Computer Vision (COL 780)
Assignment-4: Object Detection in Breast Cancer Screening

Pulkit Singal (2023AIB2064), Anshul Patel (2023AIB2072)

Link to Models: https://drive.google.com/drive/folders/1pIg2gTLL6NNkr6eu_4H-QtSGO5onqB_L

Faster R-CNN (Convolution-based Model)

Library Used:

torchvision.models.detection.fasterrcnn.fasterrcnn_resnet50_fpn()

Loading of the Data:

1. From the YOLO Dataset, all the malignant images (corresponding to which annotation files are present) are read and stored in a list.
2. From the annotation files, all the boxes are read and converted from cxcywh (normalized) format to xyxy (original) format.
3. The malignant boxes are assigned the label '12' corresponding to the 'N/A' class present in the model.
4. These boxes and labels are added to dictionary named 'targets'.

Training of the Model:

Pretrained FasterRCNN model with weights backbone ResNet50_Weights.IMAGENET1K_V1 is imported and finetuned on the given training dataset with the following training parameters:

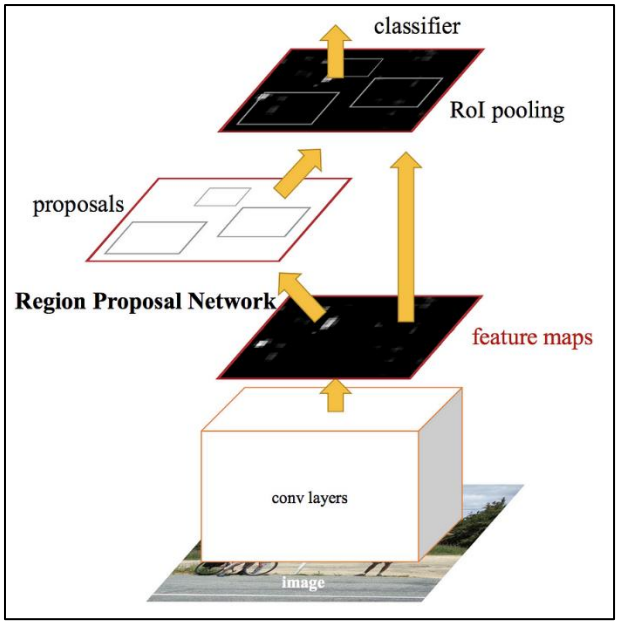
No. of Epochs: 80

Learning Rate: 0.0001

Batch Size: 8

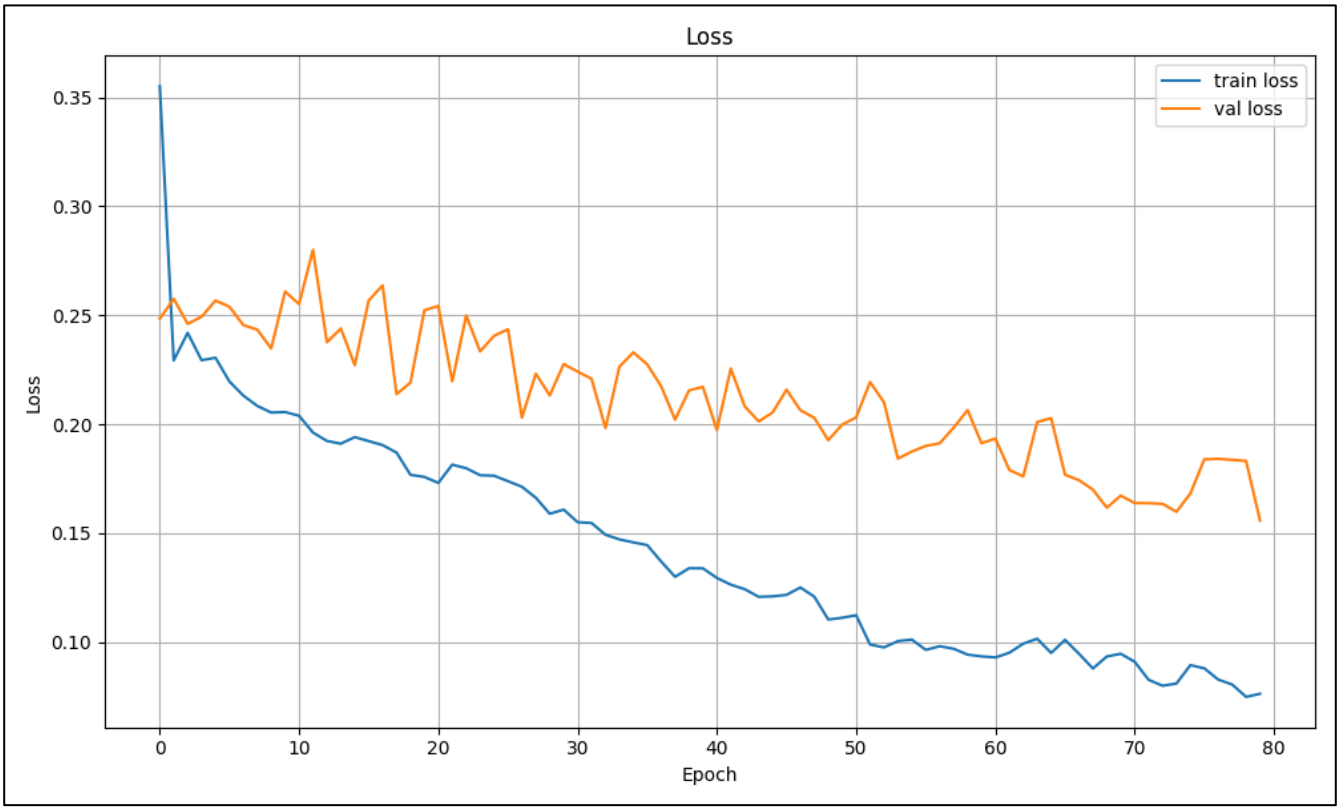
Optimizer: Adam()

Scheduler: ReduceLROnPlateau()



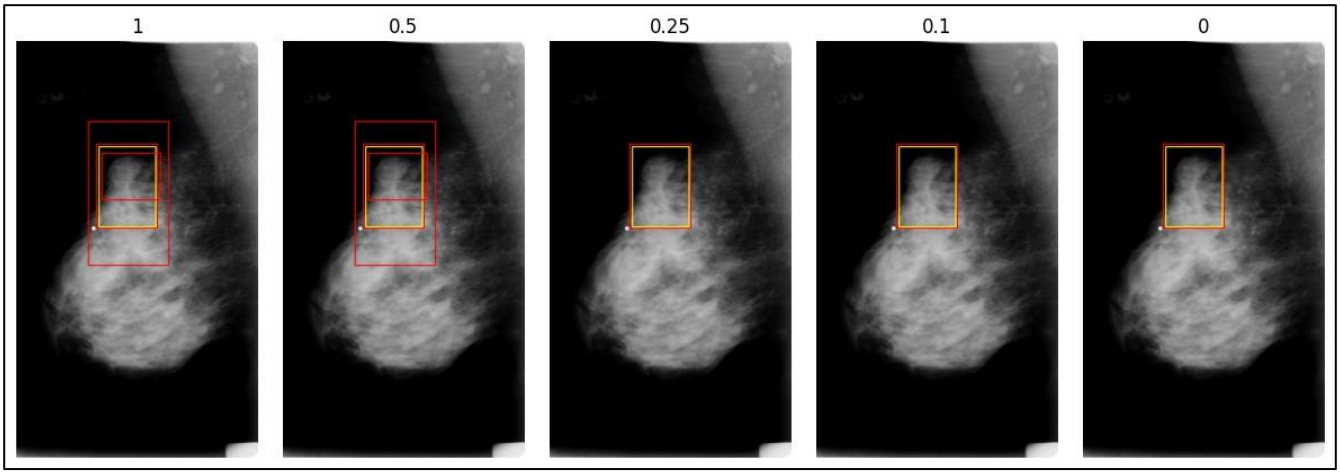
Training and Validation Loss Curves:

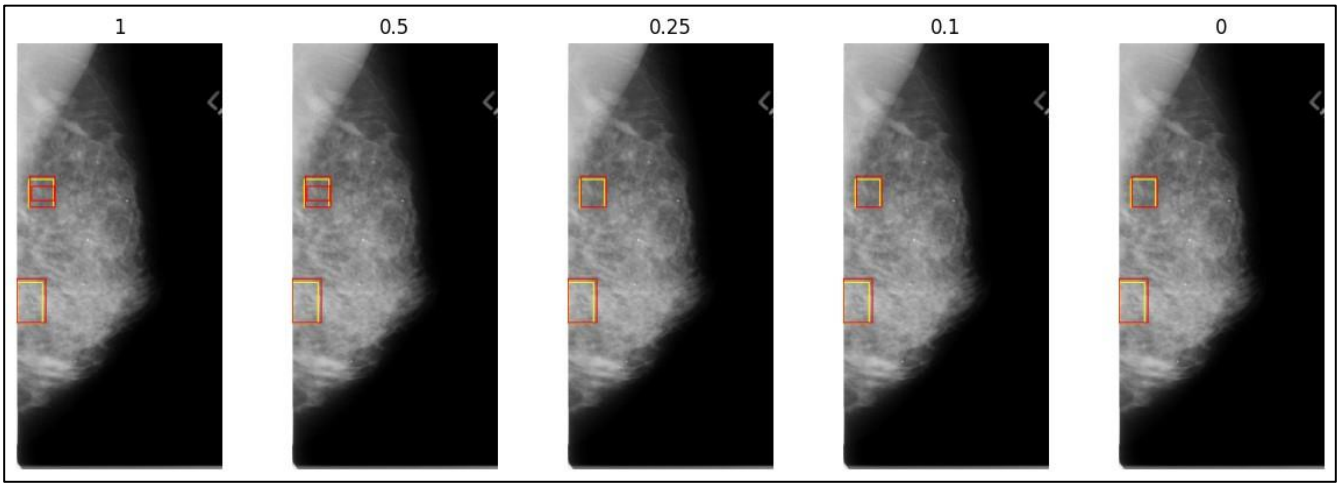
The graph below shows the plot of training and validation loss as a function of the number of epochs. It is observed that the loss value decreases with an increase in the number of epochs. This suggests that the model's performance is improving as it is trained on more data.

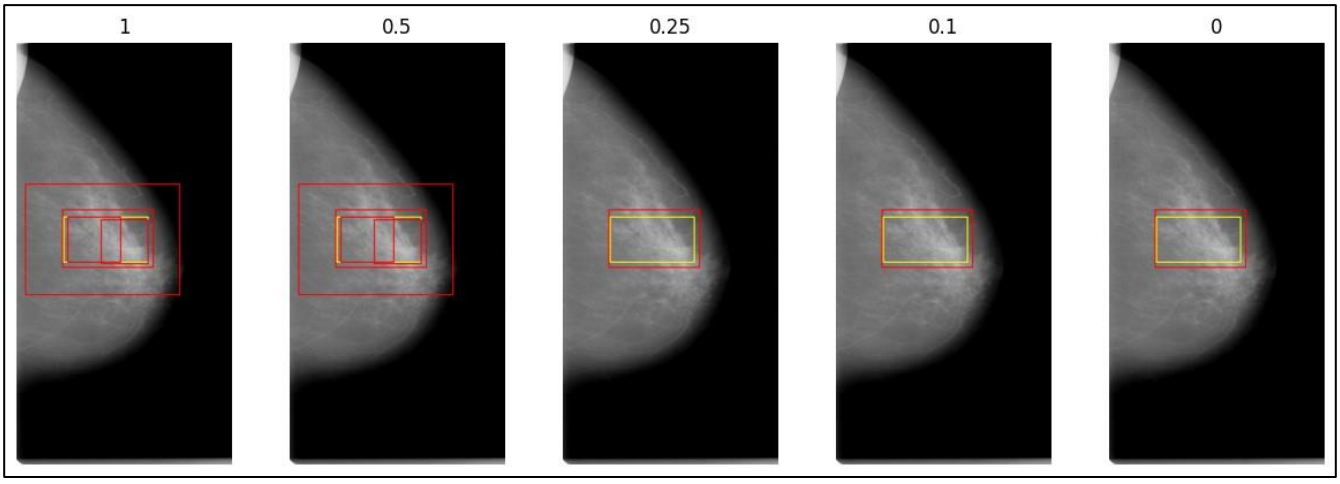


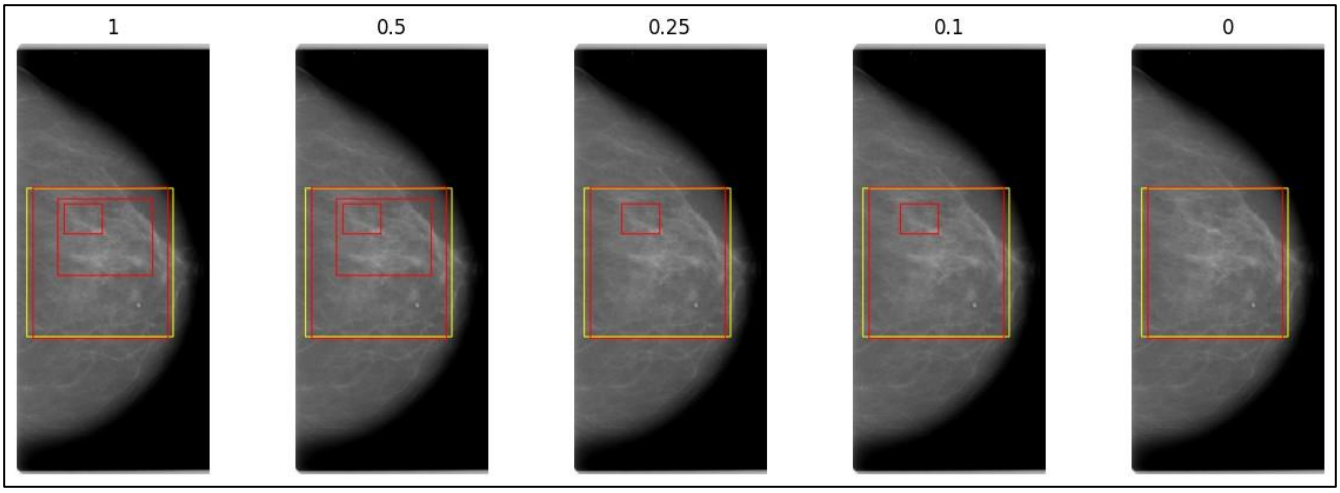
Model Predictions on Images, pre and post NMS:

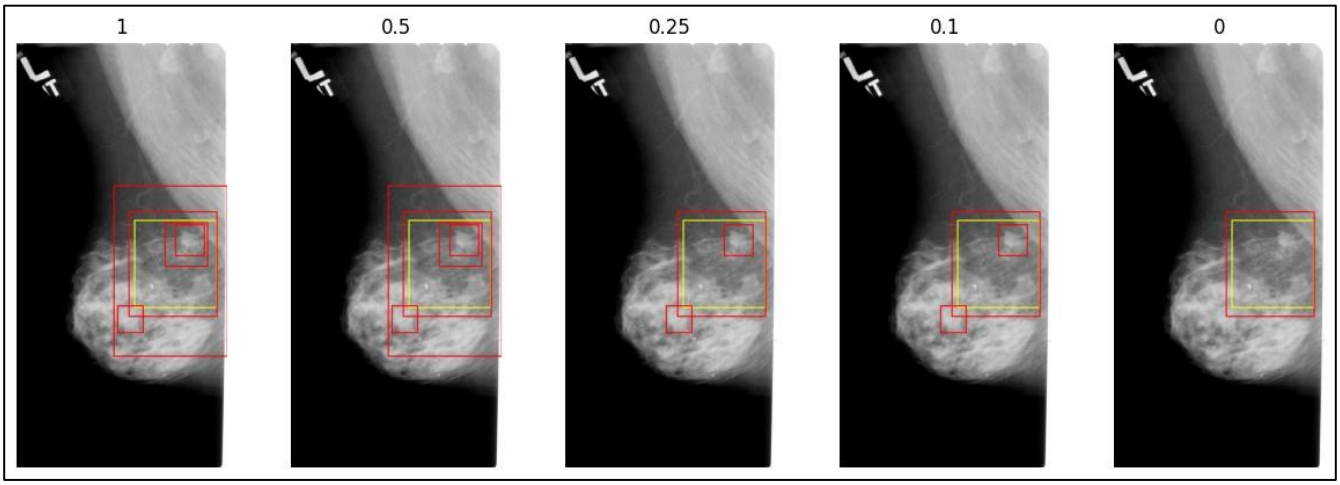
The illustration depicts the fluctuation of Non-Maximum Suppression (NMS) across the predictions generated by a detection module. The yellow bounding box signifies the ground truth for a malignant tumor, while the red bounding boxes denote predictions made by the detection model. Within this visual representation, we observe the nuanced variations in NMS. We explore five Intersections over Union (IOU) thresholds: 1, 0.5, 0.25, 0.1, and 0. A threshold of 1 indicates the inclusion of a bounding box regardless of overlaps, while a threshold of 0 signifies the removal of bounding boxes even in the presence of slight overlaps.

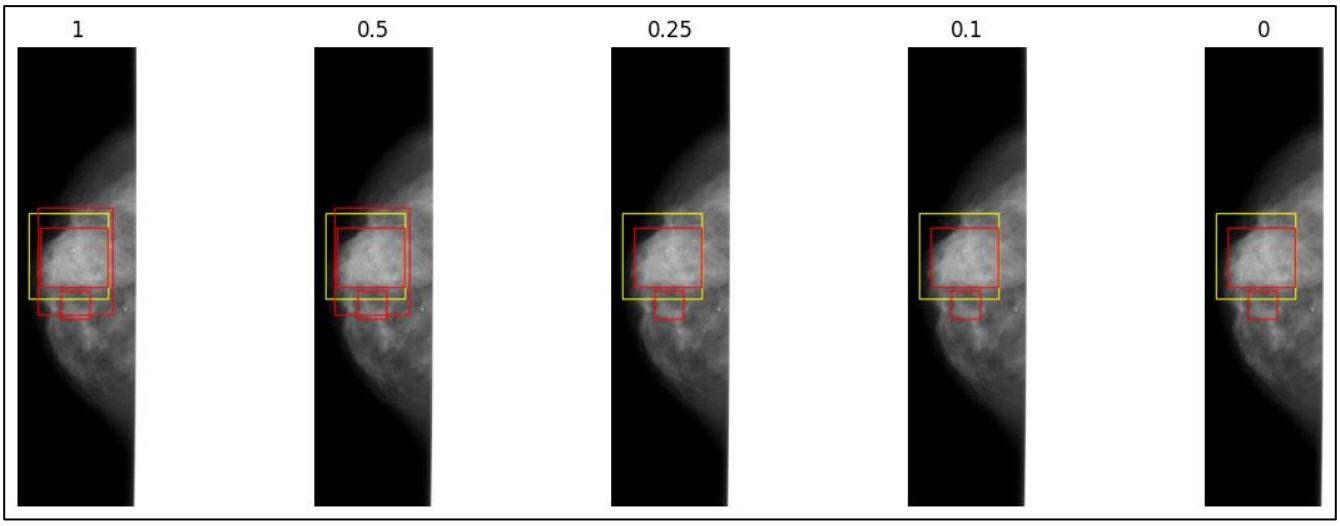

1
0.5
0.25
0.1
0

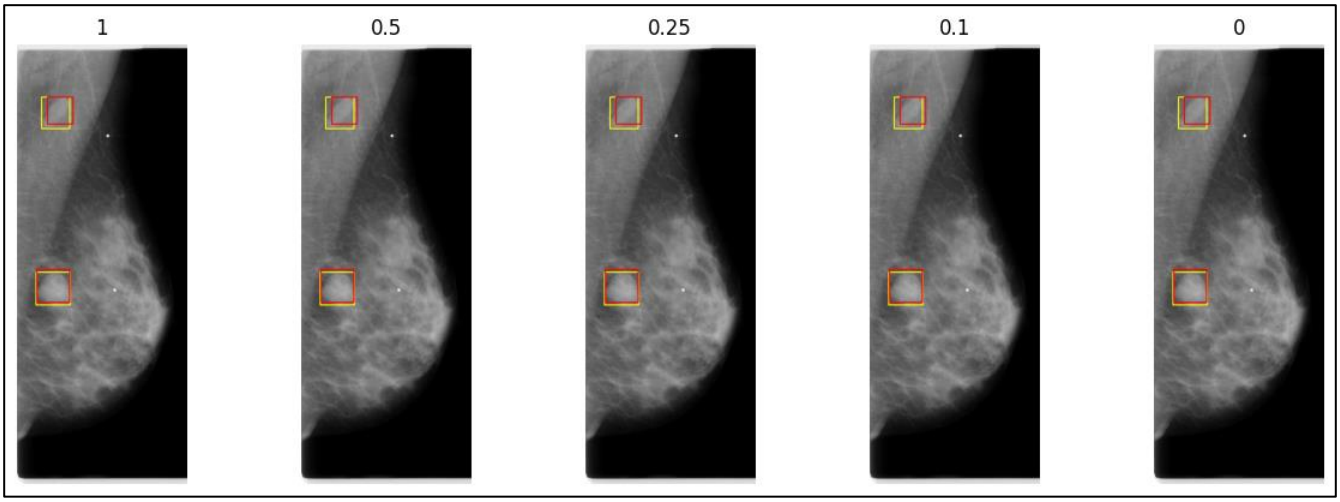

1
0.5
0.25
0.1
0

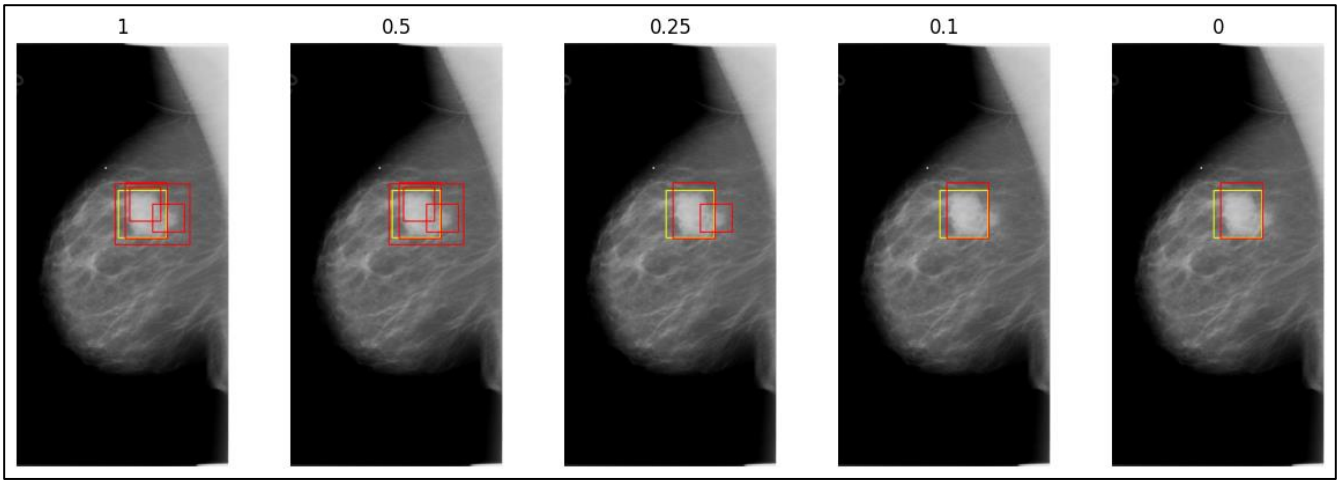

1
0.5
0.25
0.1
0

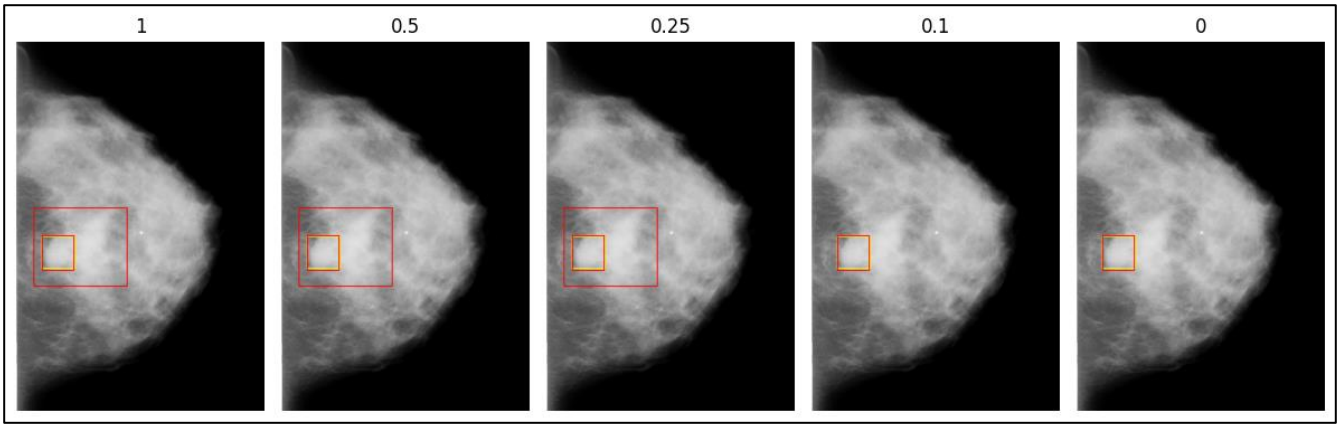

1
0.5
0.25
0.1
0

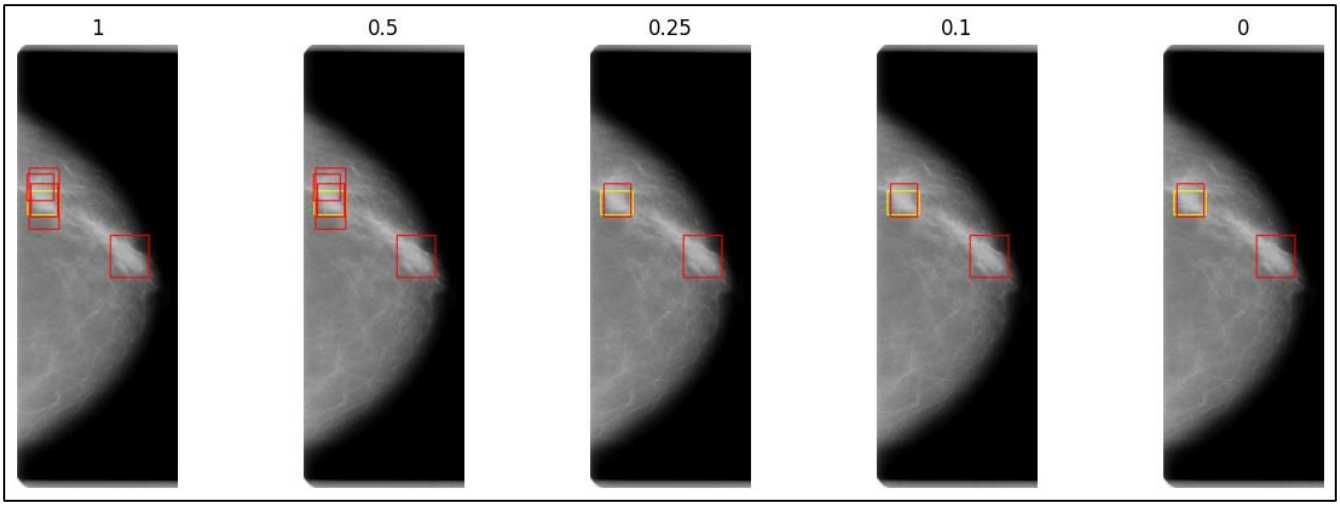

1
0.5
0.25
0.1
0


1
0.5
0.25
0.1
0


1
0.5
0.25
0.1
0

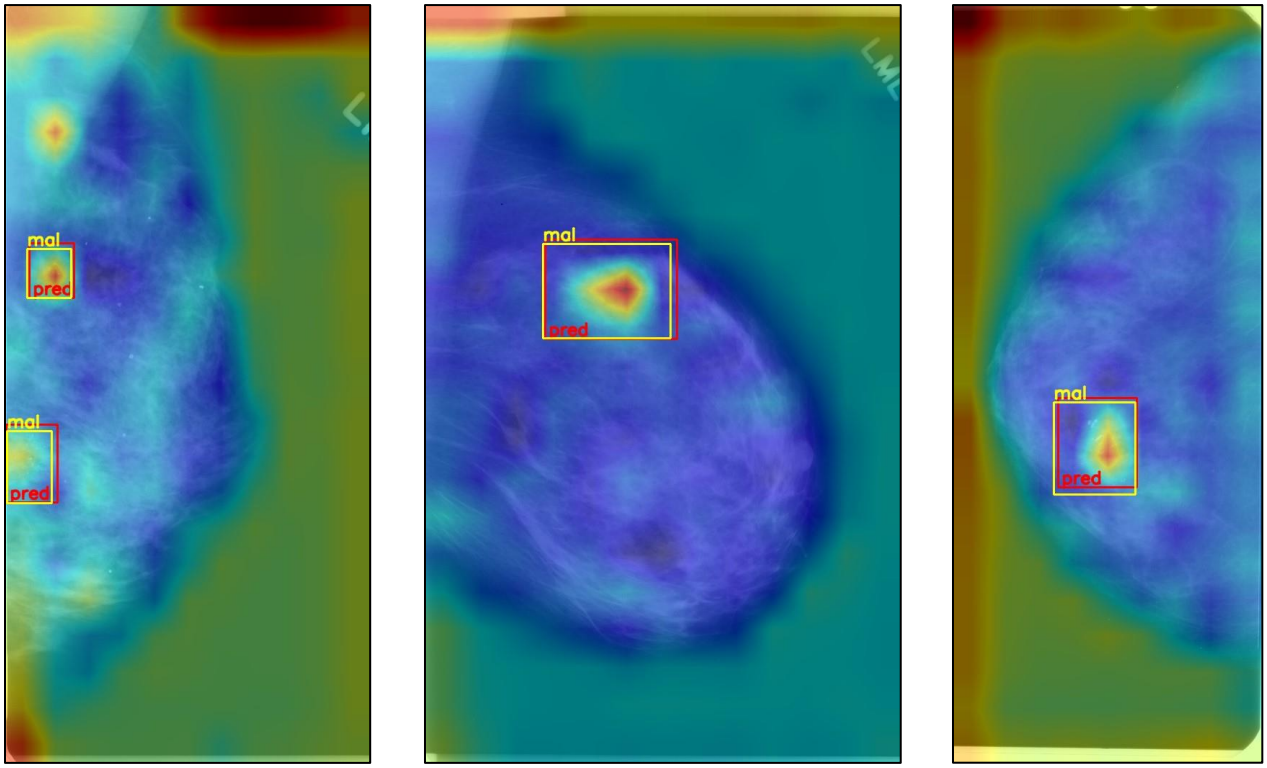

1
0.5
0.25
0.1
0

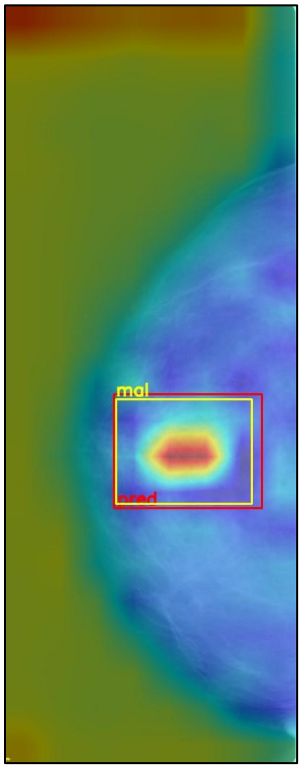

1
0.5
0.25
0.1
0

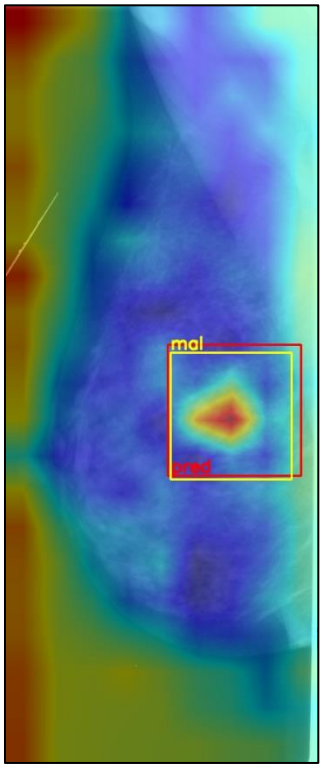


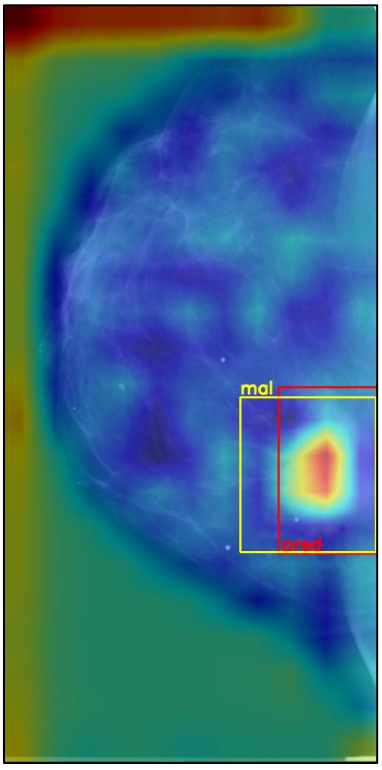
Grad-CAM/Attention Maps for Malignant Images:

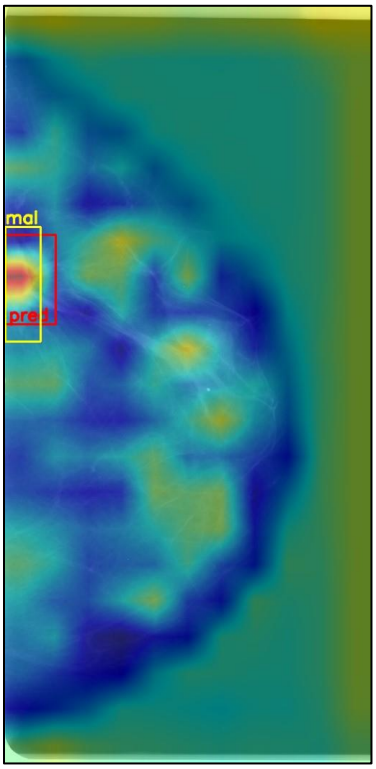
Gradient-weighted Class Activation Mapping (Grad-CAM) uses the gradients of any target concept flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept. Here, Ablation CAM, which is a gradient-free method, is used. The yellow bounding box signifies the ground truth for a malignant tumor, while the red bounding boxes denote predictions made by the detection model.

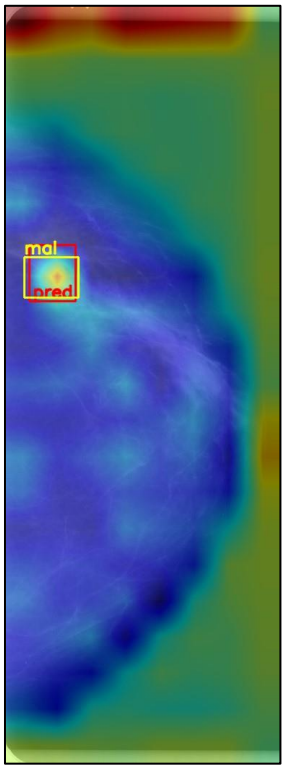


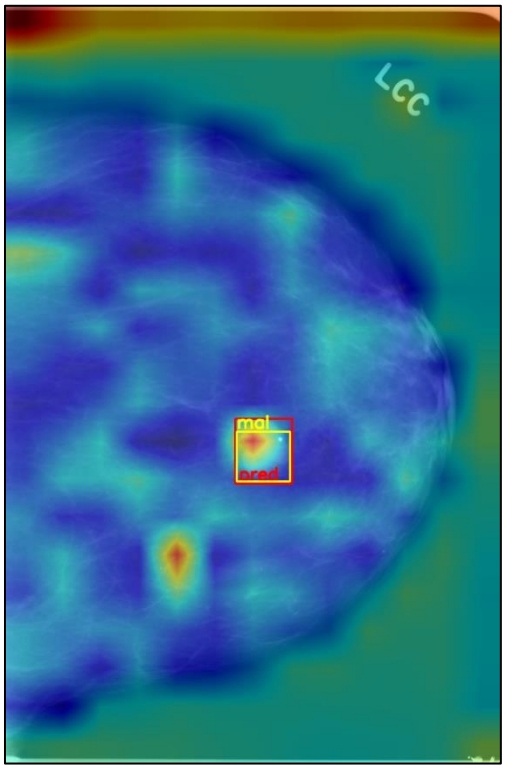

mal
pred


mal
pred


mal
pred


mal
pred


mal
pred


LCC
mal
pred

Deformable DETR (Transformer-based Model)

Library Used:

Transformers.DeformableDetrForObjectDetection

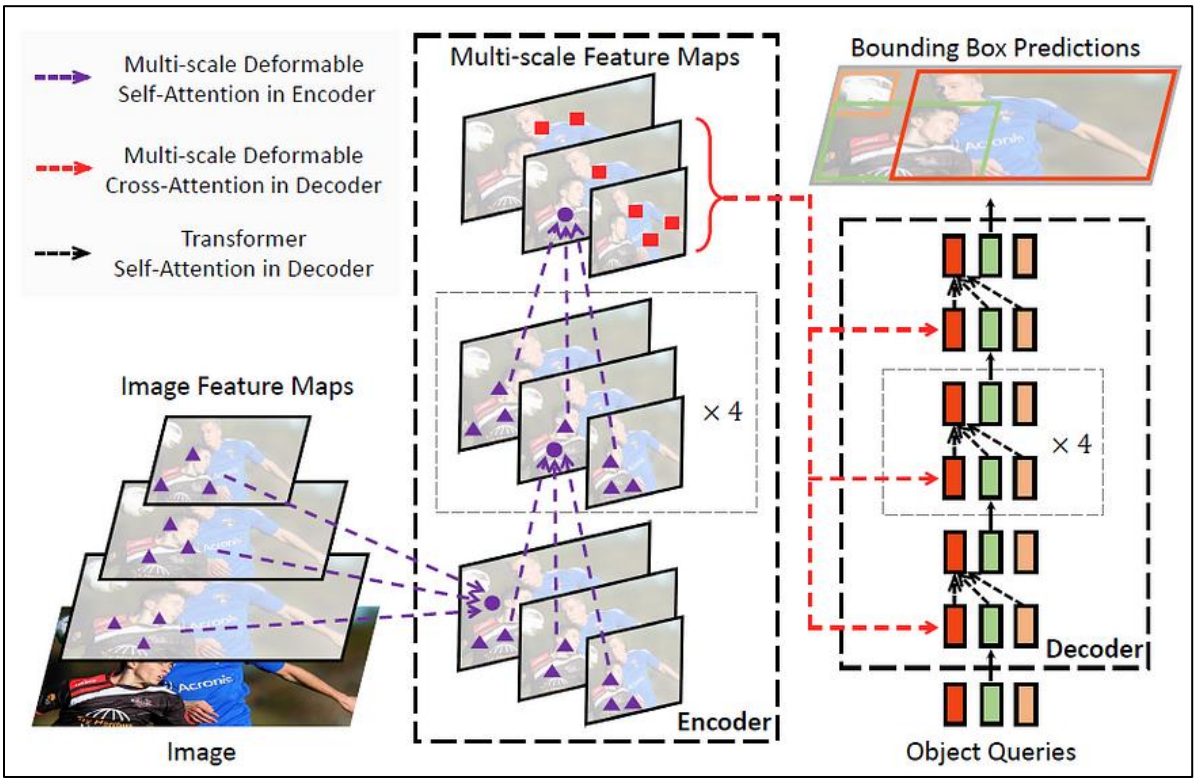
model = DeformableDetrForObjectDetection.from_pretrained("SenseTime/deformable-detr")

Loading of the Data:

1. COCO Dataset is made by reading the images and annotation json file.
2. The malignant boxes are assigned the label ‘0’ corresponding to the ‘N/A’ class present in the model.

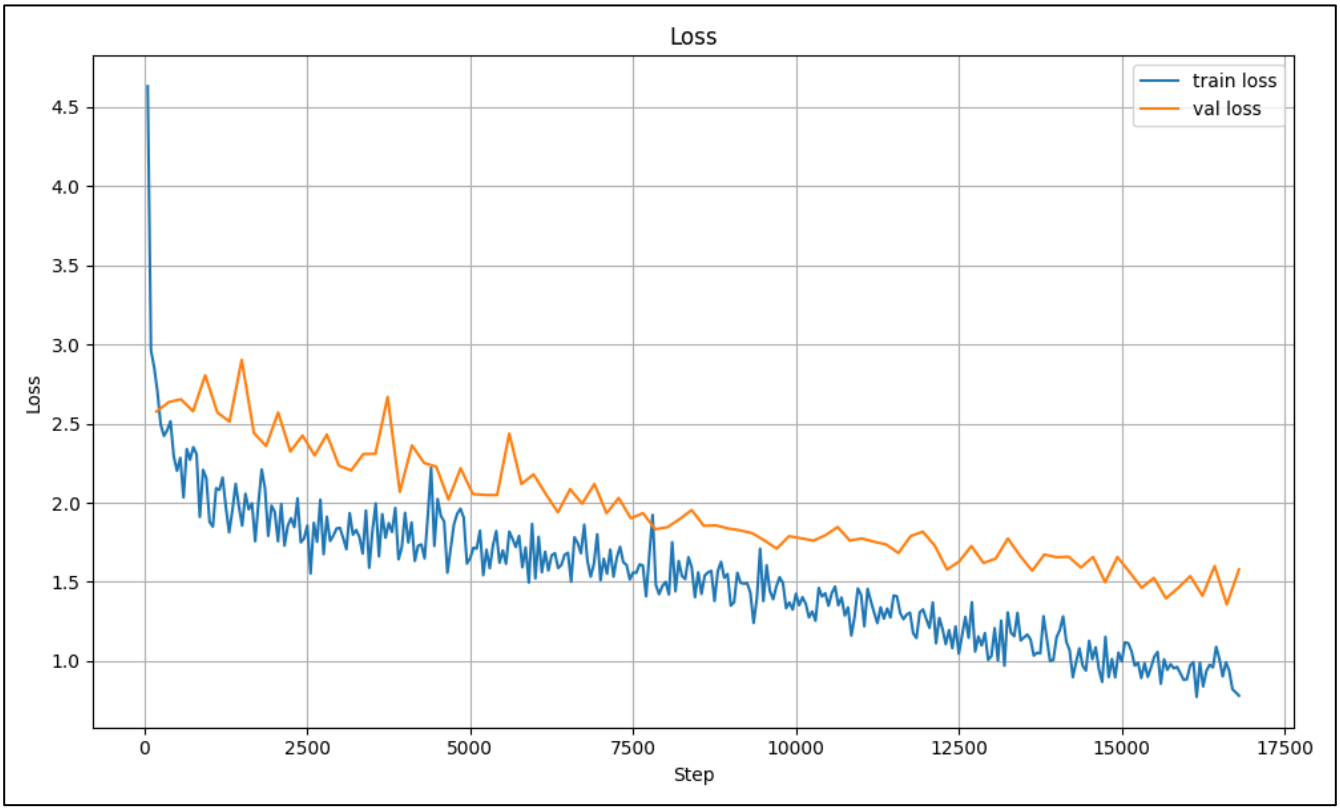
Training of the Model:

Pretrained DeformableDetrForObjectDetection model with weights "SenseTime/deformable-detr" is imported and finetuned on the given training dataset with the following training parameters:
No. of Epochs: 30
Learning Rate: 0.0001
Batch Size: 4



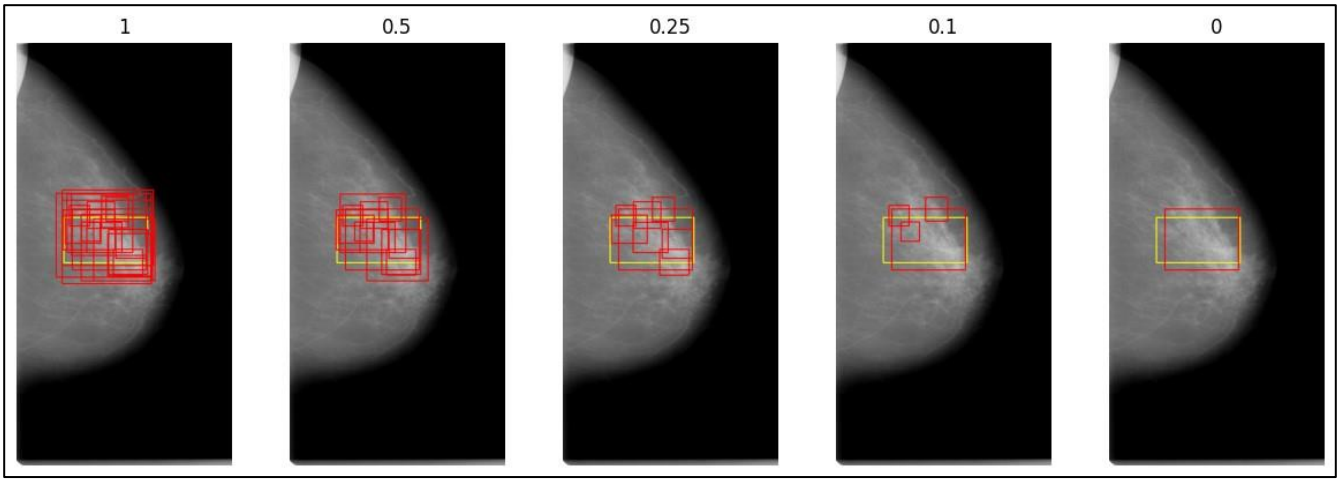
Training and Validation Loss Curves:

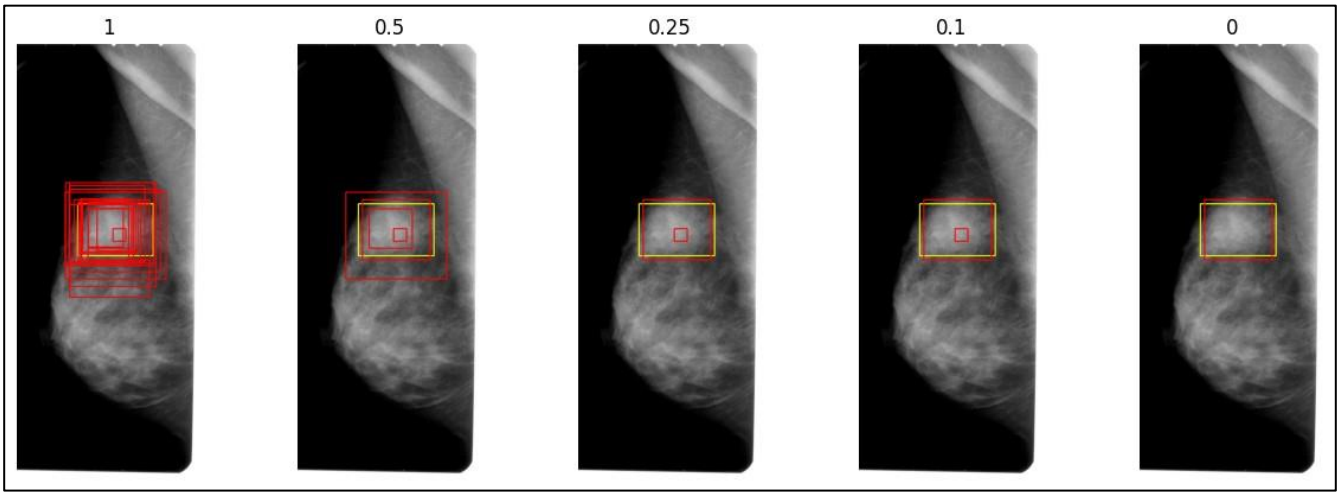
The graph below shows the plot of training and validation loss as a function of the number of epochs. It is observed that the loss value decreases with an increase in the number of epochs. This suggests that the model's performance is improving as it is trained on more data.

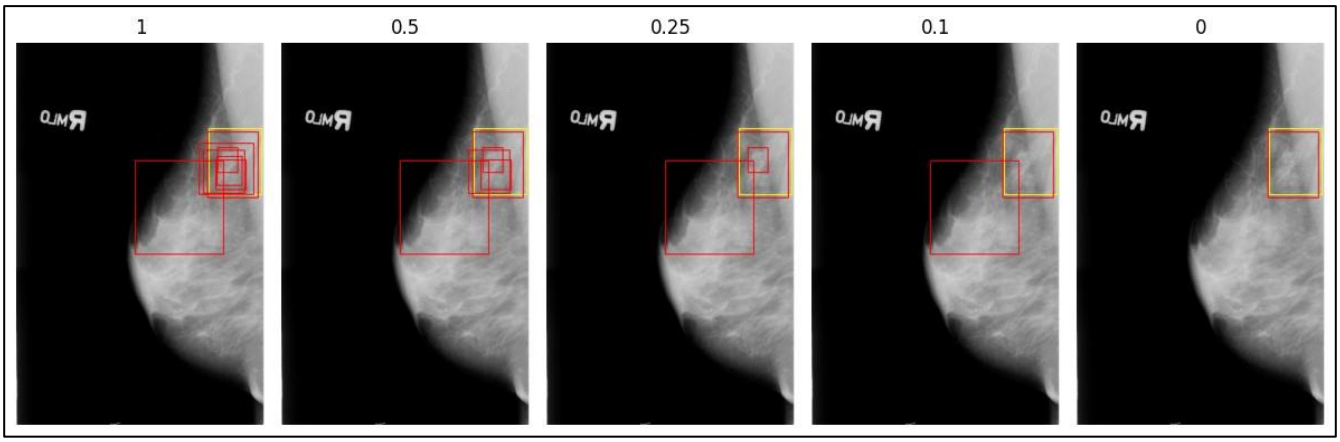


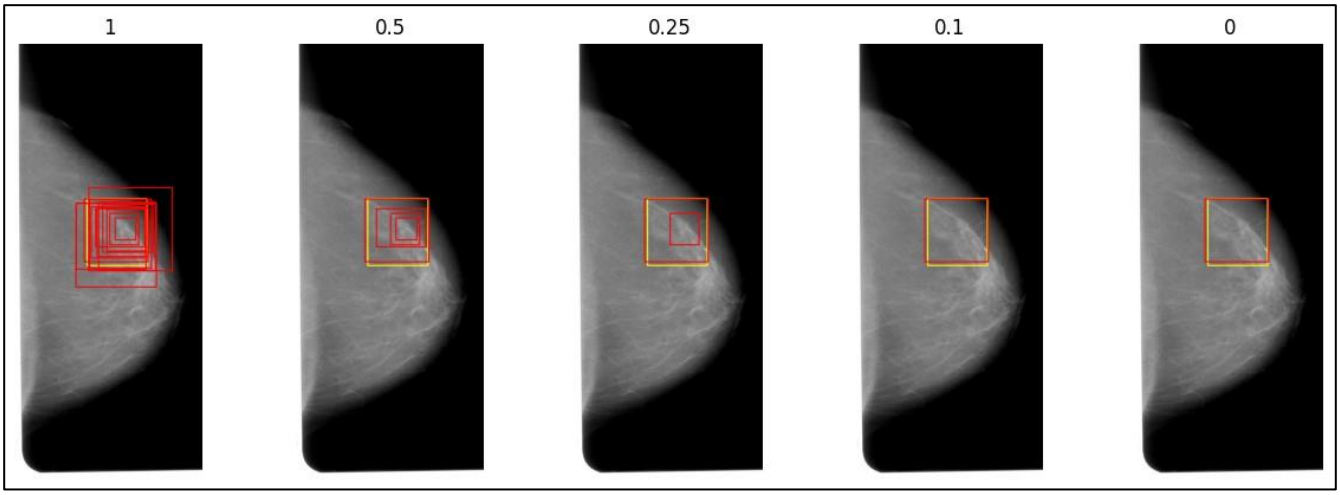
Model Predictions on Images, pre and post NMS:

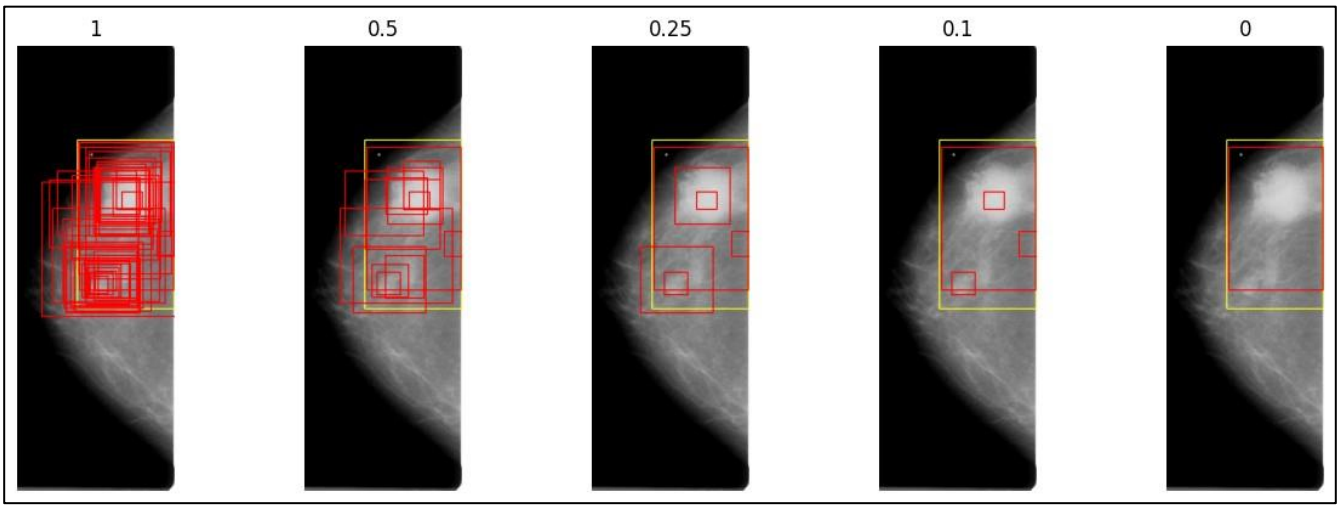
The illustration depicts the fluctuation of Non-Maximum Suppression (NMS) across the predictions generated by a detection module. The yellow bounding box signifies the ground truth for a malignant tumor, while the red bounding boxes denote predictions made by the detection model. Within this visual representation, we observe the nuanced variations in NMS. We explore five Intersections over Union (IOU) thresholds: 1, 0.5, 0.25, 0.1, and 0. A threshold of 1 indicates the inclusion of a bounding box regardless of overlaps, while a threshold of 0 signifies the removal of bounding boxes even in the presence of slight overlaps.

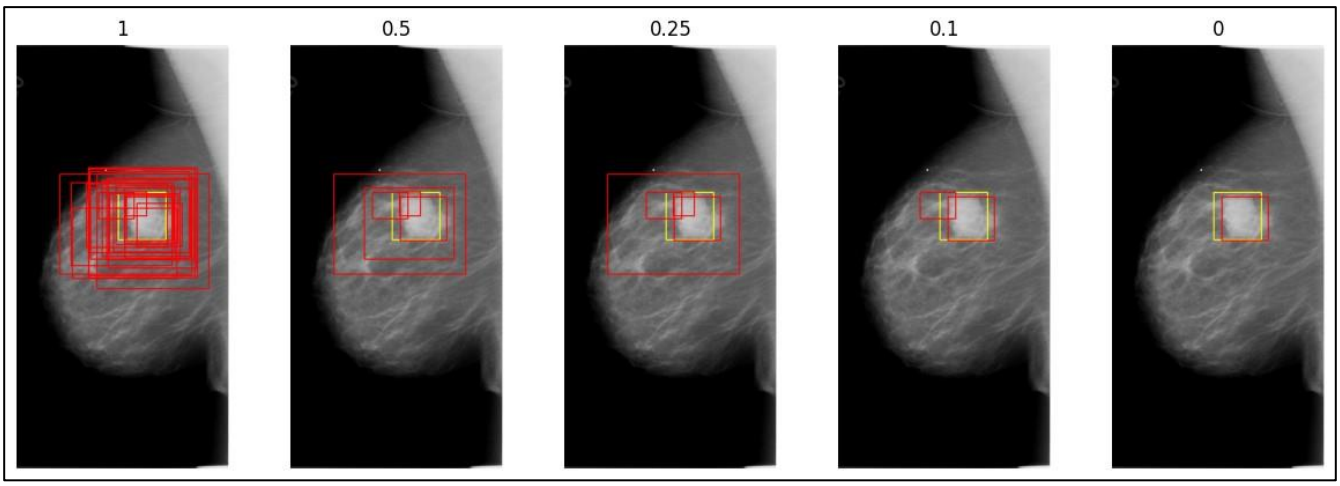

1
0.5
0.25
0.1
0

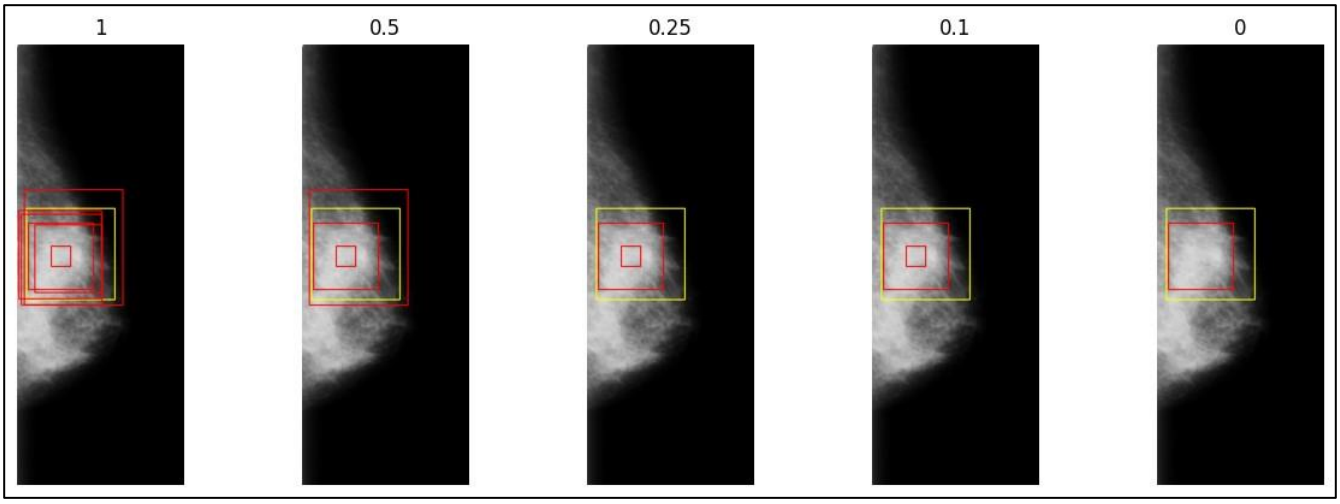

1
0.5
0.25
0.1
0

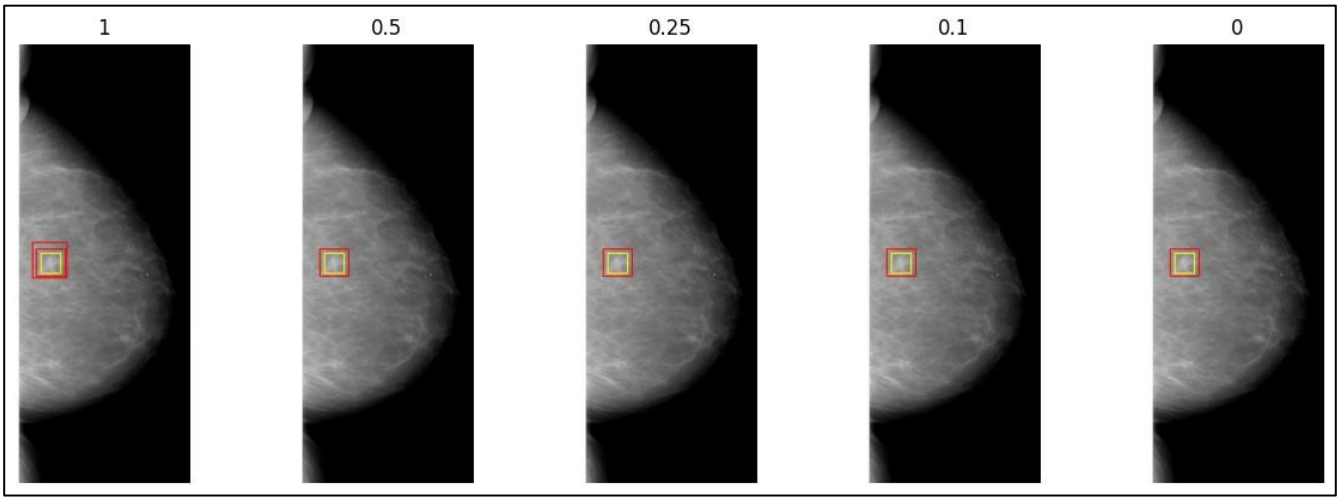

1
0.5
0.25
0.1
0

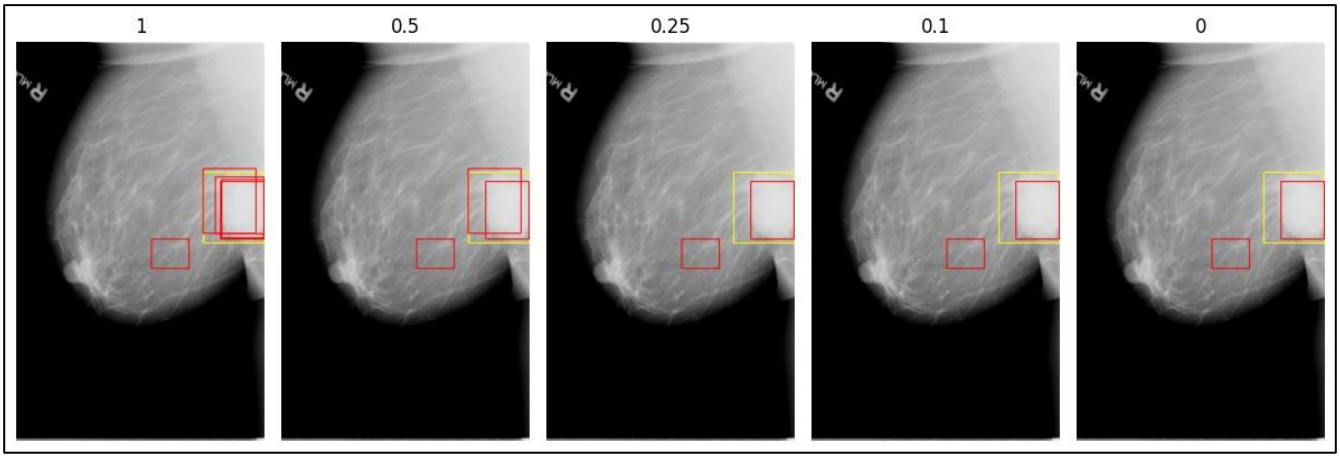

1
0.5
0.25
0.1
0

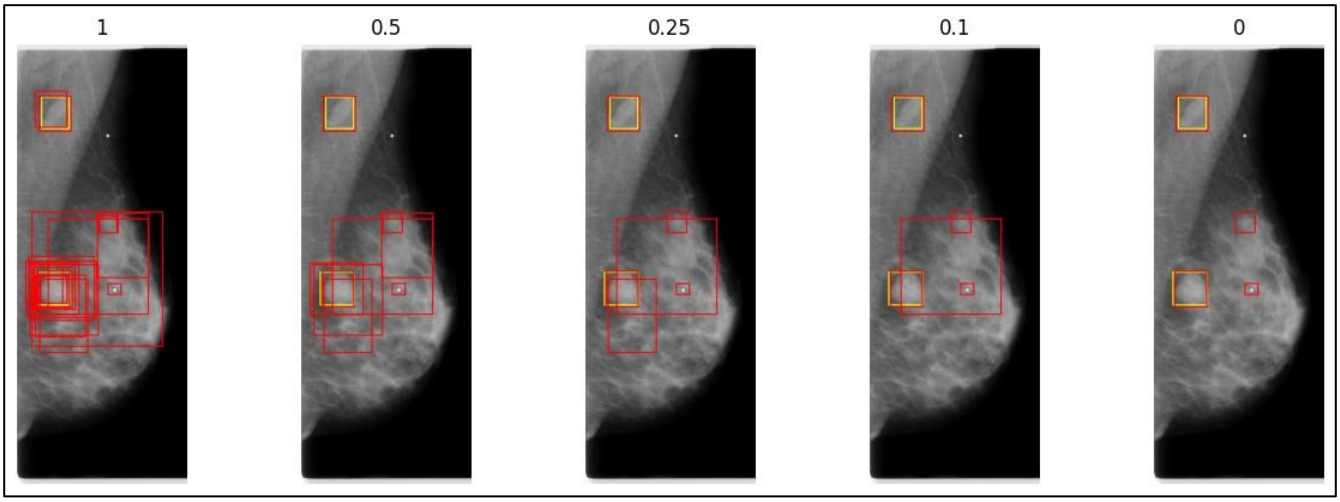

1
0.5
0.25
0.1
0


1
0.5
0.25
0.1
0


1
0.5
0.25
0.1
0

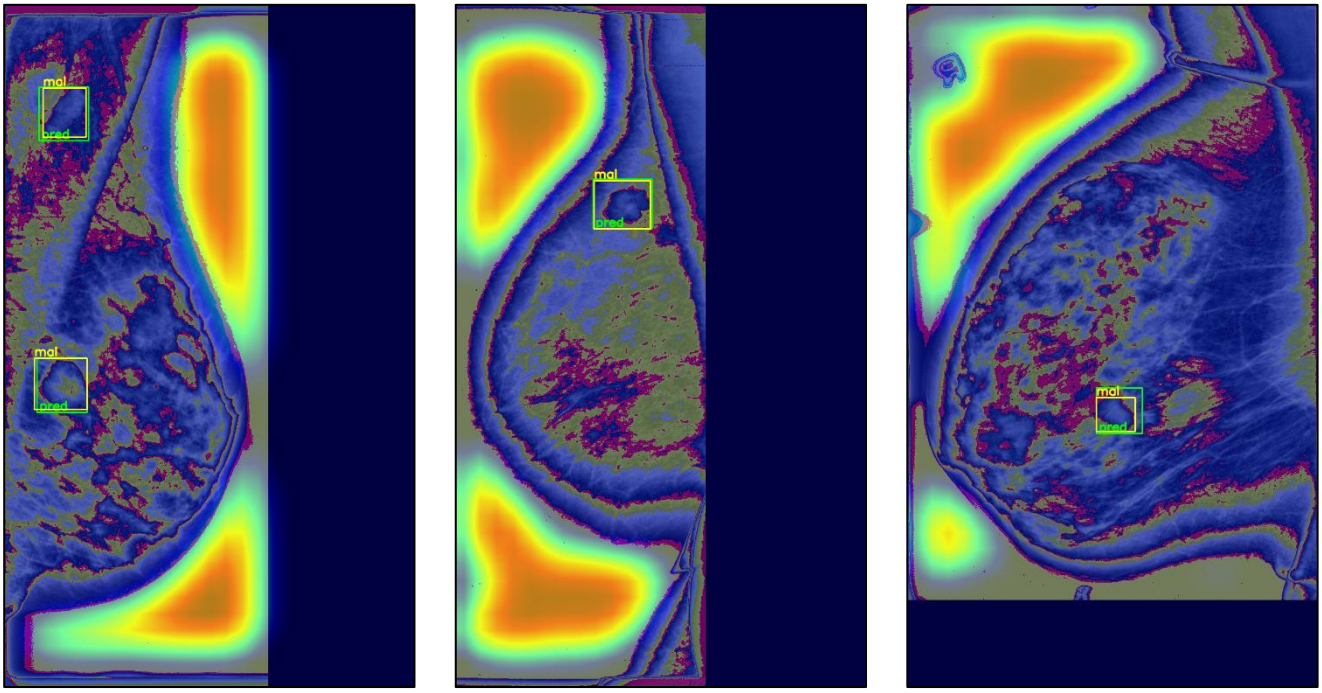

1
0.5
0.25
0.1
0

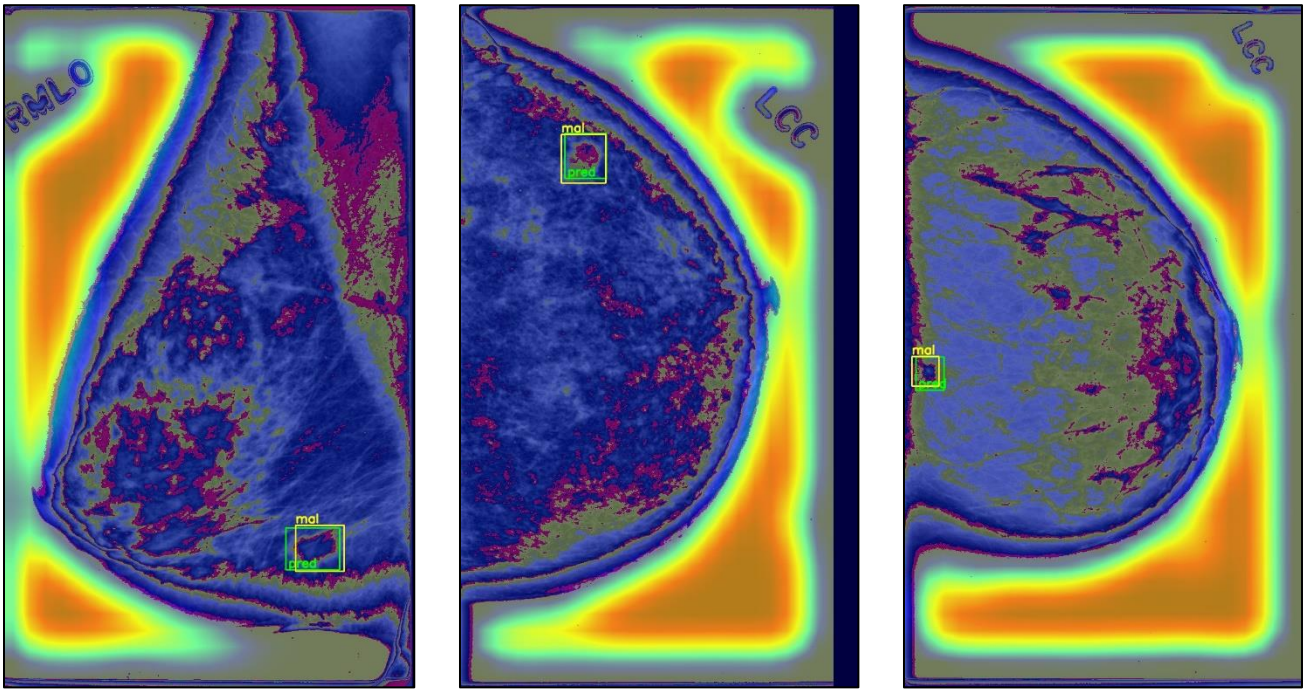

1
0.5
0.25
0.1
0

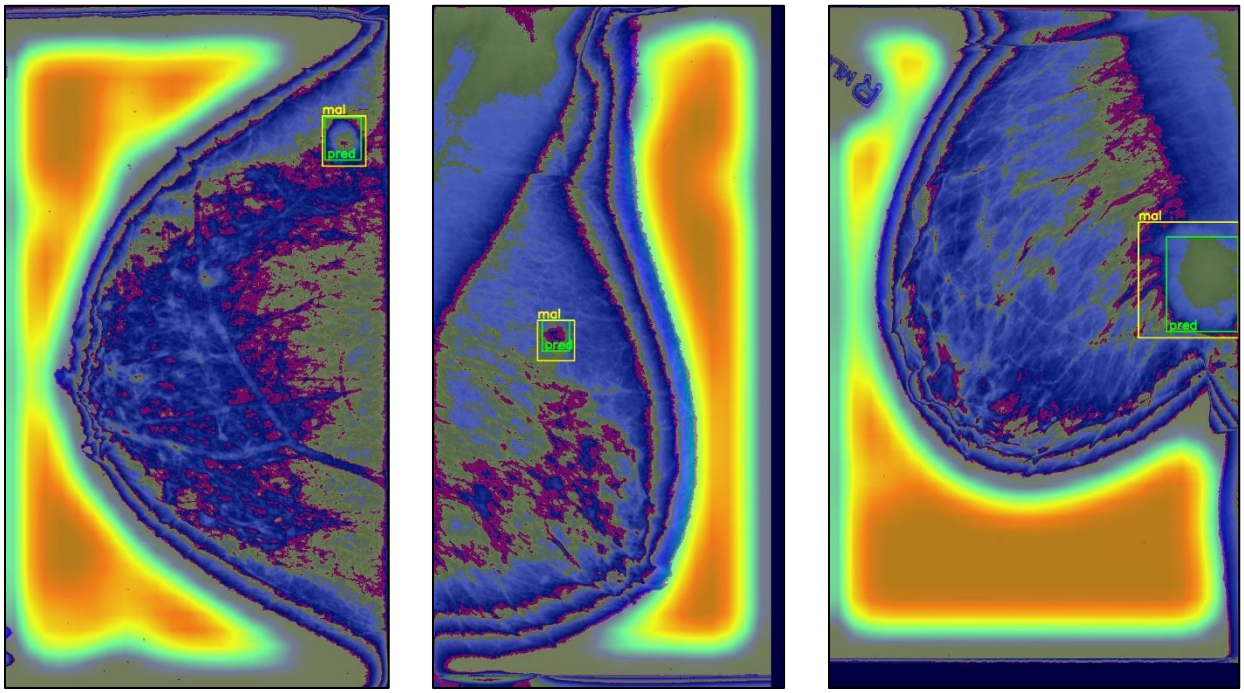


Grad-CAM/Attention Maps for Malignant Images:

Gradient-weighted Class Activation Mapping (Grad-CAM) uses the gradients of any target concept flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept. Here, Eigen CAM, which is a gradient-free method, is used. The yellow bounding box signifies the ground truth for a malignant tumor, while the green bounding boxes denote predictions made by the detection model.




RMLO
mal
pred
LCC
mal
pred
LCC
mal

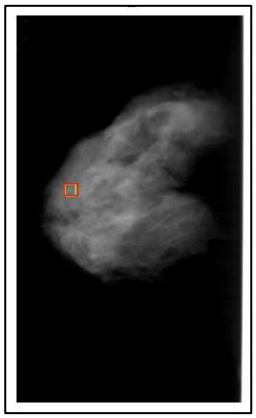

mal
pred
mal
pred
mal
pred

FROC (Free-response Receiver Operating Characteristic)

FROC plots sensitivity against average false positives per image (FPI) on the x-axis. A prediction is considered a true positive if the center of the bounding box lies within the center of the ground truth box. False positives are calculated based on all predicted bounding boxes in an image. The table below shows the Sensitivity and Threshold values at different values of FPI for the FasterRCNN and Deformable DETR models.

FPI	FasterRCNN		Deformable DETR	
	Sensitivity	Threshold	Sensitivity	Threshold
0.0	0.52587	0.98799	0.37912	0.92200
0.1	0.64346	0.98198	0.46096	0.90700
0.1	0.74318	0.96396	0.52117	0.89200
0.1	0.79680	0.93393	0.55691	0.88300
0.2	0.83349	0.86887	0.59548	0.87600
0.3	0.88335	0.52953	0.63594	0.86700
0.4	0.91345	0.21021	0.67074	0.85900
0.5	0.93509	0.08509	0.68862	0.85200
0.6	0	0	0.70743	0.84600
0.7	0	0	0.71684	0.84200
0.8	0	0	0.73189	0.83700
1.0	0	0	0.75541	0.82900
1.1	0	0	0.76105	0.82500
1.2	0	0	0.77328	0.82200
1.5	0	0	0.79210	0.81300
1.8	0	0	0.80433	0.80600
1.9	0	0	0.80997	0.80300
2.0	0	0	0.81279	0.80100
2.4	0	0	0	0

From the above table, it can be observed that at FPI = 0.3, the sensitivity of FasterRCNN is higher than that of Deformable DETR. Some of the additional cases detected by FasterRCNN model are as follows:



Calc-Training_P_00031_LEFT_CC.png

Calc-Training_P_00078_LEFT_MLO.png

References

1. Slides on Object Detection and Transformers by Prof. Chetan Arora, IIT Delhi
2. https://pytorch.org/vision/main/models/generated/torchvision.models.detection.fasterrcnn_resnet50_fpn.html#torchvision.models.detection.fasterrcnn_resnet50_fpn
3. https://github.com/NielsRogge/Transformers-Tutorials/tree/master/DETR
4. https://github.com/NielsRogge/Transformers-Tutorials/tree/master/Deformable-DETR
5. https://github.com/jacobgil/pytorch-grad-cam
6. https://doi.org/10.48550/arXiv.2010.04159