Medical image segmentation for Prostate cancer samples

Problem:

Prostate cancer (PCa) is the second most common cancer among males worldwide.Diagnosis of PCa is based on the grading of prostate tissue biopsies. These tissue samples are examined by a pathologist and scored according to the Gleason grading system. However, the system suffers from significant inter-observer variability between pathologists, limiting its usefulness for individual patients. This variability in ratings could lead to unnecessary treatment, or worse, missing a severe diagnosis.Automated deep learning systems have shown some promise in accurately grading PCa. If we can improve those techniques and develop better models for detecting PCa on images of prostate tissue samples, and estimate severity of the disease using the most extensive multi-center dataset on Gleason grading . Patients can avoid risk of missing cancers or unnecessary treatment.

Each individual image of biopsies is quite large. It is very important to efficiently locate areas of concern to zoom in. In this project, we will generate masks, which are the segmentation (pixel wise location and class of data in the image) and attempt to predict masks for unseen (test) data.In other word we want to identify the location and sizes of different Pca in new biopsies samples

Data :

https://www.kaggle.com/c/prostate-cancer-grade-assessment/data

The dataset includes the original image, mask and csv files.

The image dimensions are quite large (typically between 5.000 and 40.000 pixels in both x and y).Each slide has 3 levels you can load, corresponding to a downsampling of 1, 4 and 16.Biopsies can be in different rotations and colors, which has no clinical value . We will use the data which already resized to a smaller dimensions

Almost each image has an associated mask with label information, which indicate which parts of tissue are healthy or cancerous, that will help on localization of samples.The mask values depend on data provider:

Radboud: Prostate glands are individually labelled, Valid values are:

0: background (non tissue) or unknown

1: stroma (connective tissue, non-epithelium tissue)

2: healthy (benign) epithelium

3: cancerous epithelium (Gleason 3)

4: cancerous epithelium (Gleason 4)

5: cancerous epithelium (Gleason 5)

Karolinska: Regions are labelled, Valid values are:

1: background (non tissue) or unknown

2: benign tissue (stroma and epithelium combined)

3: cancerous tissue (stroma and epithelium combined)

Cvs file provide image level label information ,data provider and so on

Data Wrestling & Analysis:

a) Missing Data

len(image)-len(mask)=100

There are 100 mask is missing ,since we have enough data, we will just simple delete those

b) Load,Transform and Normalize

Create customized dataset function to load image and mask and transform them to tensors

c) Image Visualization

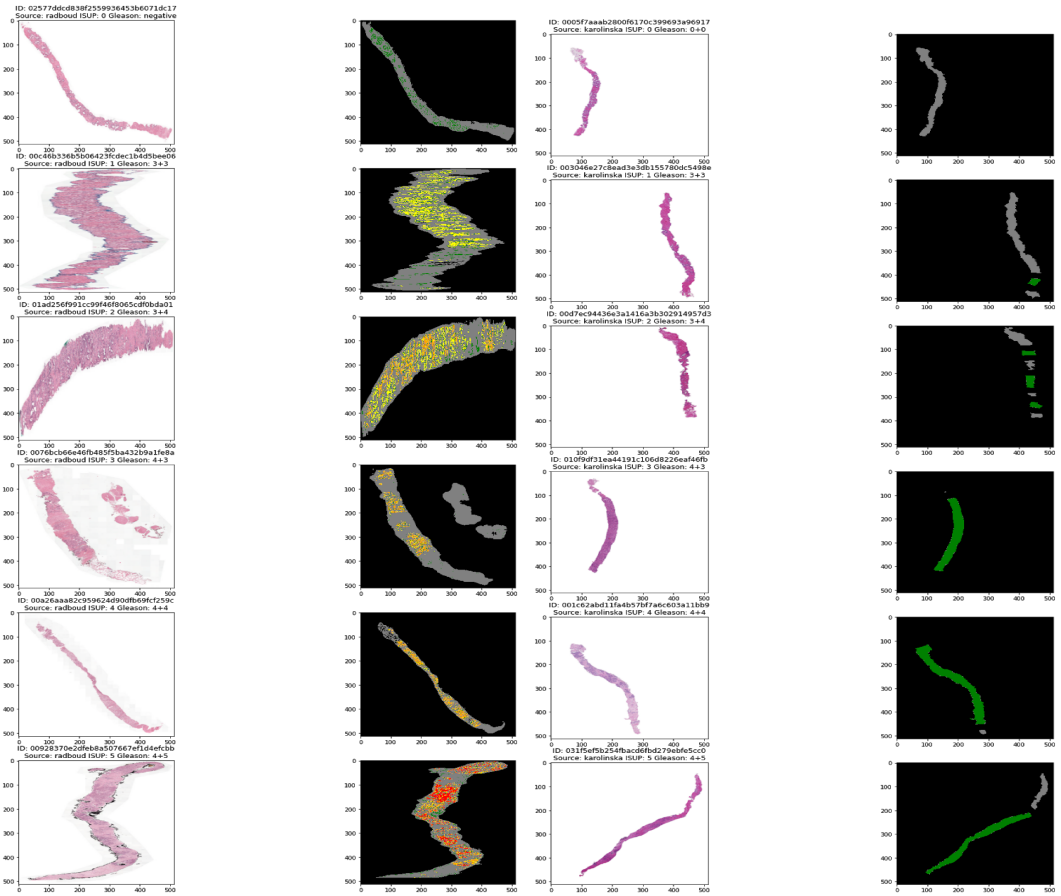
Use skimage to show the image file.

For the mask file I only take out the values of Red Level and apply color map assigning each label a distinct color. Since the label information is stored in the red (R) channel, the other channels are *set to zero.*

Below are some sample images and associated masks.

Left one is from r center, right one from k center, you can see that they grade it in different ways.

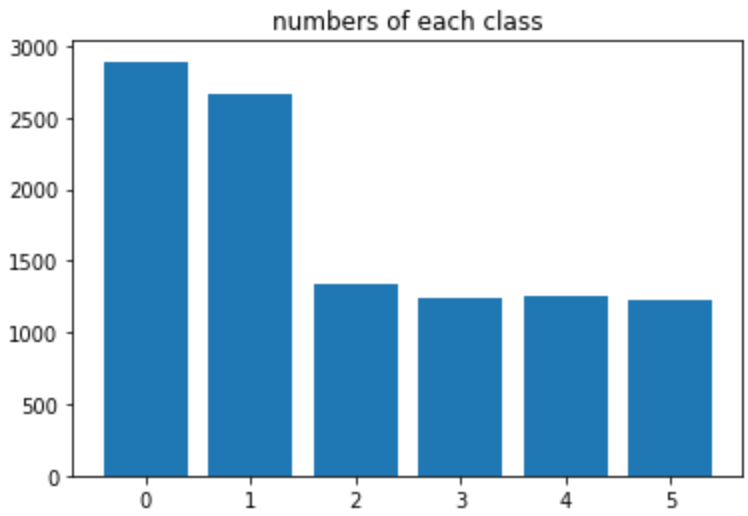
The class 6 sample from r center has the richest colors , that means even the cancer developed into the highest level ,It still has some healthy tissues,it may contain all level tissues.



Images and their corresponding masks from radboud center(left) and Karolinska center(right) for 6 different classes

d) Label Distribution

From the plot,we can see that the isup_grade 0 and 1 i.e no cancer, are way more than the rest of classes, in other words the multiclass are not balanced that is normal in most medical datasets, the cancer class will always be underrepresented and that's also the most important challenge when performing machine learning tasks on Medical DATA.



Summary: Our dataset is a large and unbalanced image dataset,It has two different label systems. We will use the ranboud center grading system, since it includes extra information on the severity of cancer. Unet might be a good semantic segmentation model for this, since it has the advantage of using fewer images to obtain good results,training the data using the highest level class sample(since those samples most likely include all 6 labels) might be good enough.

Image Segmentation

1)Semantic segmentation: label *each pixel* of an image with a corresponding ***class*** of what is being represented3).

2)U-net Structure:U-net is used in many image segmentation tasks for biomedical images,it has the advantage of requiring fewer training samples and Precise segmentation.The architecture contains two paths. First path is the contraction path (encoder) which is used to capture the context in the image. The encoder is just a traditional stack of convolutional and max pooling layers. The second path is the symmetric expanding path (decoder) which is used to enable precise localization using transposed convolutions.

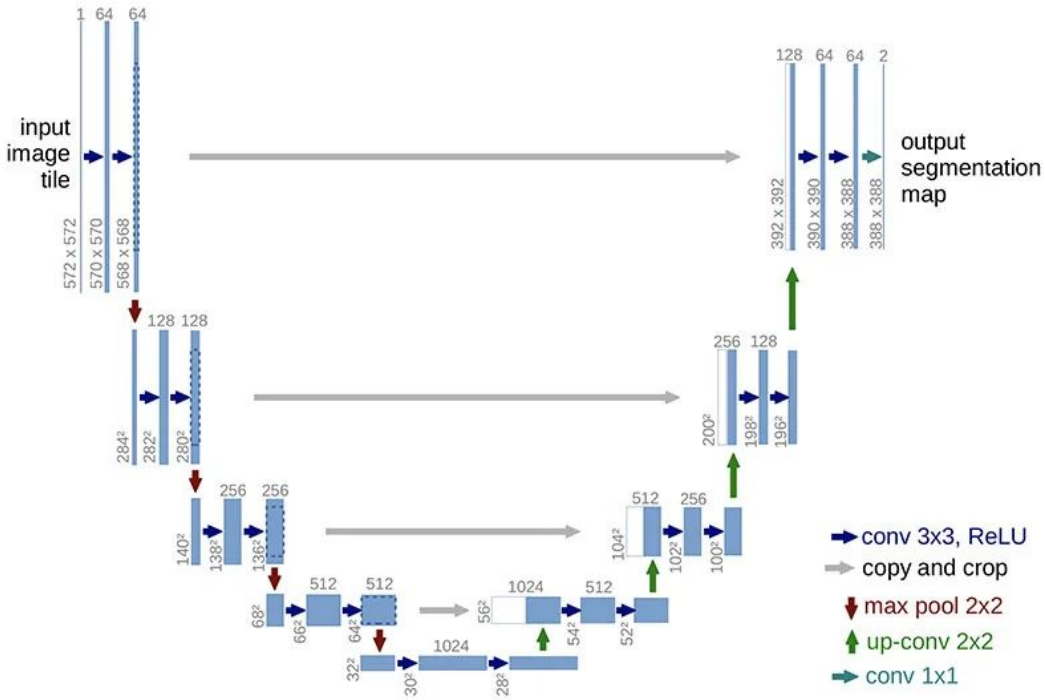
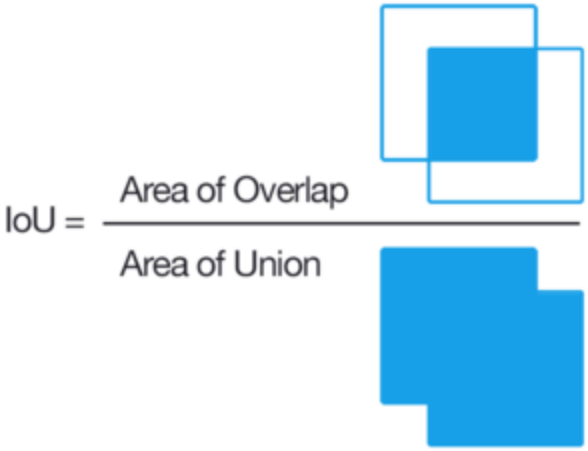


Fig 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. 1)

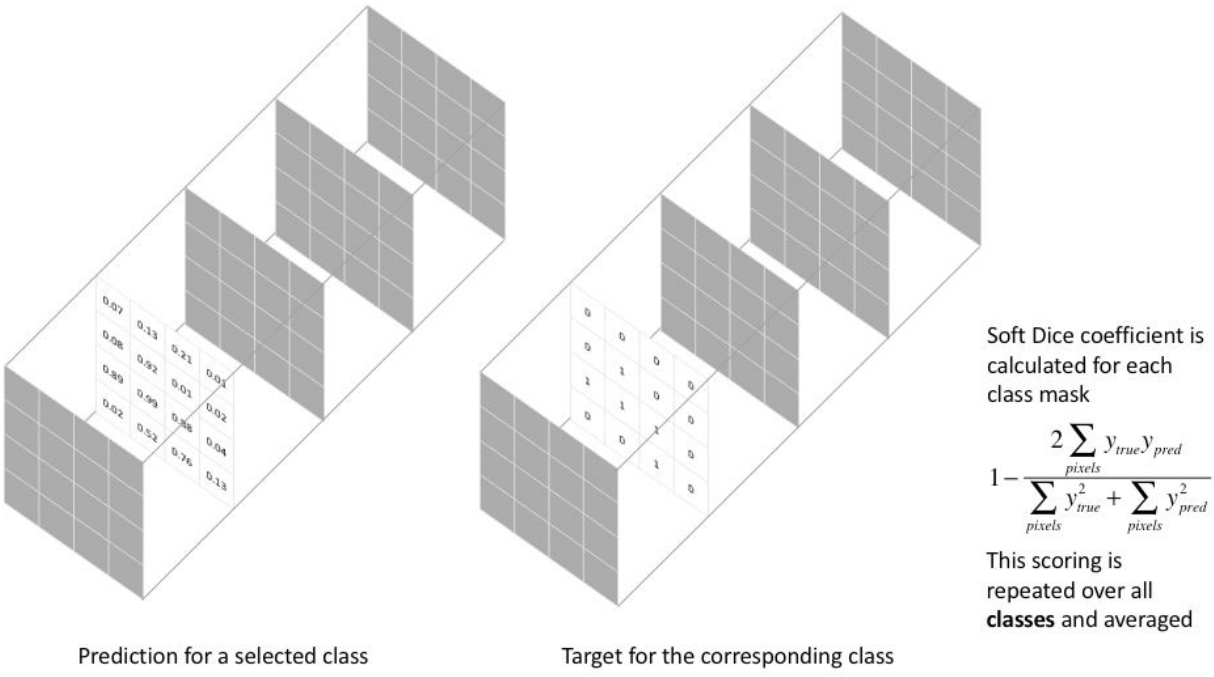
3) Evaluation Metrics

NLLLOSS:The negative log likelihood loss. It is useful to train a classification problem with multi classes. It works well for this dataset

IoU(Jaccard Index):The Intersection-Over-Union (IoU), also known as the Jaccard Index, is one of the most commonly used metrics in semantic segmentation.IoU 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,



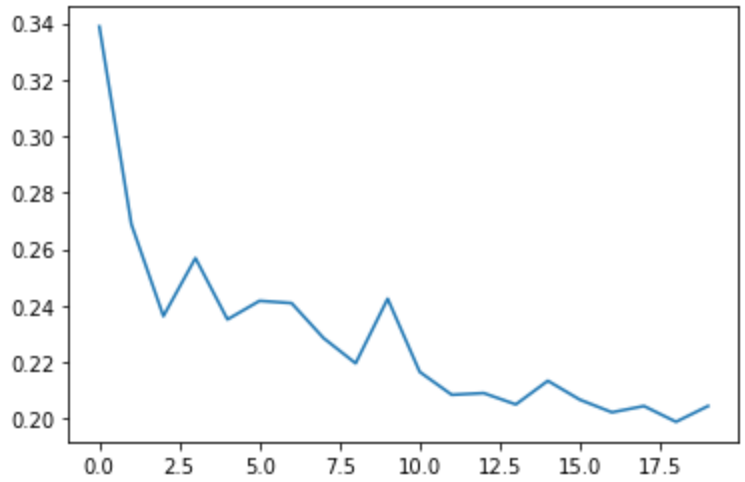
Soft Dice Coefficient : 2 * the Area of Overlap divided by the total number of pixels in both images. For multiclass2):



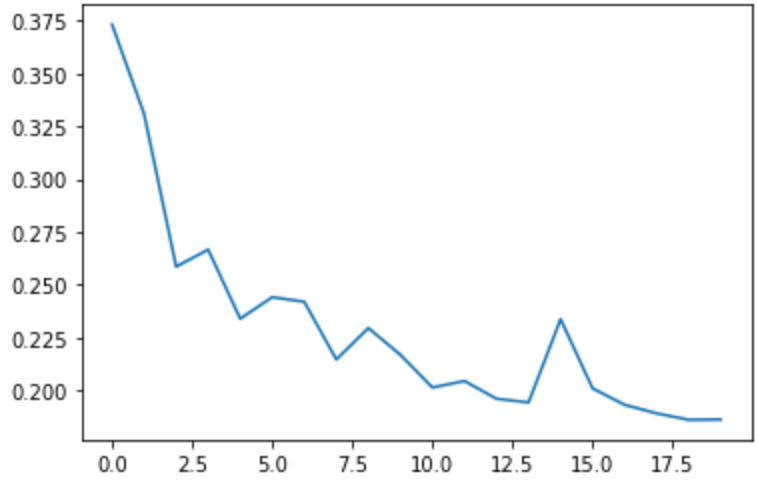
IoU and Soft Dice coefficient loss are quiet similar, we will only use Soft Dice coefficient as evaluation metrics for this project

As mentioned before we will use ranboud center grading system to train our model, we create two dataset grader(which only include data from r center) and its subset grade5(which only include level 5 data)

From the training Loss plots, we can see that actually they perform quite similarly ,grader(running_loss `0.21492987511807987`) performing is a little bit better than grade5(0.22823013762801225) for the valid set. That Is probably because the validate dataset we use is from the grader testset which is more similar to the grader training set, they has the same distribution.



a)20 Epoch of running loss for all classes dataset



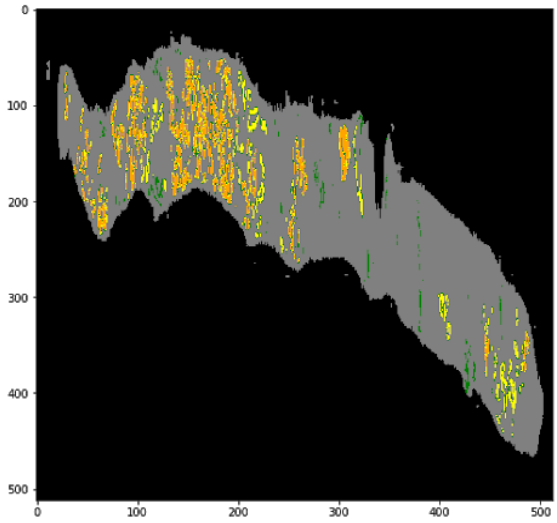
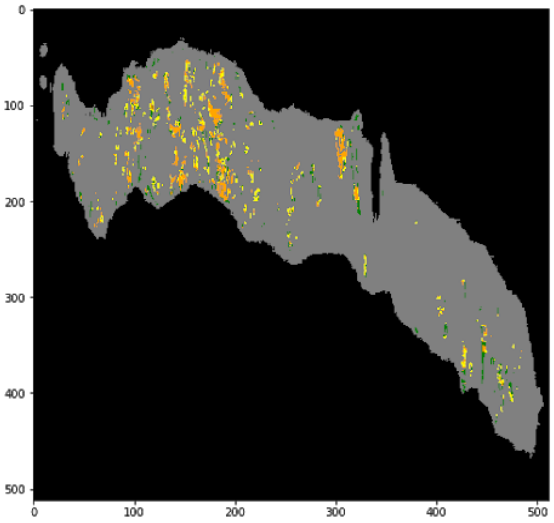
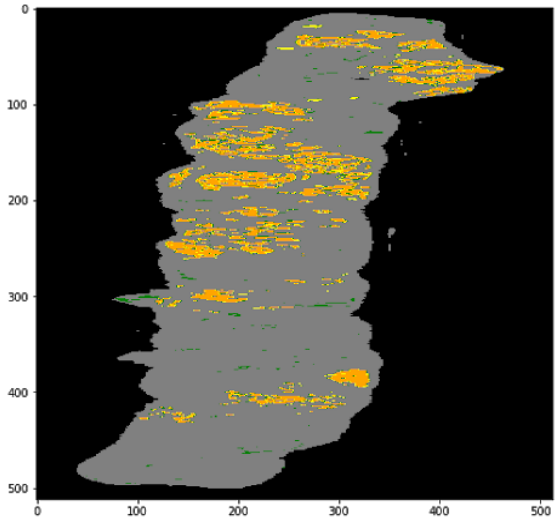
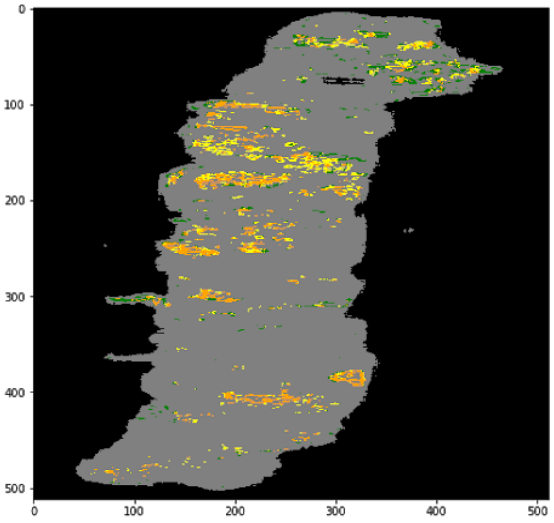
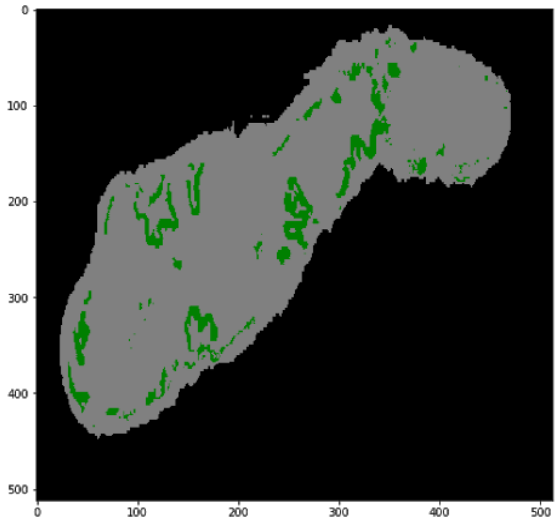
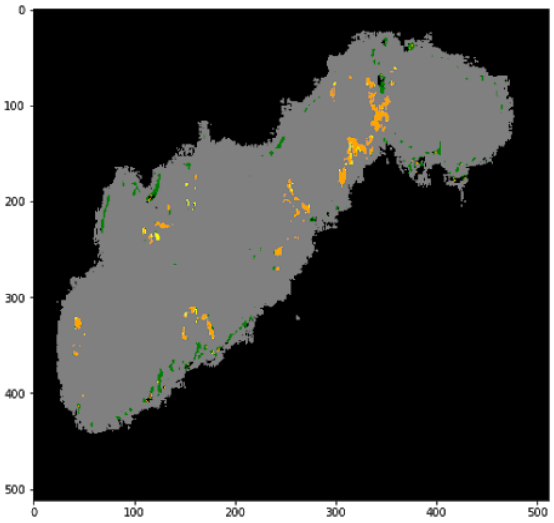
b)20 Epoch of running loss for class5 only dataset

Soft Dice loss vs NLLLOSS

NLLLOSS was used for both training and testing data in this project
Soft Dice loss was used as an extra evaluation metric for testing data only.
Comparing two results(`0.1933531503291718 vs 0.830966999828632`), you can see that even when NLLLOSS is pretty small , Soft Dice loss show a relatively large number, that is because SDL measure area of Overlap divided by the total number of pixels in both images. So if time allowed ,we better use SDL for future tunning.

4) **Visualization valid datasets**

Below we show the groudtruch mask and our mask after training 100 epcho on grader data use it to predict on our test data, you can see that they are quiet close

Left:ground truce label mask right : model predict mask

Final thought: 1)since we use resized dataset, if we use high resolution data , may get better results.
2) further measure of false positive and false negative pixels , as these are very critical in medical diagnosis.

1)Olaf Ronneberger, Philipp Fischer, and Thomas Brox ‘U-Net: Convolutional Networks for Biomedical Image Segmentation’
2)Eli steven, luka Antiga,Deep Learning with PyTorch
3)some medium souce add later