Camelyon 16 - Medical Imaging

Shantanu Jain
slj2142

Problem Statement

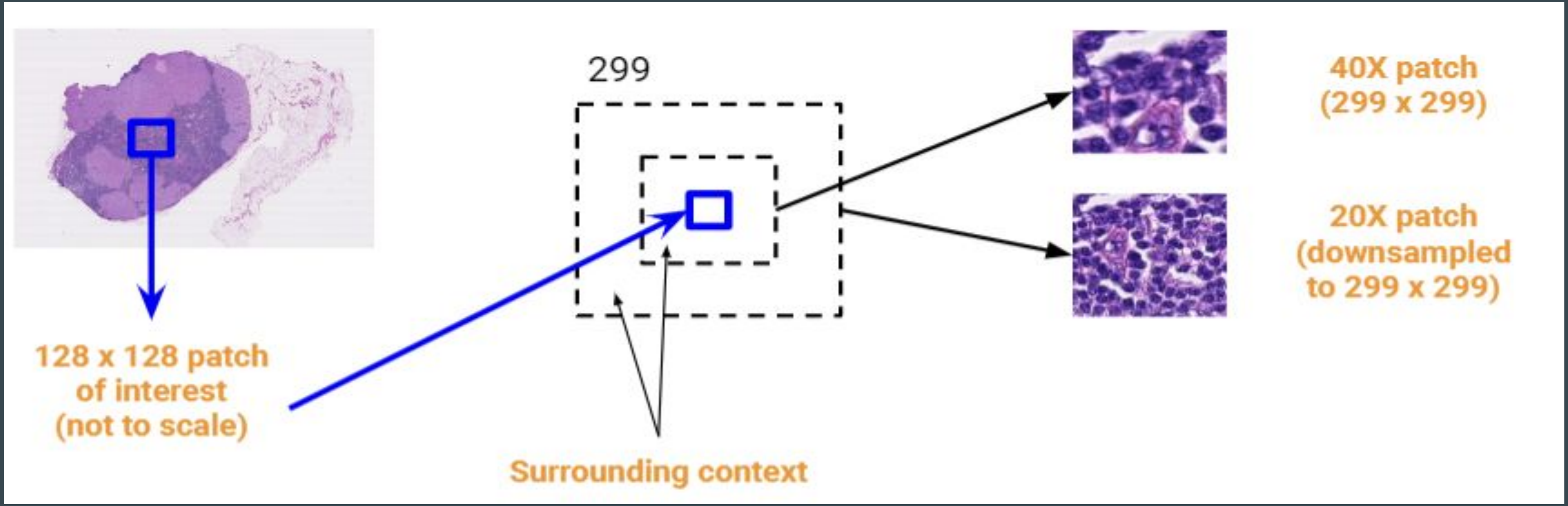
- The task is to predict cancer in gigapixel pathology images.
- Given a collection of data, develop a model that outputs a heatmap showing regions of a biopsy image likely to contain cancer.
- The challenge is from the CAMELYON 16 dataset, whose subset of 21 slides will be used during the entire project for generating the model.

Agenda

- Dataset Preprocessing
  - Load dataset
  - Patching on a zoom level
  - Gray Area removal
  - Centre Alignment in case of multiple zoom levels
- Custom Model
  - One Zoom Level
  - Performance
  - Prediction and maps
- Inception
  - Two Zoom Levels
  - Performance
  - Predictions and heatmaps

Dataset Preprocessing

- Patching (since cancer tissues pertain to small section of the slide)
- Gray Region Removal (for efficiency purposes)
- Centre Alignment (for multiple zoom levels)
- Train Set: 19 slides, Test Set: 2 slides.



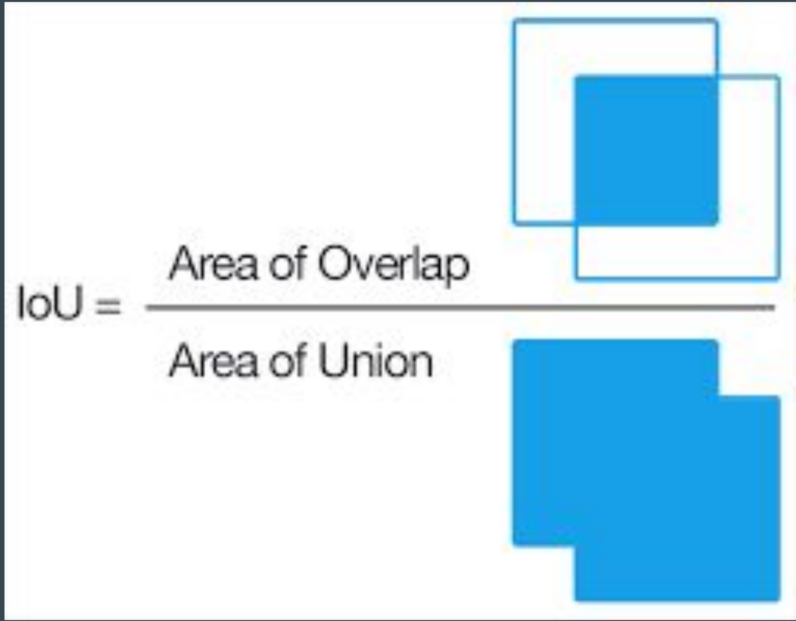
Metrics Used for Evaluation

Intersection Over Union (IOU) Score:

- This metric is generally used in image segmentation task for determining the accuracy of the predictor.
- We can leverage it here as well, since we need to determine how our mask corresponds with the ground truth mask.

F1-Score:

- Since, we are dealing with a highly imbalanced dataset, accuracy metric becomes a weak indicator.
- Hence, comes F1-score which is a measure of the precision and recall of the model.



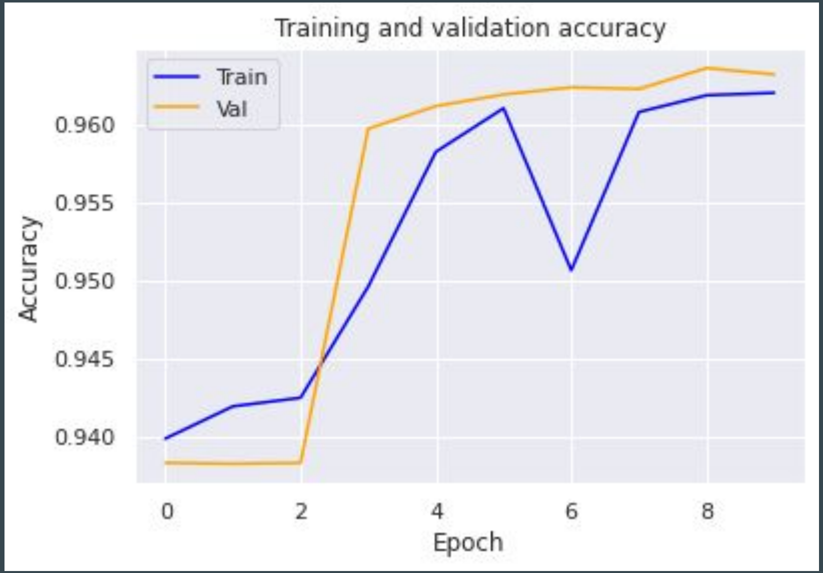
Custom Model with One Zoom Level

- Level = 4
- Patch shape = (32, 32, 3)
- Epochs = 10
- Thresholding at 0.5.
- No Data Augmentation

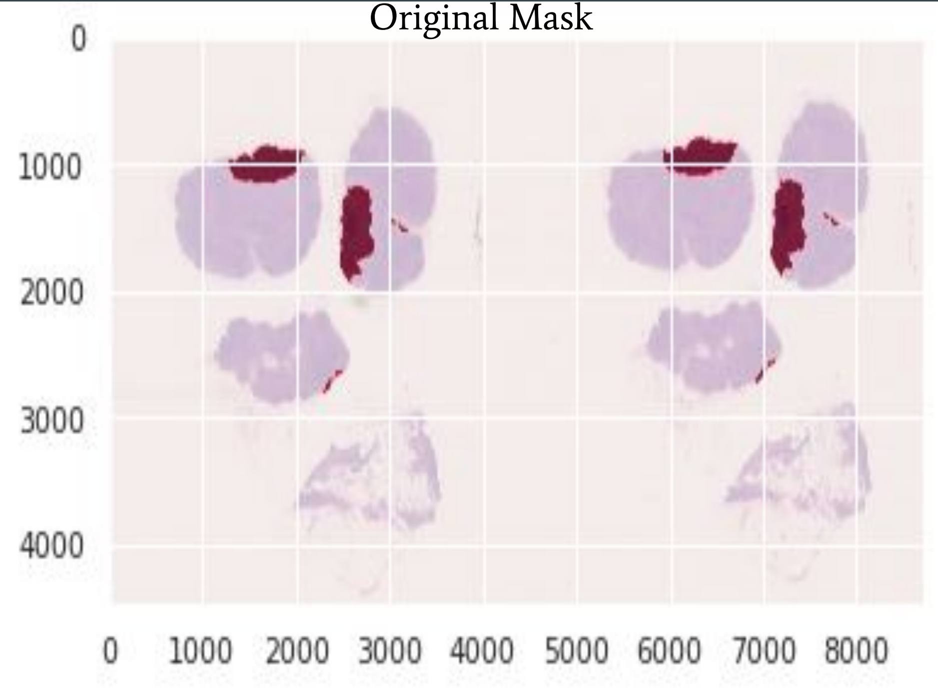
```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 30, 30, 16)        448
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 15, 15, 16)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 13, 13, 64)        9280
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 6, 6, 64)          0
_________________________________________________________________
dropout (Dropout)            (None, 6, 6, 64)          0
_________________________________________________________________
flatten (Flatten)            (None, 2304)              0
_________________________________________________________________
dense (Dense)                (None, 1)                 2305
=================================================================
Total params: 12,033
Trainable params: 12,033
Non-trainable params: 0
```

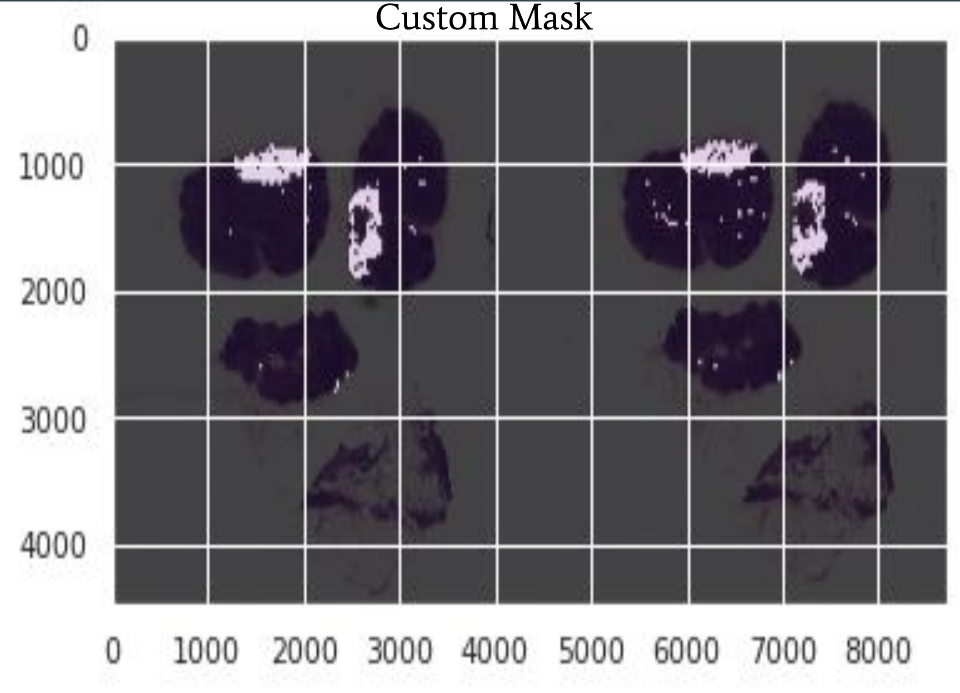
Custom Model: Performance

- Accuracy and Loss metrics (epochs = 10)
- Intersection Over Union (IOU) measure: 0.57

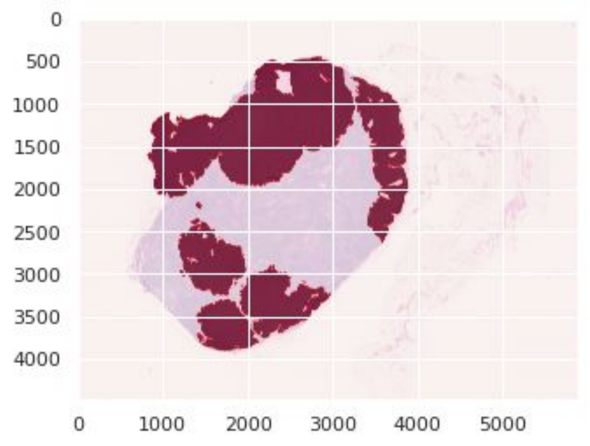


Custom Model: Prediction

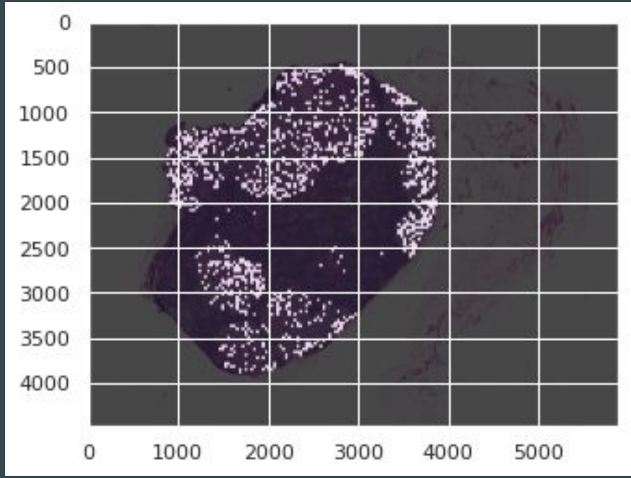




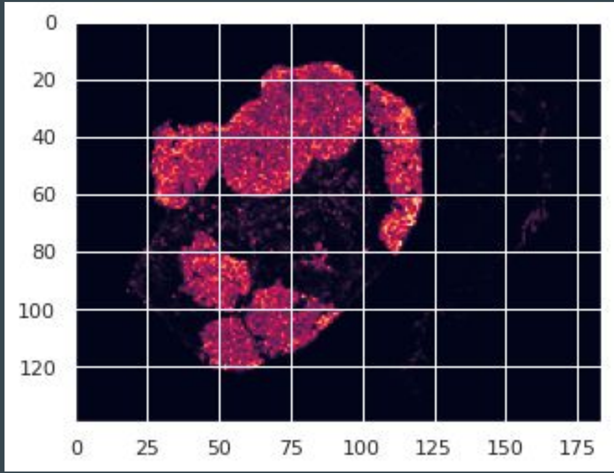
Custom Model: Prediction



With Threshold = 0.5



No Thresholding



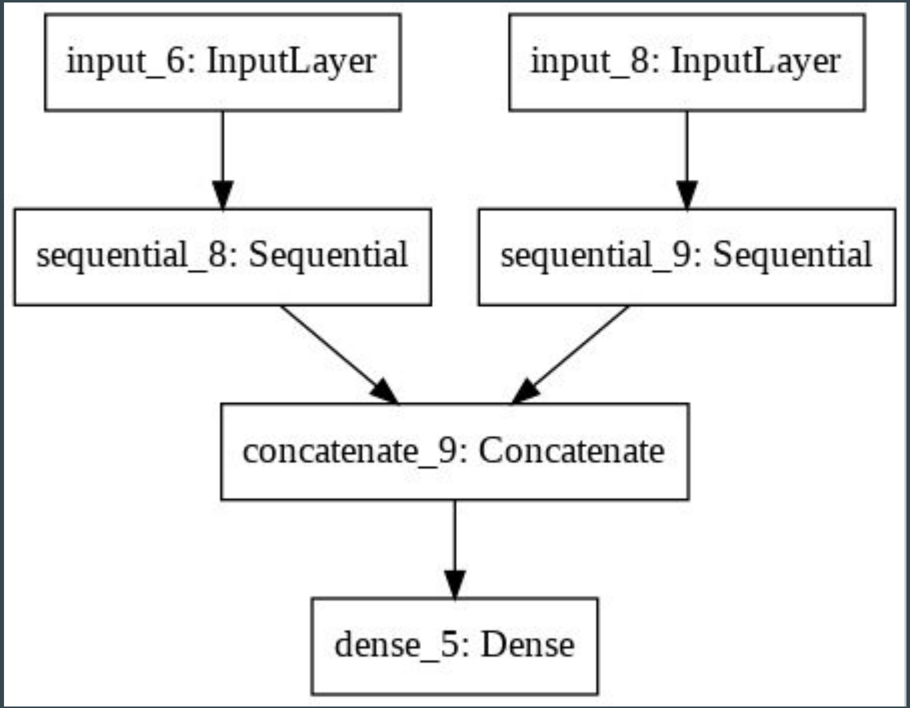
F1-Score and IOU

- Intersection Over Union: 0.57

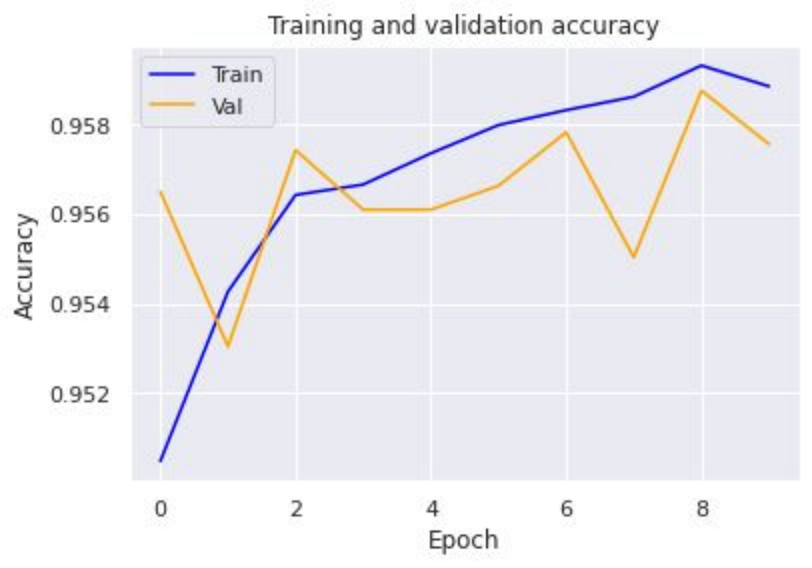
	precision	recall	f1-score	support
0	1.00	0.99	0.99	37481
1	0.59	0.94	0.73	599
accuracy			0.99	38080
macro avg	0.80	0.96	0.86	38080
weighted avg	0.99	0.99	0.99	38080

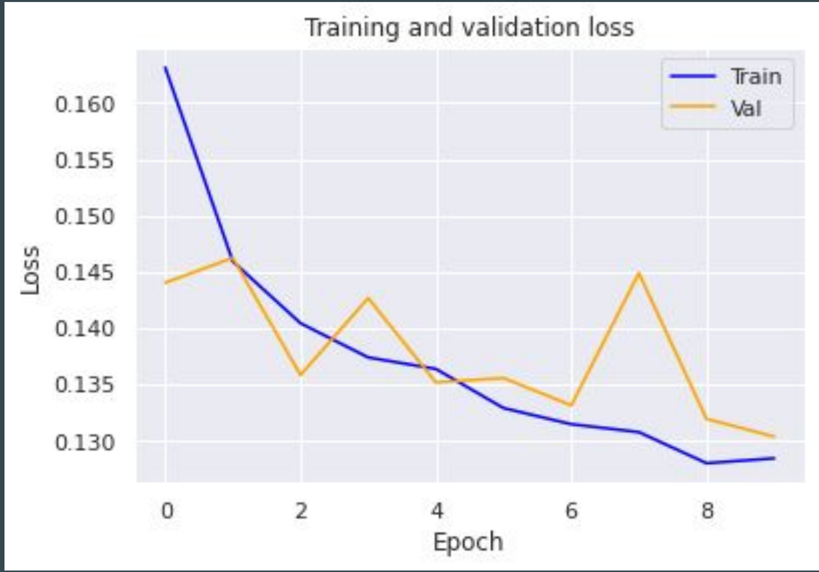
Inception Model with Two Zoom Levels

- Level 3 and Level 4
- Patch Shape: (75, 75, 3)
- Epochs = 10
- Thresholding at 0.5
- Data Augmentation



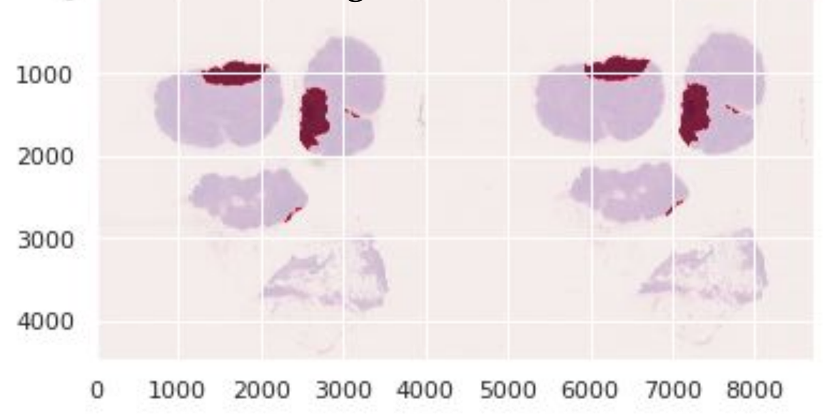
Inception Model: Performance



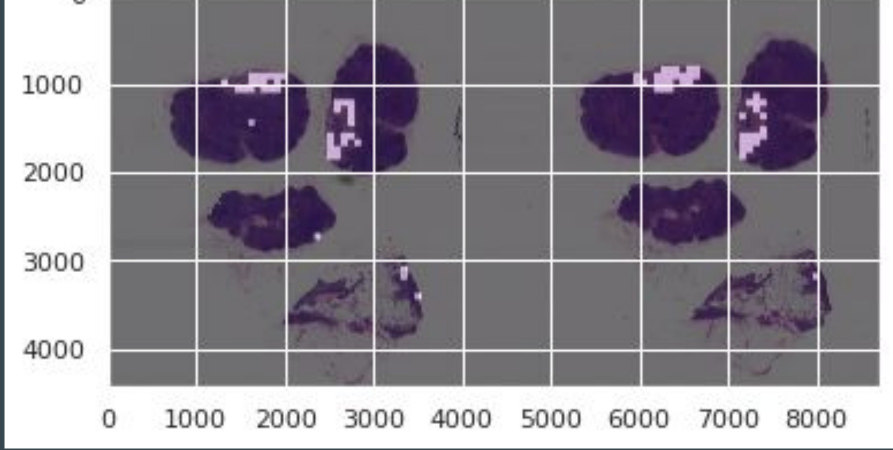


Inception Model: Prediction

Original Mask

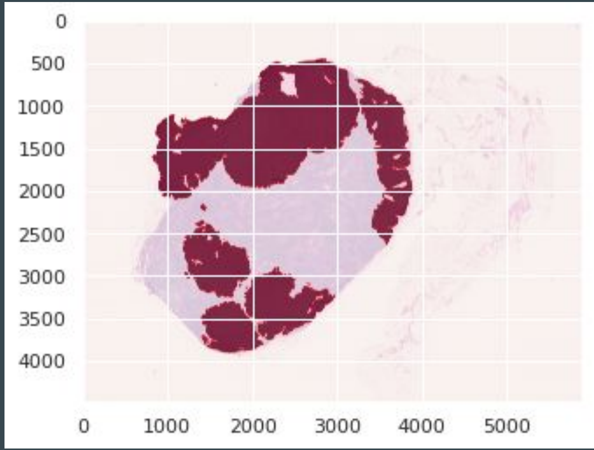


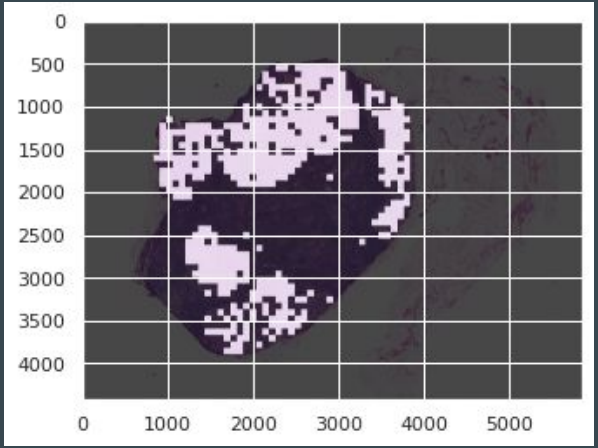
Custom Mask



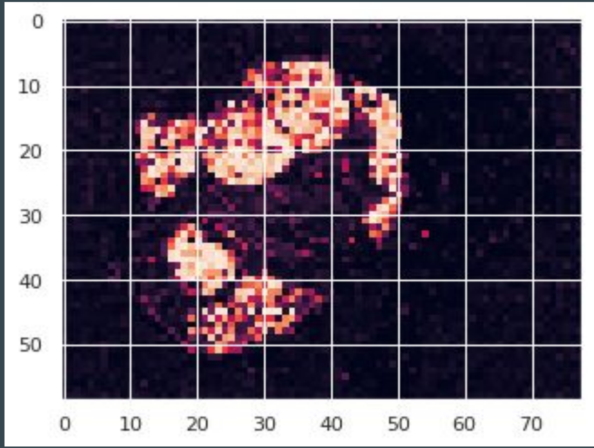
Inception Model: Prediction

With Threshold = 0.5





No Thresholding



F1-Score and IOU

- Intersection Over Union Score: 0.33

	precision	recall	f1-score	support
0	1.00	0.98	0.99	6763
1	0.34	0.94	0.50	81
accuracy			0.98	6844
macro avg	0.67	0.96	0.75	6844
weighted avg	0.99	0.98	0.98	6844

Conclusion

- In terms of accuracy, f1-score and IOU measure, the custom model and InceptionV3 gave similar performance on the 2 slide test set that we had.
- But InceptionV3 performed better, since it was able to achieve the same level of predictability, with only 30065 patches, as compared to 146570 patches for the custom model.
- Hence, even though the performance is similar, InceptionV3 outperforms the custom model in terms of data complexity.

Thank You!