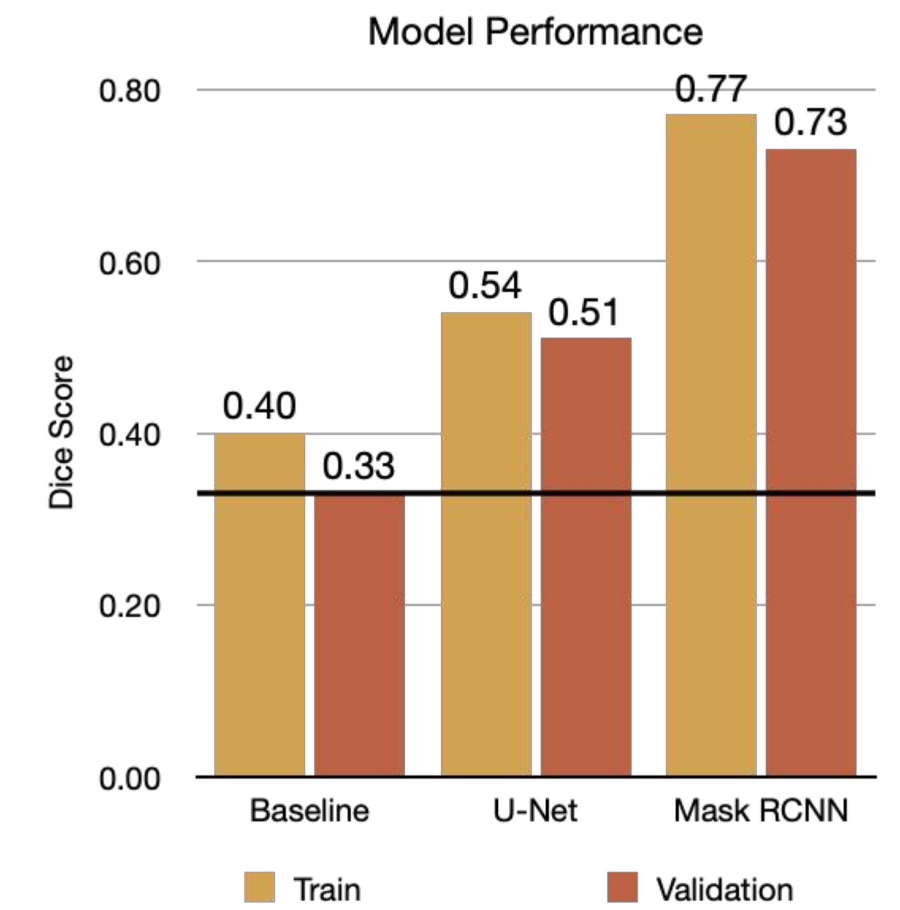
For U-Net, we experiment with the following hyperparameter combinations; bolded column results in the highest Dice score.

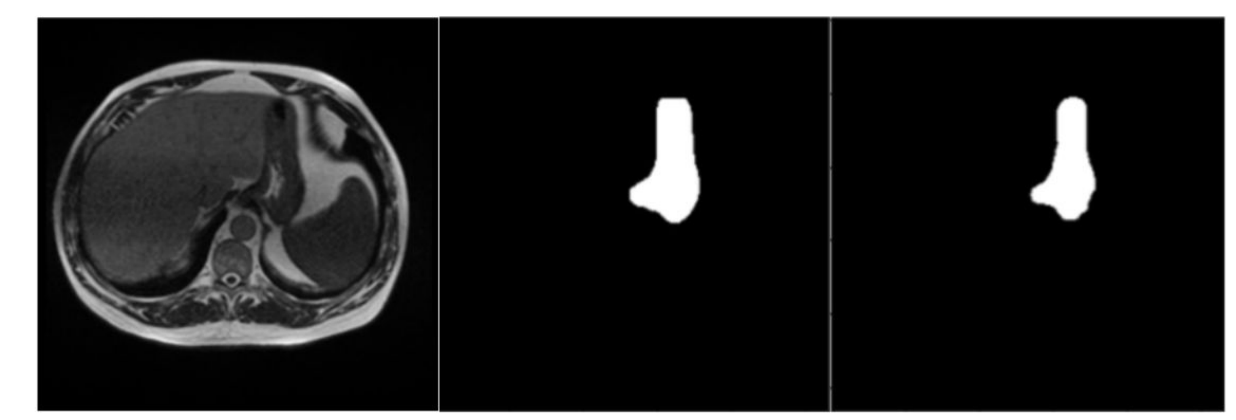
Parameter Name	Default	Comb. 1	Comb. 2	**Comb 3.**	Comb. 4	Comb. 5
batch size	1	32	32	**32**	64	64
learning rate	1e-5	1e-5	1e-4	**5e-5**	5e-5	1e-4
max epochs	15	15	15	**15**	15	15
scaling	0.5	0.5	0.5	**1**	0.5	1
depth	4	4	4	**3**	3	3

For Mask R-CNN, following is the best performing Mask R-CNN specific hyperparameter combination:

Parameter Name	Value
Backbone	ResNet-50-FPN
Train RoIs Per Image	64
Max GT Instances	3
Detection Min Confidence	0.7

Following are the Dice score comparison for each model after training & evaluation and visualized mask prediction result examples.





(a) Original MRI Scan (b) Predicted Stomach Mask (c) True Stomach Mask

Conclusion & Future Work

In this project, we implement and modify the U-Net model and the Mask R-CNN model, and apply them to the GI tract image segmentation problem. Our best U-Net model and Mask R-CNN model achieve Dice scores of 0.51 and 0.73 on the validation set respectively, which both are significant improvements compared to the baseline Dice score result of 0.33. Mask R-CNN performs better than the U-Net model with potential reason of higher complexity in loss function.

We hypothesize several future directions may further improve our results and network design:

- Formulate a more sophisticated loss function for U-Net
- Ensemble the U-Net architecture with Mask R-CNN, potentially using U-Net as the backbone
- Incorporate data augmentation procedure