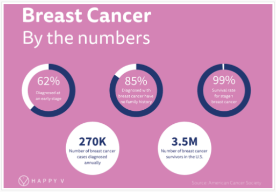
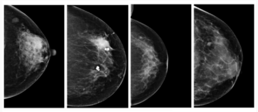
Breast Cancer Tumor Detection via Faster R-CNN

by Takara Truong & Ann Wu

1 Introduction



In breast cancer, early detection is paramount in improving patient outcome and cost. However, detection resources (such as radiologists) often do not meet the demand.

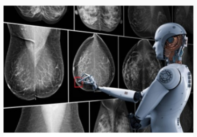


Examples of mammograms with cancer identified by AI but missed by both radiologists (left two panels) and mammograms with cancer identified by radiologists but missed by AIs (right two panels). (Courtesy: JAMA Network ©2020 American Medical Association)

AI has stepped up to meet this demand, with 2020 state-of-the-art algorithms showing performance matching Bayes error (of a human radiologist).

2 Project Goal

Our goal is to implement a CNN that takes mammography images as input, and provides binary labels of "benign" or "malignant" as output, to our best possible accuracy.



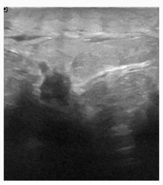
Source: https://blogs.nvidia.com/blog/2018/02/01/making-mammography-more-meaningful/

3 Dataset

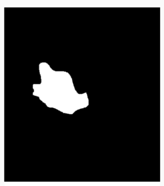
The Breast Ultrasound Images Dataset (2018) contains 780 PNGs of breast cancer ultrasound scans from women between ages 25-75 years old. Images are on average 500x500 pixels.

category	num samples
benign	437
malignant	210
normal	133

Source: https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset



BUSI dataset sample #124_img: malignant tumor ultrasound image



BUSI dataset sample #124_mask: malignant tumor mask image

Each image has a corresponding mask image of identical size, with the region of interest (i.e. tumor) represented in white pixels.

We extracted the bounding boxes from the mask images into a CSV file

Screenshot of extracted metadata.csv

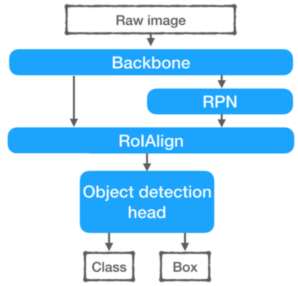
Then we shuffled the dataset, and divided it into training-validation-test datasets with a 90-5-5 split.

4 Methods

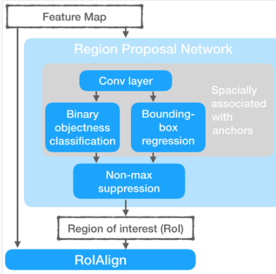
We observed in our literature review that best-in-class methods for breast cancer detection tended to have a two-stage approach:

1. Segmenting image into **regions of interest**
2. **Labeling each region of interest** with probability of existing tumor

We chose to use a **Faster R-CNN** architecture due to having a region proposal network.



Faster R-CNN architecture (https://arxiv.org/abs/1506.01497)



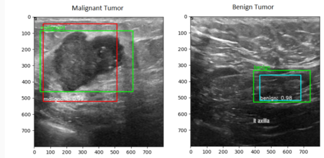
Region Proposal Network (RPN) architecture (https://arxiv.org/abs/1506.01497)

Architectural decisions:

- **Backbone**: VGG-16
- **RPN**: 22500 initial anchors --> max of 2000 proposals
- **Object Detection Head:** FC --> Softmax
- **Learning rate:** 0.0001, **Opt**: Adam, **Num epochs**: 13000
- **Loss function:**

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*)$$

5 Experiments & Analysis



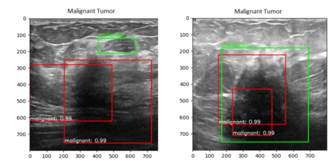
Example classifications of a malignant tumor (left) and benign tumor (right) from our model. Green is for ground truth, red is for malignant tumor classification, and blue is for benign tumor classification.

One key experiment was an ablation study to see the effect of removing the 2nd bounding box adjustment component in the object detection head (1st such component is in the RPN).

Model	Train Acc	Val Acc	Test Acc	Score
With bbox adj	50.13%	43.75%	41.16%	**43.99**
W/o bbox adj	45.24%	48.27%	43.41%	**45.72**

Experimental results: model accuracy before and after removing the 2nd bounding box adjustment component

$$score = \frac{train + 2 * val + 2 * test}{5}$$



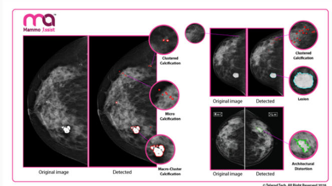
Test set showing confusing malignant ground truths for two similar images with similar regions of interests.

Upon analysis of failed labels, we discovered confusing ground truth bounding box labels that likely contributed to lowering the accuracy.

6 Conclusions & Future Work

Through this work, we learned about the complexities of achieving a well-performing RPN on greyscale ultrasound images.

Given time and resources, we would like to experiment with other benchmark datasets for breast cancer detection with our model, as well as experiment further with the architecture of the Faster R-CNN components!



Source: https://www.biospectrumindia.com/news/74/11942/telerad-tech-launches-ai-product-for-early-stage-breast-cancer-detection.html