Detect Cancer in Gigapixel Pathology Images

ADL Course Project

Jing Qian (jq2282)

May 19, 2019

I. INTRODUCTION

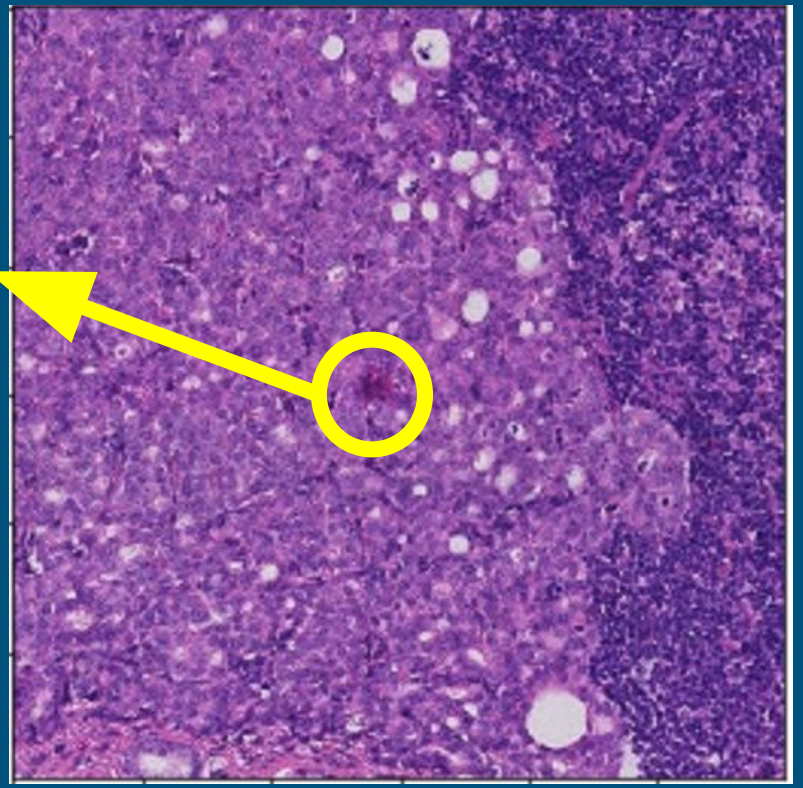
How do we detect the tumor?

Use **CNN**

High accuracy!

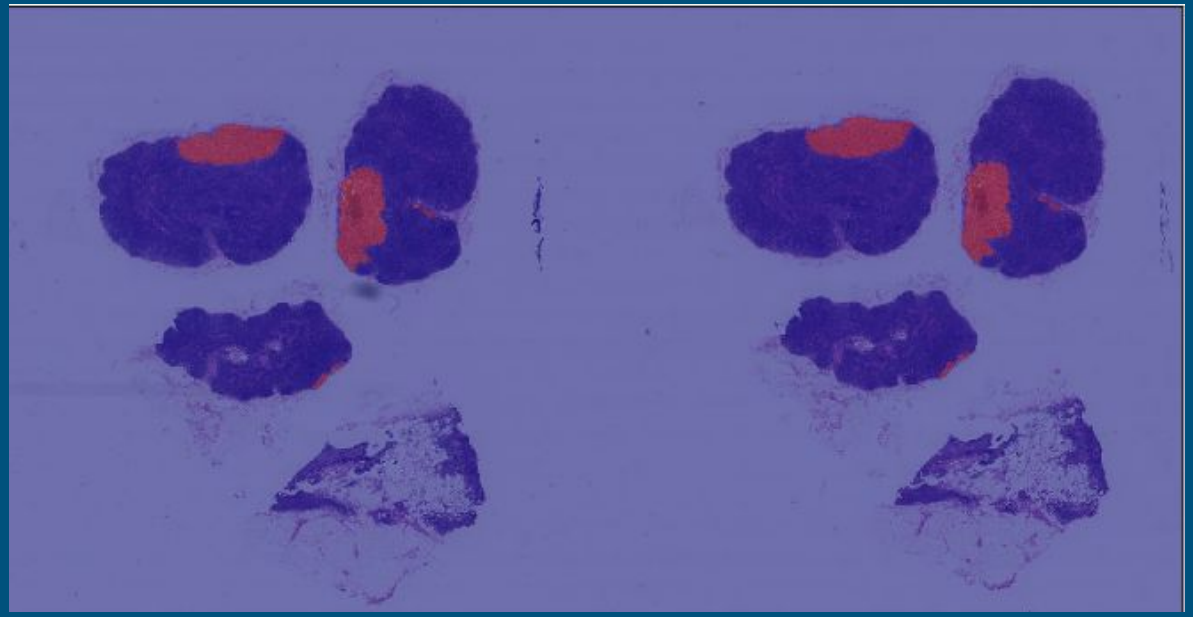
Time saving!!

Helpful tool for physicians!



Data Source: CAMELYON 16 challenge || **Method Reference:** Google group

II. DATA EXPLORATION



Varied Slides!

II. IMAGE PROCESSING

	Train	**Inference**
Slide No.	078, 084, 091	016, 101, 110
Patch extraction	Random (200/slide)	Sliding window
Remove background	Yes	No
Cut edge	Yes	Yes

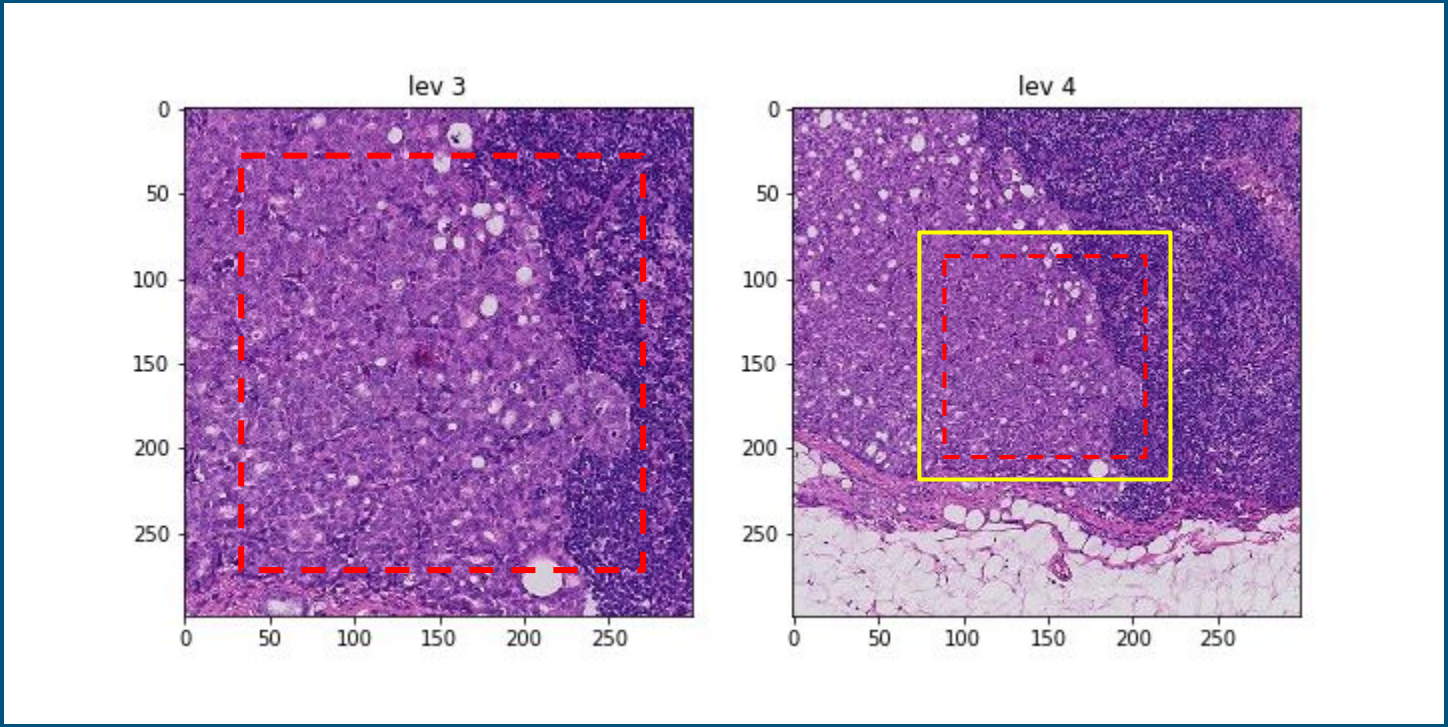
- Extract 299*299 patches from slides.
- Label tumor by center 128*128 (lev3) region (different from paper)
- Balanced training set.

III. MODEL SELECTION

	Inception V3	Train Acc	Val Acc	Speed
Pretrain (Imagenet)	Yes	95%	93%	Quick
Pretrain (Random Ini)	Yes	83%	80%	Quick
+Fine Tuning	Yes	81%	NA	Medium
Self-Defined CNN	No	50%	NA	Slow
Multi-Scale (+lev4)	Yes	87%	89%	Medium

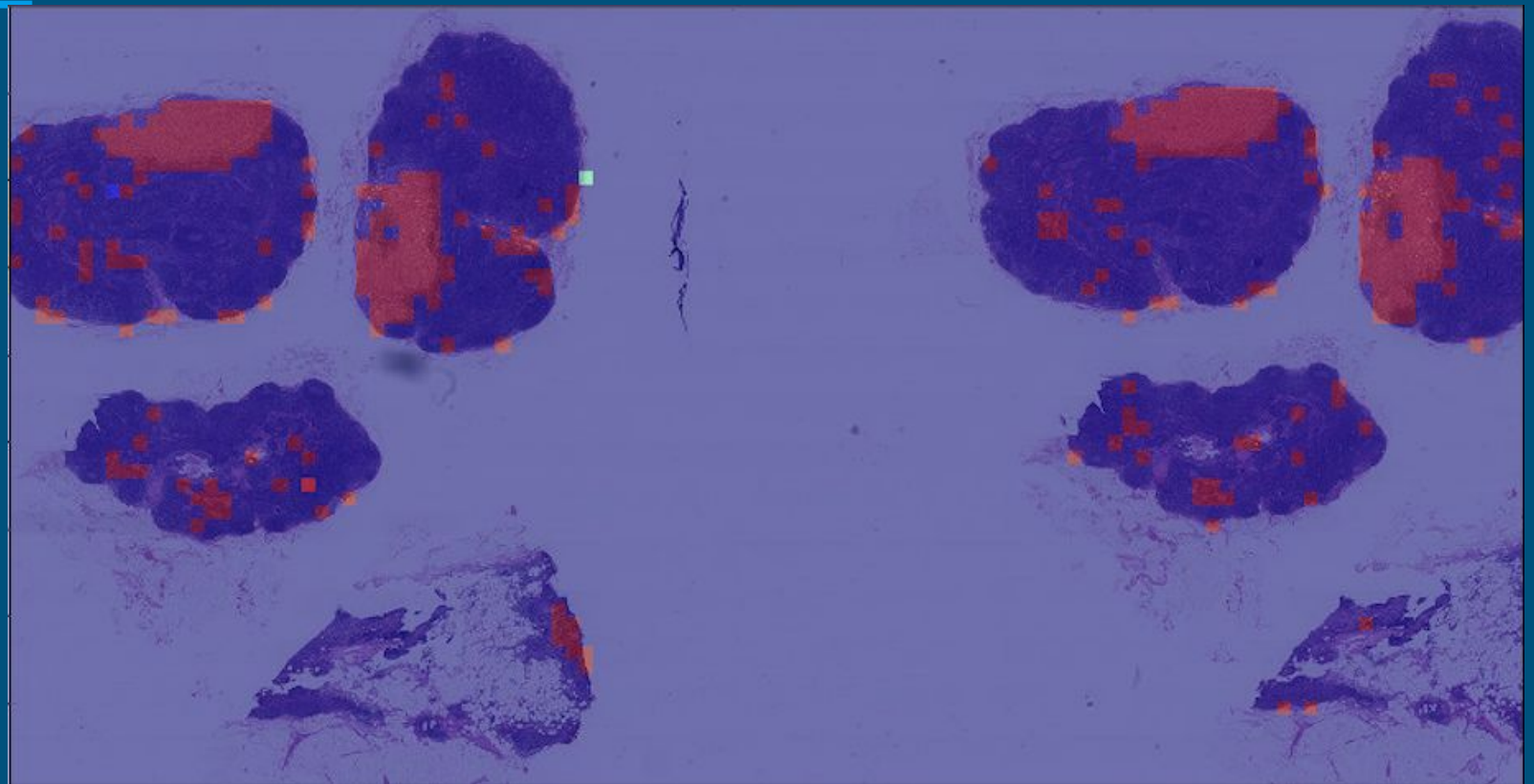
Single-scaled models use the same training set from level 3 (downsampled by 8).

Illustration of Multi Scale



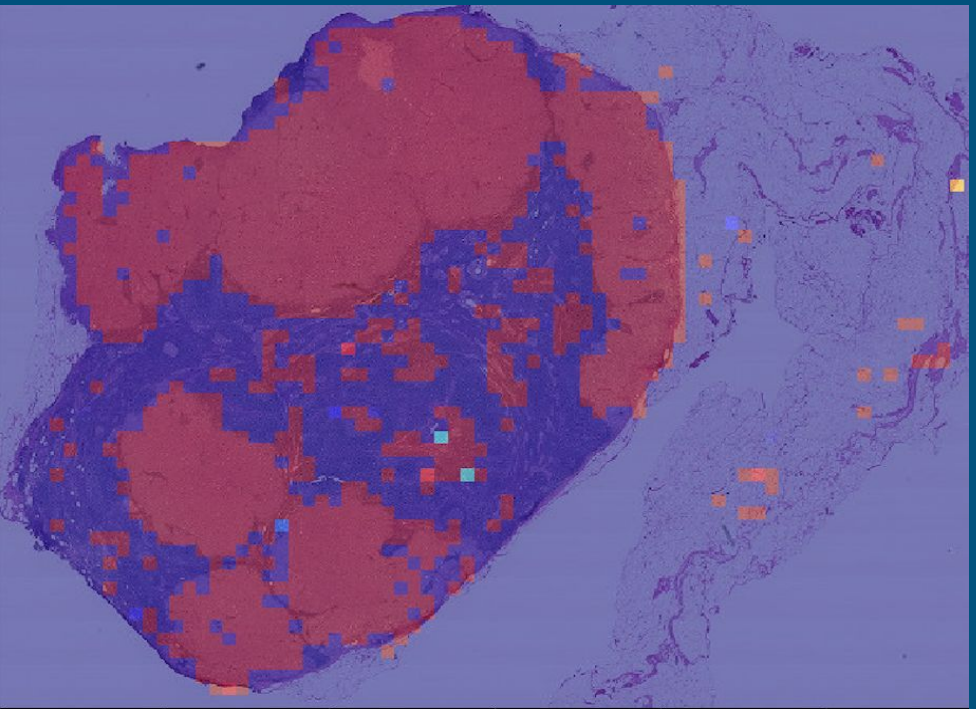
IV. MODEL PERFORMANCE

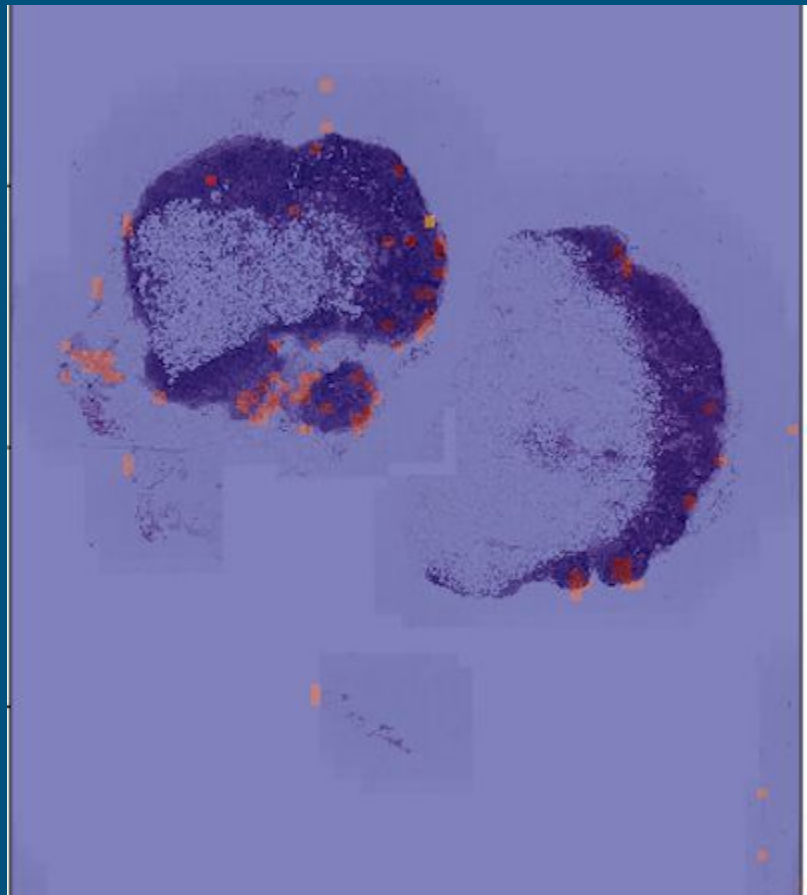
Heatmap + sample image (slide 101)



IV. MODEL PERFORMANCE

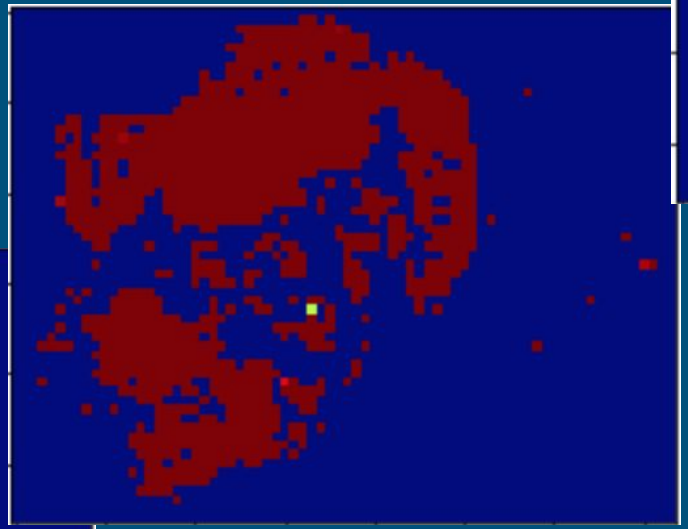
Heatmap + sample image
slide 110 (left), slide 023 (right)



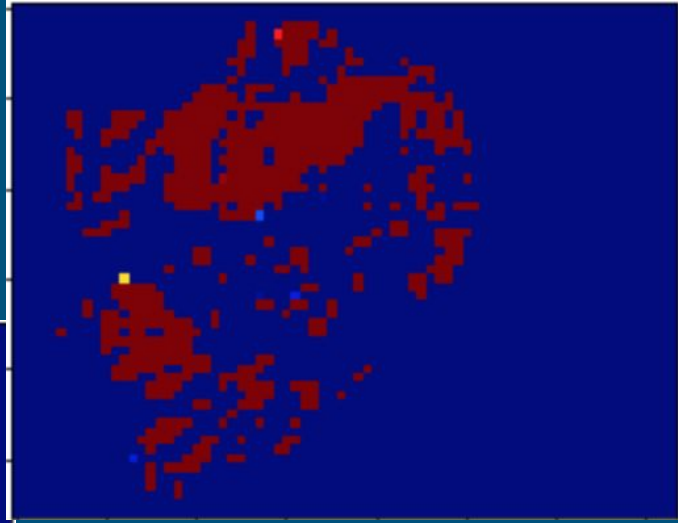


IV. MODEL PERFORMANCE

Single Scale Model



Multi-Scale Model



Ground Truth

Heatmap of slide 110
Single Scale vs. Multi-Scale

IV. MODEL PERFORMANCE

1.ROC Curve: TPR vs. FPR

False Positive Rate = FP/(FP+TN)

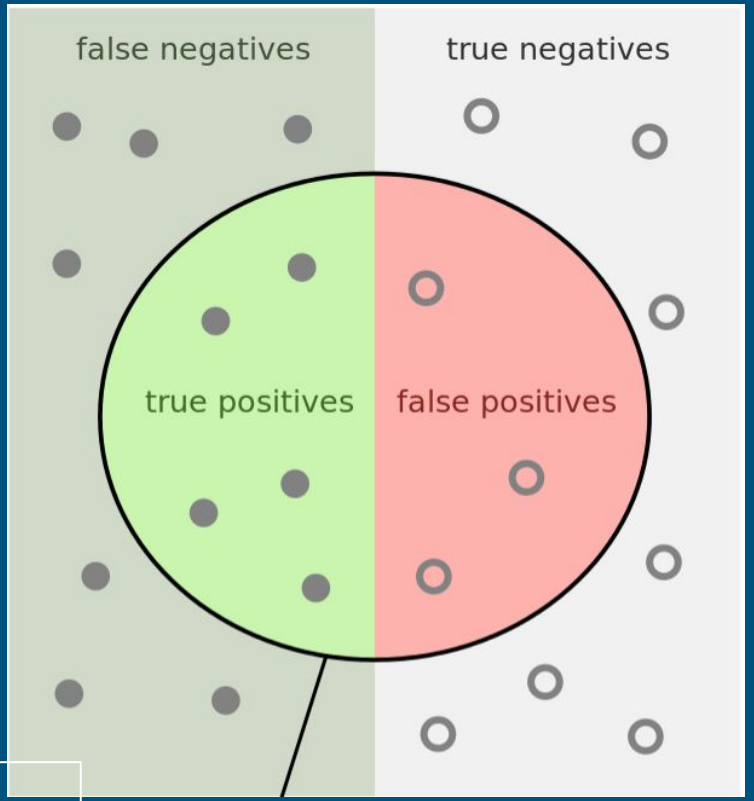
True Positive Rate = TP/(TP+FN)

2.Precision vs. Recall Curve:

Precision = TP/(TP+FP)

Recall = TP/(TP+FN) = TPR

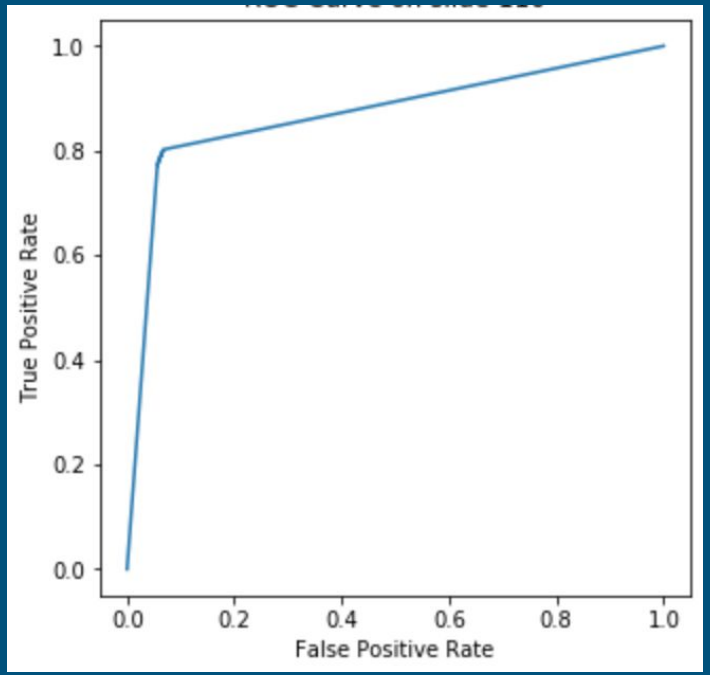
Better for **Unbalanced** data!



Pic. from Wiki

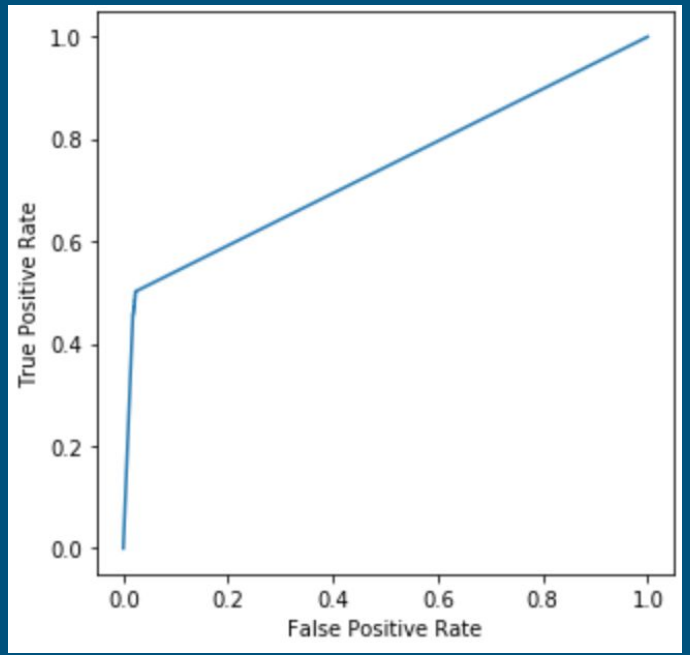
IV. MODEL PERFORMANCE

1.ROC Curve



Left:
Single scale

Right:
Multi-scale



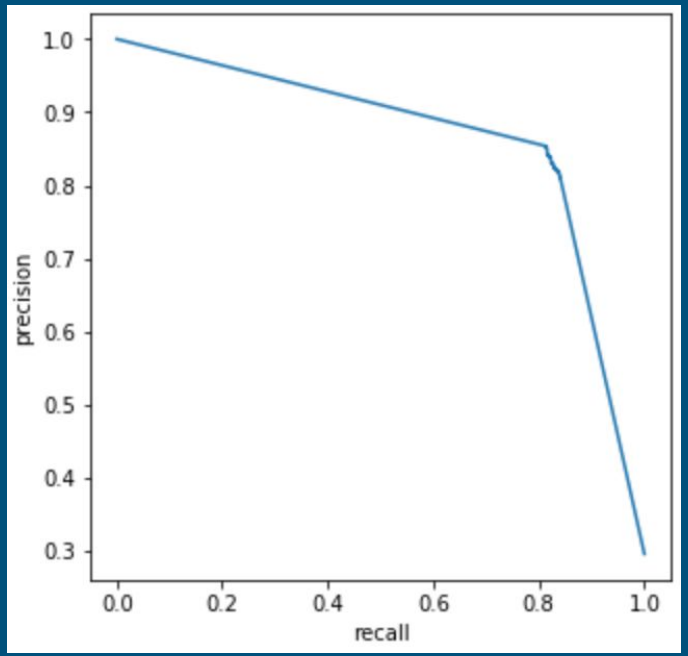
AUC Score = 0.87

AUC Score = 0.74

IV. MODEL PERFORMANCE

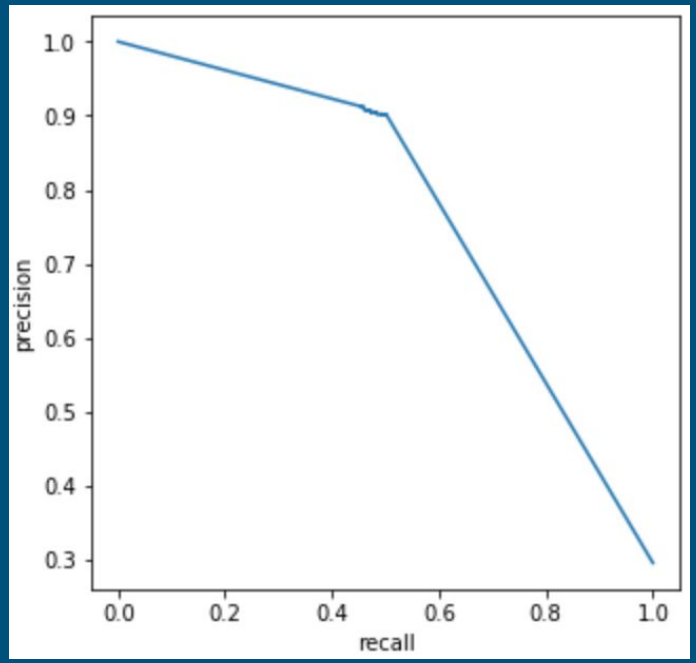
2.Precision vs. Recall Curve

Better differentiation ability!



Left:
Single scale

Right:
Multi-scale



F1 Score = 0.81

F1 Score = 0.61

V. DISCUSSION

1. Pretrained Inception V3 with single high resolution scale had best performance of all models.
2. Small training data got reasonable prediction accuracy.
3. Unbalanced data may be one main reason for prediction error and precision-recall curve (or F1 score) is a good evaluation metric.
4. Future work to improve the performance:

- Use more training data.
- Do more experiments to find optimal model and parameters.
- Advanced method to extract tissue regions in training.