Model Card and App Info

Revision Record

App name: Mammogram Mass Analyzer
Model names: exp50_best.pt and exp51_best.pt

App deployment status: Prototype
Version: 0.0
Date: 18-Nov-2022
Created by: vbookshelf
Notes: Released for demonstration.

Known Issues

1- The latency (prediction time) is around 10 seconds per image. This is because the CPU is being used for inference. This is also because two models are being ensembled during inference. Inference would be faster with a GPU.

-x-

Purpose

This desktop app uses computer vision to detect masses on mammograms. It takes full field digital mammograms in dicom format as input. It analyzes each mammogram and then displays an image with bounding boxes drawn around any masses. The predictions are made by two ensembled Yolov5l models. They were fine tuned on data from the VinDr-Mammo dataset.

Patient Data Security

This is a desktop application.
- Patient data never leaves the user’s pc or laptop.
- There's no tracking.
- The code is fully accessible and therefore auditable for malware.

Input

The app accepts mammograms only in dicom format. Multiple files can be submitted at the same time.

Output

The app outputs images with bounding boxes drawn around the detected masses.

VinDr-Mammo Dataset Summary

- A large-scale full-field digital mammography dataset of 5,000 four-view exams
- 20,486 total mammograms in dicom format
- 18,232 mammograms contain no abnormal regions.
- 2254 mammograms contain abnormal findings.
- Abnormal findings include: mass, calcification, asymmetries, architectural distortion and others.
- Only findings with BI-RADS 3, 4, or 5 are annotated with bounding boxes.
- Contains many Density C and Density D images.

Paper:
VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography
https://arxiv.org/abs/2203.11205

Dataset on Physionet:
https://physionet.org/content/vindr-mammo/1.0.0/

Creating the train and val datasets

To create the training and validation datasets I used only 1426 mammograms from the VinDR-Mammo dataset. This was made up of:

- All mammograms that contain masses (1226 images)
- 200 normal images

The train test split was approx. 90/10.
Total train images: 1280
Total val images: 146

Train target distribution:
Mass: 1100
Normal: 180

Val target distribution:
Mass: 126
Normal: 20

Train density distribution:
DENSITY C 1035
DENSITY B 158
DENSITY D 81
DENSITY A 6

```
Val density distribution:
DENSITY C    113
DENSITY D     17
DENSITY B     16
```

Validation Results

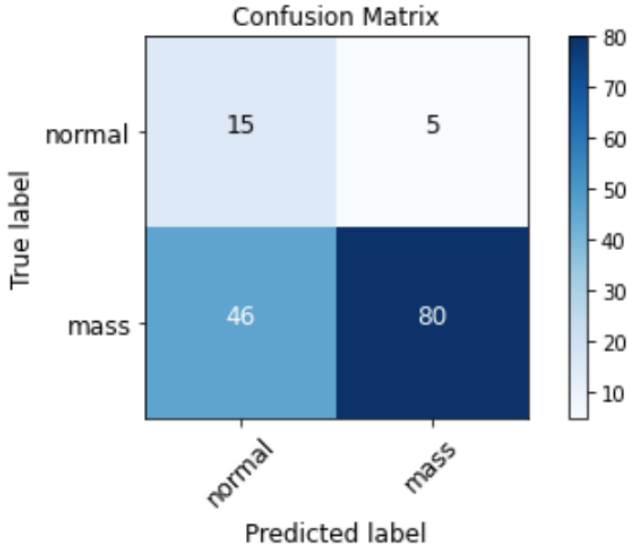
These are the results produced by two ensembled Yolov5l models. Both models were fine tuned on the same training data. A different augmentation setup was used for each.

```
Val accuracy: 0.65
Val recall: 0.63
```

The val dataset contained a large number of mammograms with Density C. These are images of high density breasts that may be challenging for a human to analyze. When assessing the model's quality, this is an important point to keep in mind. At first glance, an accuracy of 0.65 may not look exciting, but if this score equals or surpasses what a human radiologist could achieve on high density breast images, then this model's performance is actually very good.

Guideline used to calculate accuracy:
If the model correctly detected a mass on an image, and the IOU between the true and predicted bounding boxes was greater than 0, then the model was deemed to be correct. It's not important for the app to precisely annotate masses i.e. draw good bounding boxes. Its main purpose is to guide the eye of the radiologist to a region of interest.

Confusion Matrix



Classification Report

	precision	recall	f1-score	support
normal	0.25	0.75	0.37	20
mass	0.94	0.63	0.76	126
accuracy			0.65	146
macro avg	0.59	0.69	0.56	146
weighted avg	0.85	0.65	0.71	146

Hardware

- 2x RTX A5000 GPUs
- Trained on vast.ai

Misc Info

1- The model was trained using full field digital mammography images (FFDM). If a user submits screen-film mammography images (SFM), the model could produce poor results.

2- In practice users could submit images of varying quality. These images could have been taken using different types of x-ray machines or even be photos of x-ray films taken with cellphone cameras. These and other real-world factors could reduce the accuracy of the model.

3- The app can easily be configured to use a GPU.

4- I used data augmentation and model ensembling to improve accuracy and reduce overfitting.

Documentation

All code is available on GitHub:
https://github.com/vbookshelf/Mammogram-Mass-Analyzer

License

Contact

Email: contact -at- woza -dot- work