Assignment –4

2023AIB2073, 2023AIB2074

Part-1: Convolution based model

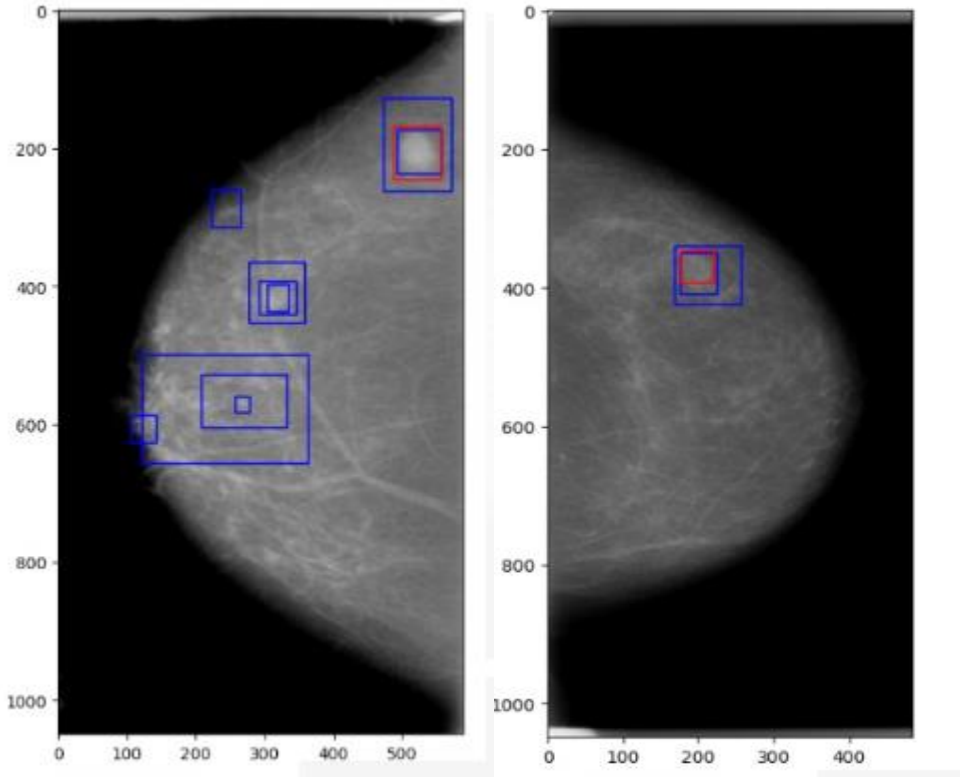
Model used: **Faster RCNN**. Pretrained model has been fine tuned.

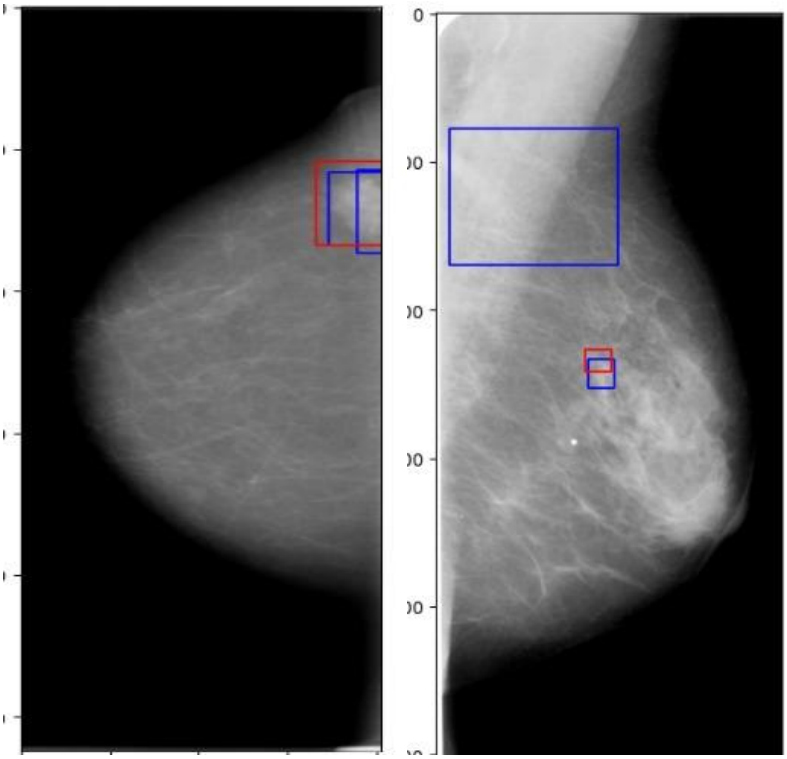
Training parameters and optimizer: **Adam Optimizer** used with 0.0001 learning rate **Early stopping** is used with patience=10.

Data Preprocessing: YOLO dataset is used for this part. Images **normalized** with ImageNet std mean and variance. **Image resizing** was also tried but it didn't improve performance.

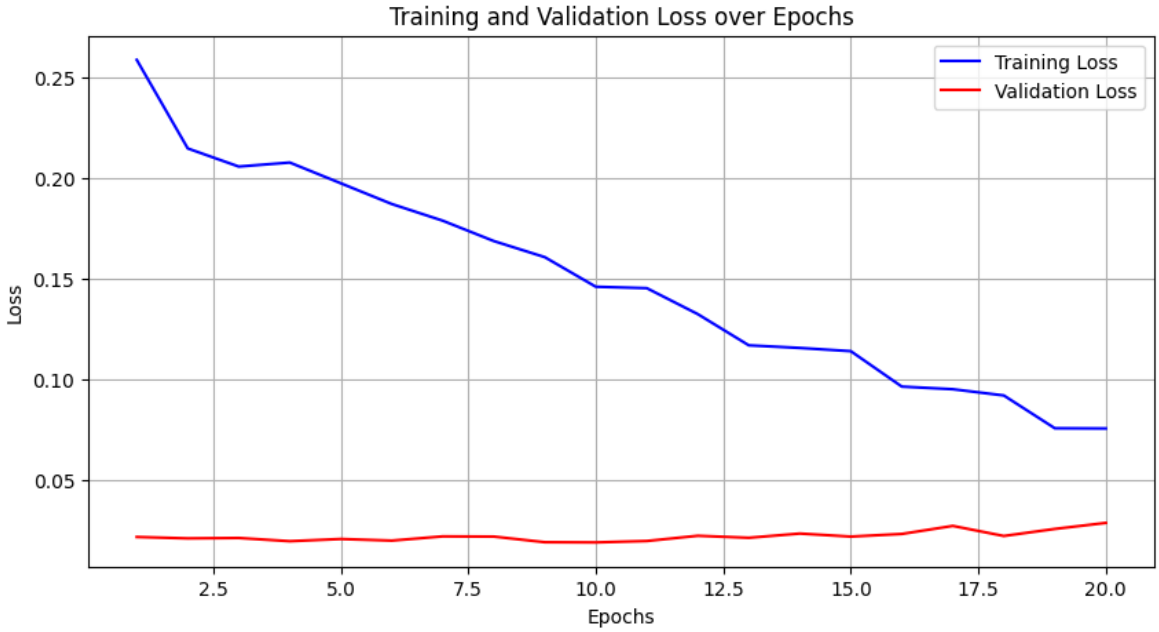
Data visualizations:

Ground truth bounding box is shown in red and Predicted bounding box is shown in green color





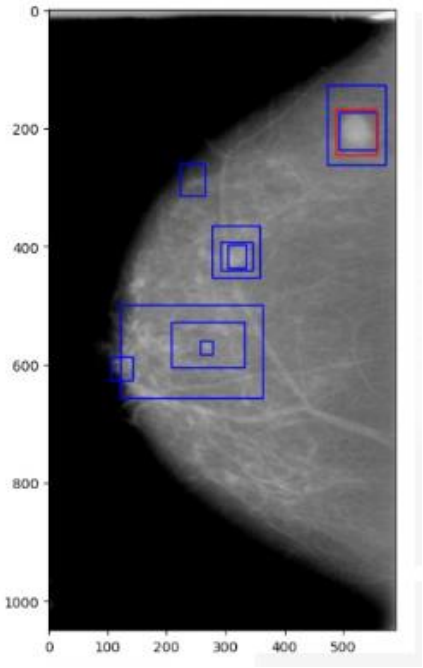
Training and validation loss curves

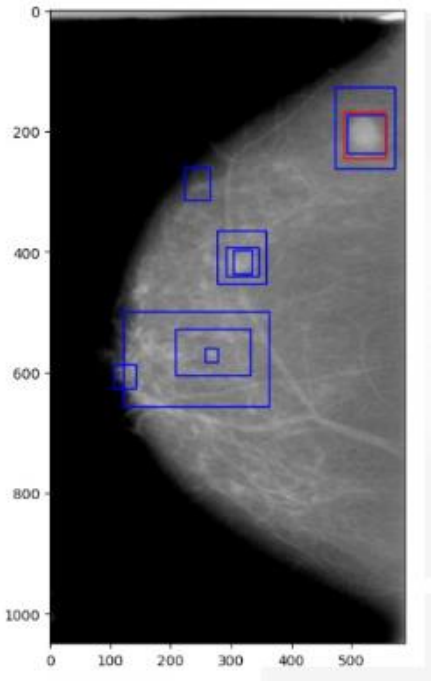


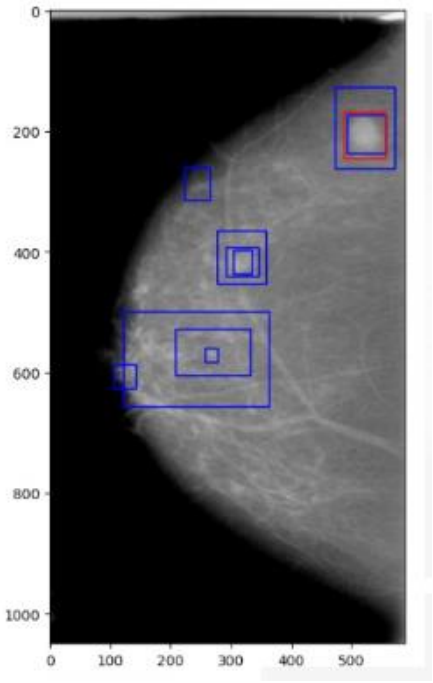
Effects of image pre-processing techniques: Performance improved with Image normalization.

Model predictions on validation set images, pre and post-Non-Maximum Suppression (NMS)

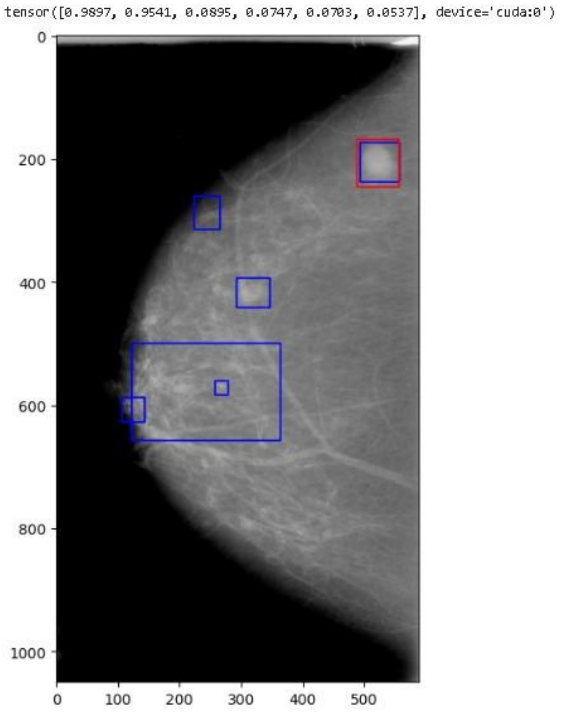
Model predictions examples pre NMS:



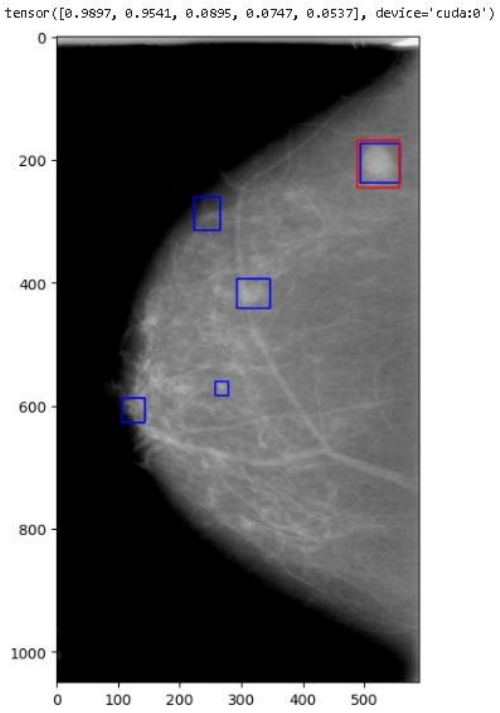




NMS=1 NMS=0.5

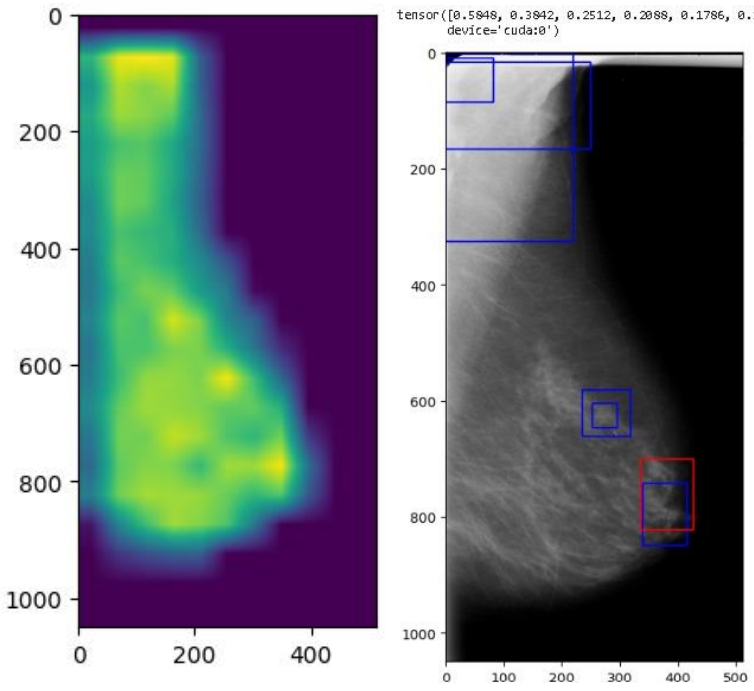


NMS=0.2



NMS=0

Grad-CAM/Attention Maps visualizations of model predictions:



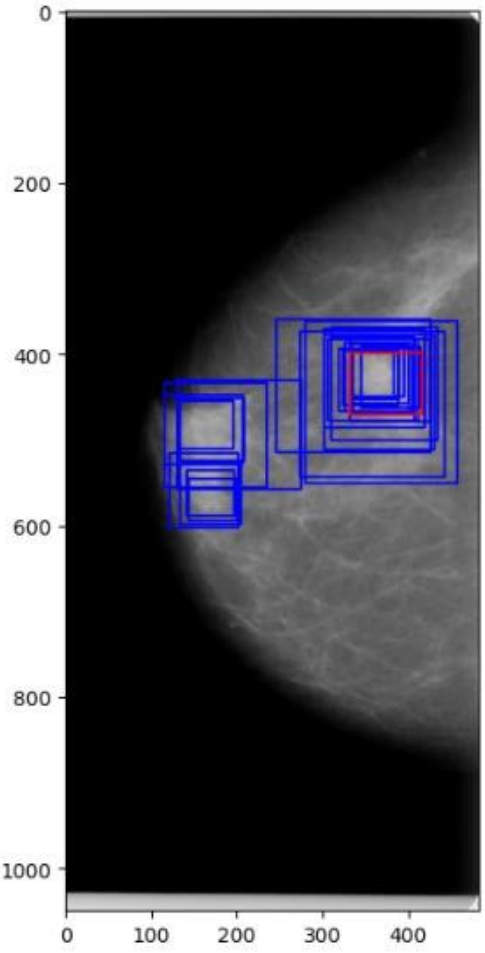
Part-2: Tarnsformer based model

Model used: **Deformable DETR**. Pretrained model has been fine tuned.

Training parameters and optimizer: **Adam Optimizer** used with 0.00001 learning rate and 0.0001 **weight decay** (regularization). **StepLR scheduler** used with step size=2, gamma=0.9). **Early stopping** is used with patience=10.

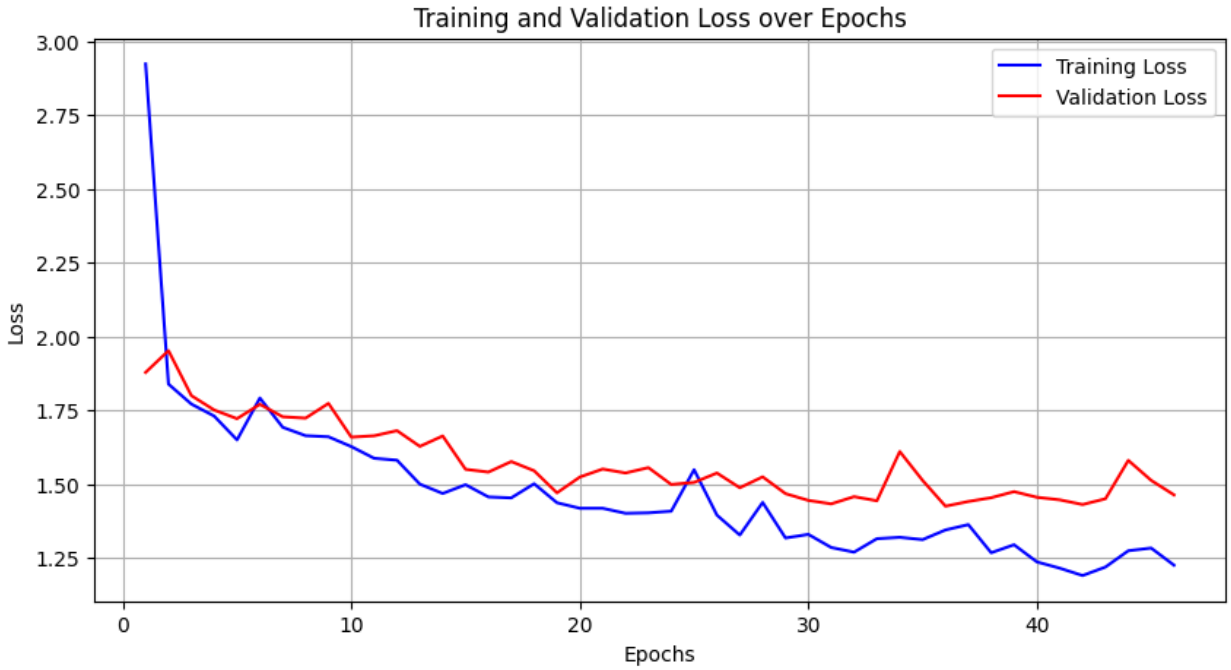
Data Preprocessing: COCO dataset is used for this part. DetrImageProcessor used for Image processing.

Data visualizations:



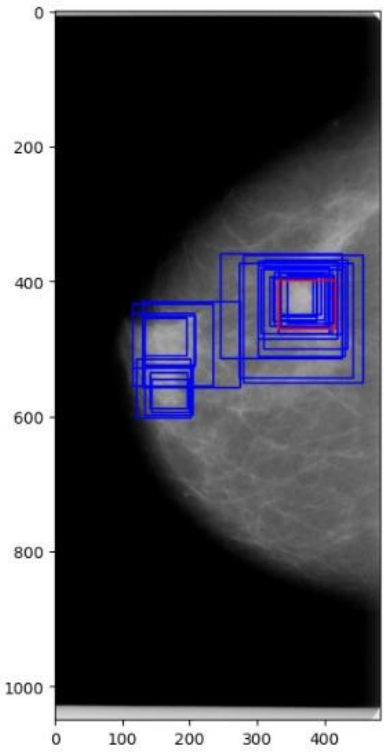
Effects of image pre-processing techniques: Performance improved with Image normalization.

Training and validation loss curves

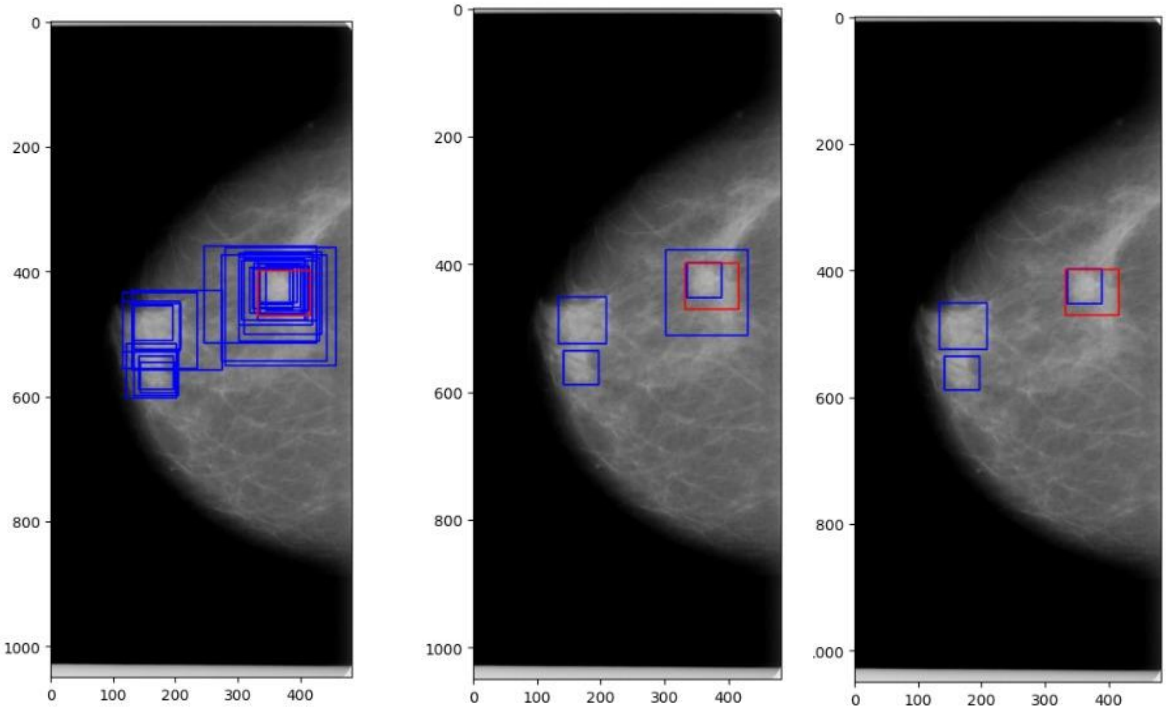


Model predictions on validation set images, pre and post-Non-Maximum Suppression (NMS)

Model predictions examples pre-NMS:

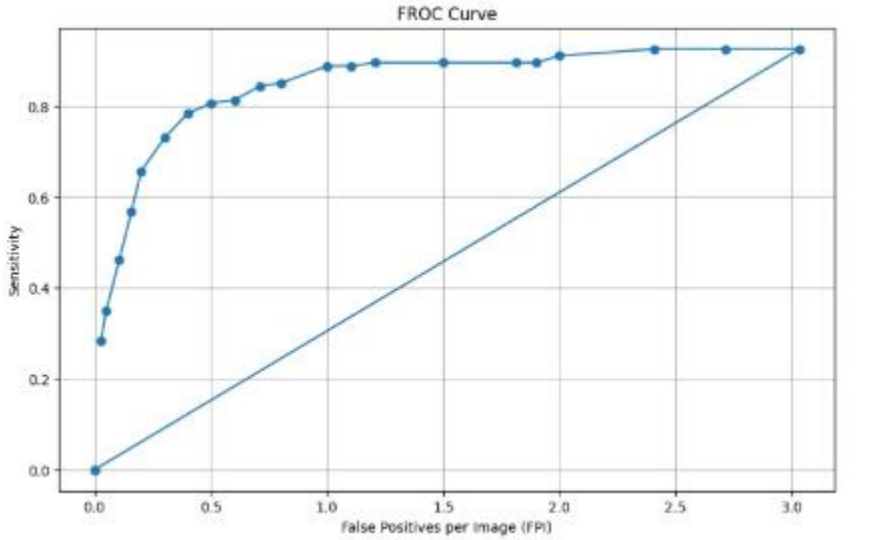


Model predictions examples post NMS:

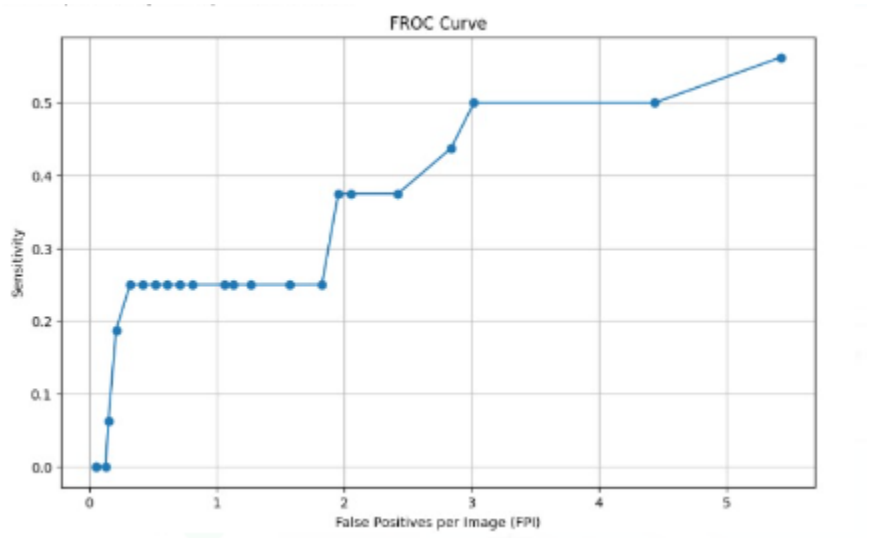


NMS = 1 **NMS = 0.2** **NMS = 0**

Comparision the performance of the two trained models:



Faster RCNN

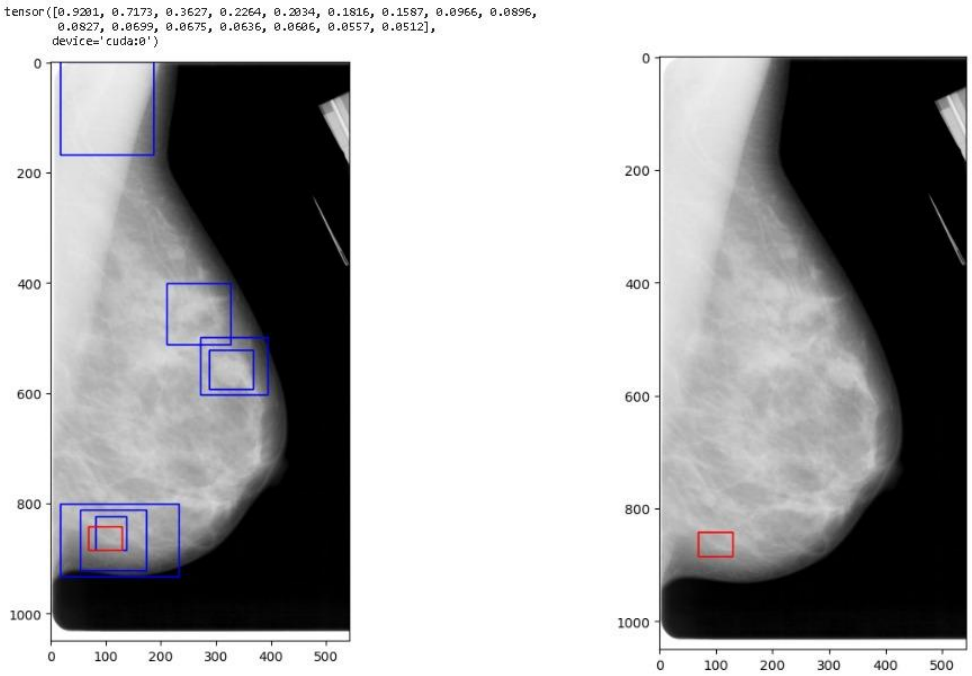


Deformable DETR

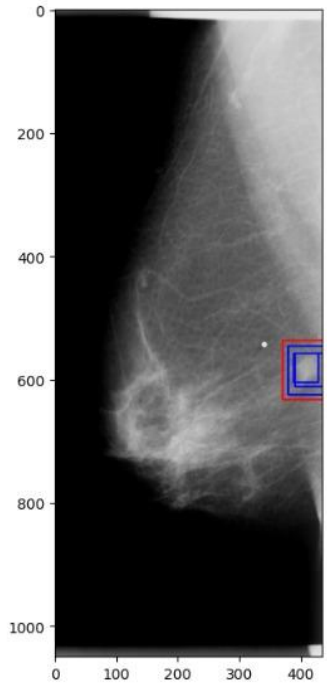
Remarks on comparison: As per FROC curves, performance of Faster RCNN is better than Deformable DETR.

Reasons: Number of parameters in Deformable DETR is huge. Training data is small for training this model.

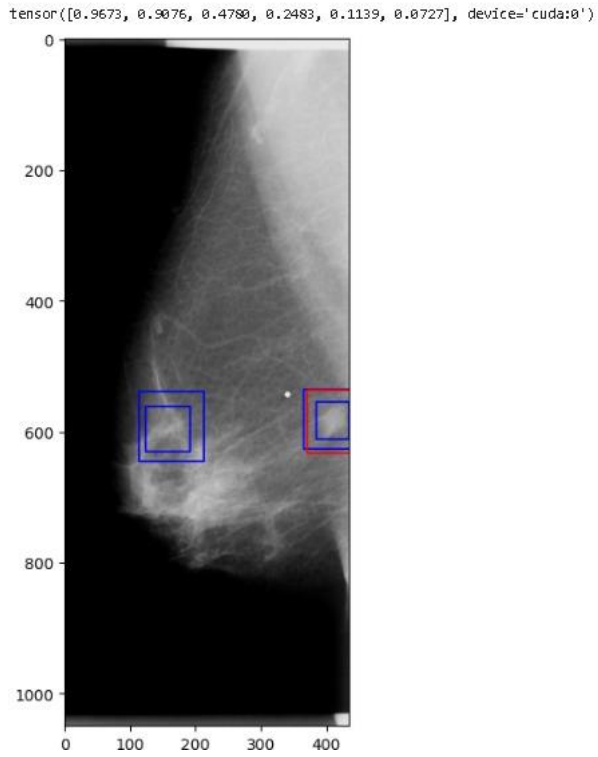
On same FPS of 2 , rnn has threshold of 0.114 and transformer had threshold of 0.9049 . so we kept the same confidence threshold to observe the bounding boxes on an image . RNN created some extra boxes on an image.



But transformer could not even identify the location. So RNN has high false positive with high accuracy but transformer did not perform well due to high no. of parameters and less no. of training examples .



transformer on same fps of 2



rnn on fps of 2 , showing extra locations .