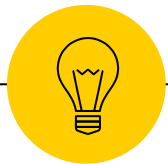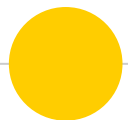# Segmentation of cancerous breast sonograms through the utilization of U-Net

Daral Maesincee (6437982 EGBE/M)
EGBE601 Medical Signal Processing and Instrumentation

Significance

- 7.8 million women worldwide are living with breast cancer → growing prevalence
- One of the leading causes of death among women → 685,000 deaths in 2020
- ***Ultrasonography for breast cancer***
  - Non-invasive, relatively harmless diagnostic tool
  - Exploration of suspicious lesions detected by preliminary physical examinations/mammography
  - Identification of different breast cancer types
  - *Problem*: unsuitable screening technique
    - Inability to produce clear images of deeper regions and differentiate early signs of cancer
- ***Goal:*** utilize signal processing to segment various breast sonograms for improved prediction accuracy of the disease through the convolutional network architecture, U-Net
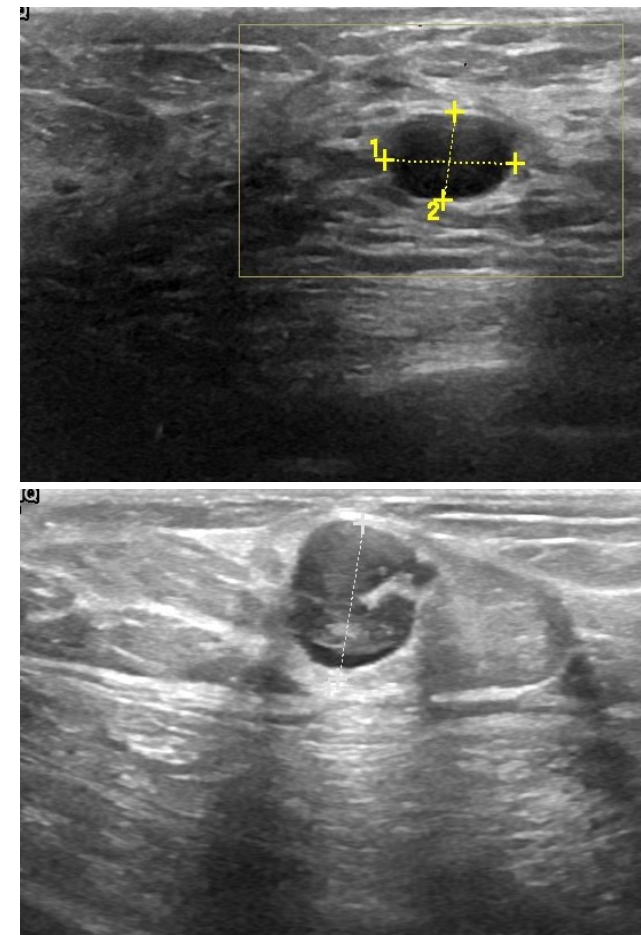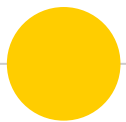
U-Net
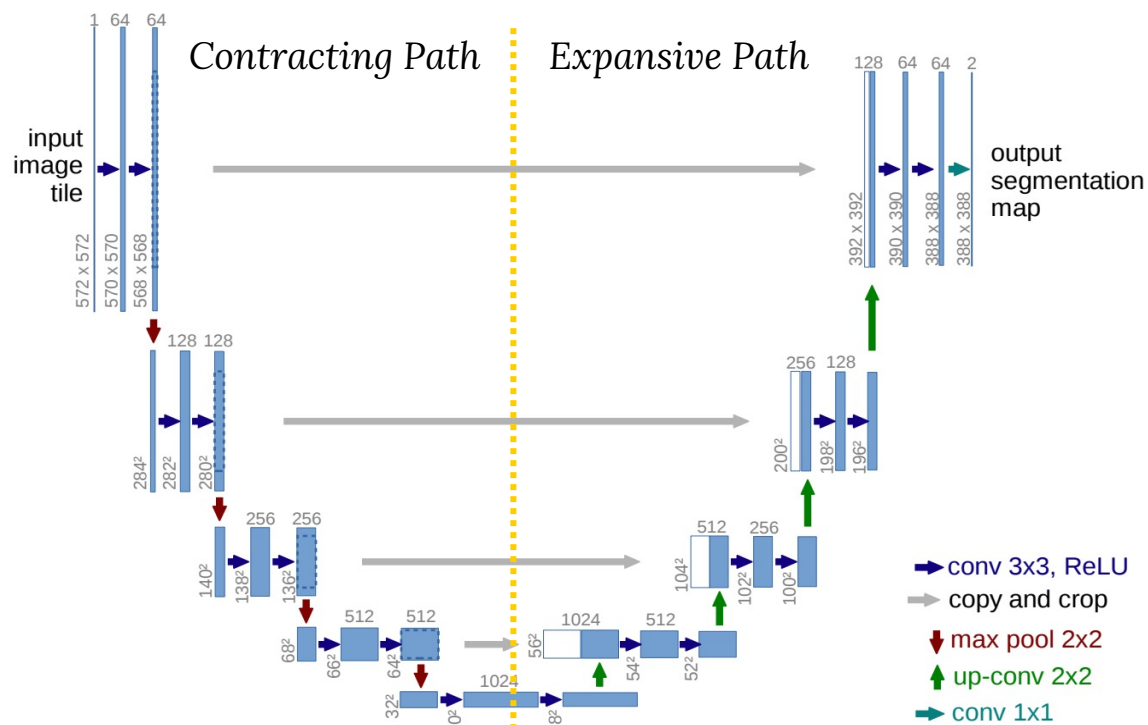
- "Fully convolutional network" by Long, Shelhamer, and Darrell → U-Net architecture
- ***Purpose***: semantic (image) segmentation
- ***Applications***: medical imaging, object detection, biometric recognition, surveillance system
- U-Net = contracting path + expansive path
  - ***Contracting path***
    - Input → two 3x3 convolutions → rectified linear unit (ReLU) → 2x2 maxpooling operation with stride 2 → downsampling → x2 number of feature channels
  - ***Expansive path***
    - Upsampling of feature map → 2x2 convolution → 1/2 number of feature channels → concatenation → two 3x3 convolutions → ReLU → 1x1 convolution → output

Tools & Libraries

- NumPy
- Pandas
- Matplotlib
- Tensorflow
- Keras
- Scikit-Image

Dataset

- Source: Kaggle
- 600 female patients aged 25 – 75 years → 780 breast ultrasound images
- Average image size: 500*500 pixels
- Directories: *images + masks*
  - Normal (266 files)
  - Benign (891 files) } *train + test*
  - Malignant (421 files)
- Creating new directories:
  - benign.png + malignant.png → Images
  - benign_mask.png + malignant_mask.png → Masks

Dataset_BUSI_with_GT (3 directories)

About this directory

The data reviews the medical images of breast cancer using ultrasound scan. Breast Ultrasound Dataset is categorized into three classes: normal, benign, and malignant images. Breast ultrasound images can produce great results in classification, detection, and segmentation of breast cancer when combined with machine learning.
GT==Ground Truth, present at the bottom of each image.

benign
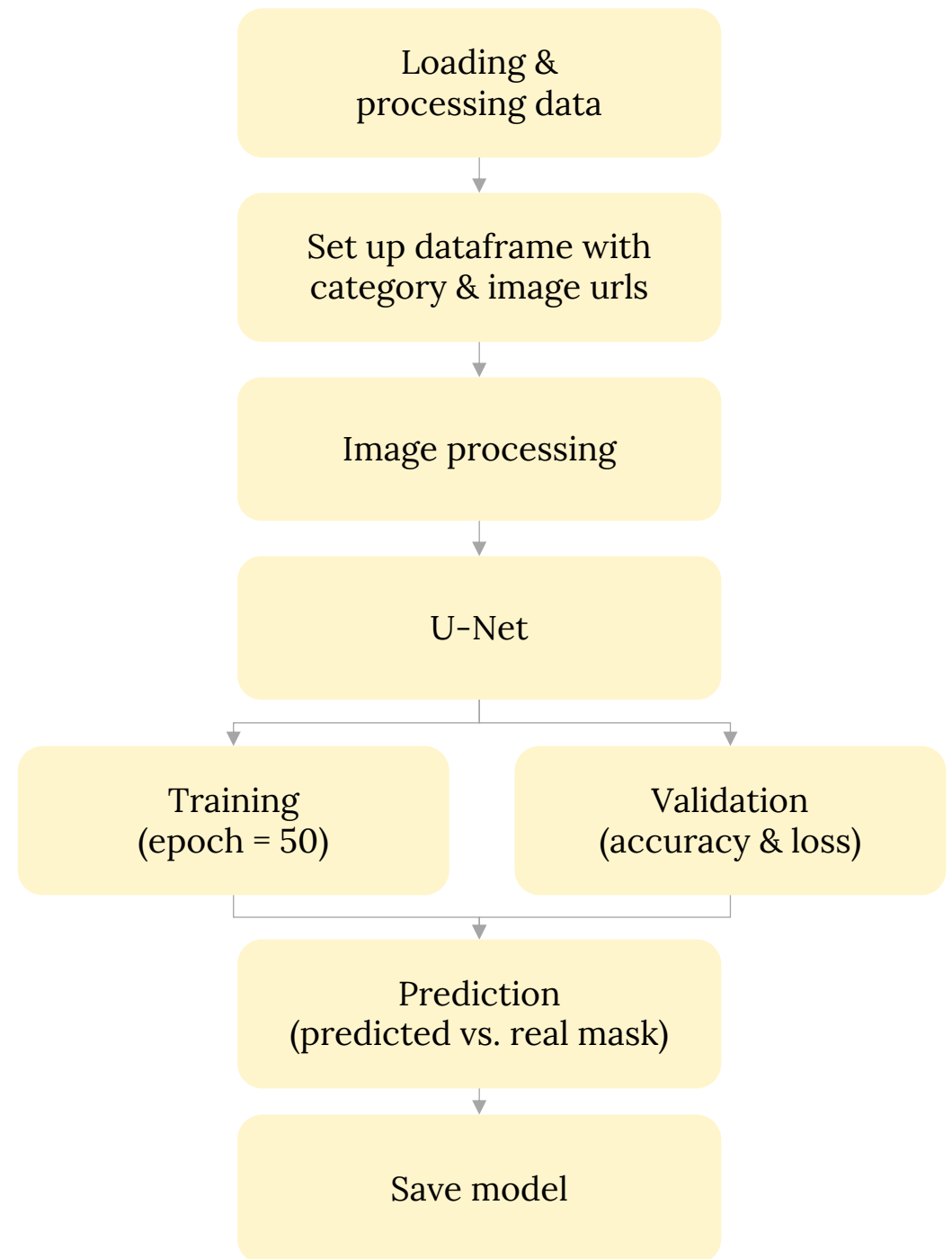891 files

malignant
421 files

normal
266 files

benign (1).png
334.35 kB LEFT BREAST

benign (1)_mask.png
240 B

malignant (103).png
348.93 kB

malignant (103)_mask....
1 kB

Training Dataset

Loading & processing data

↓

Set up dataframe with category & image urls

↓

Image processing

↓

U-Net

↓

Training (epoch = 50)

Validation (accuracy & loss)

↓

Prediction (predicted vs. real mask)

↓

Save model
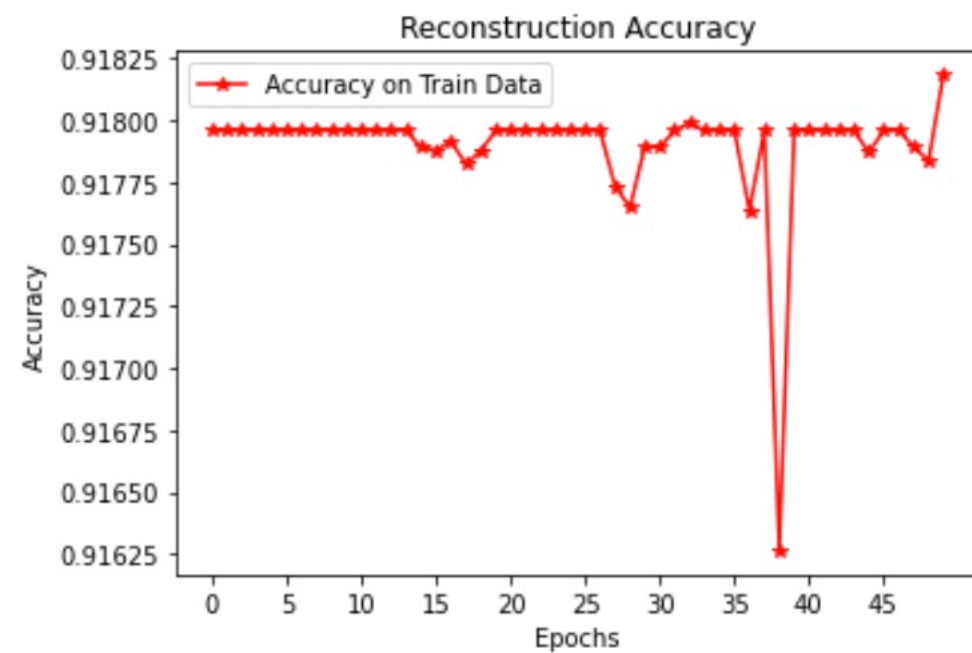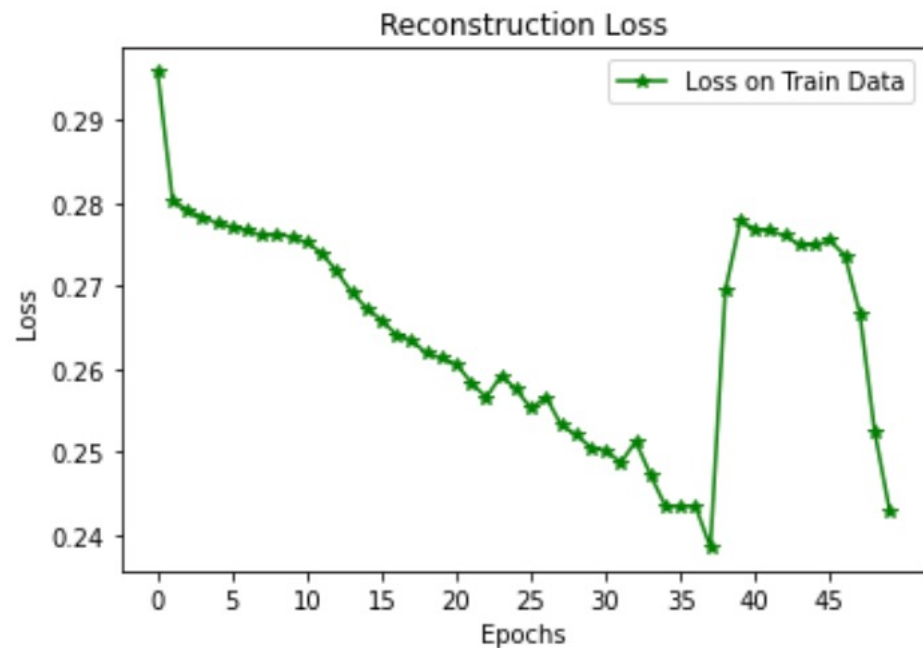
Epoch 50/50
539/539 [==============================] - 825s 2s/step - loss: 0.2429 - accuracy: 0.9182 - auc_1: 0.9553

Testing Dataset

Loading &
processing data

↓

Setup dataframe with
category & image urls

↓

Image processing

↓

Saved model

↓

Validation

↓

Prediction
(predicted vs. real mask)
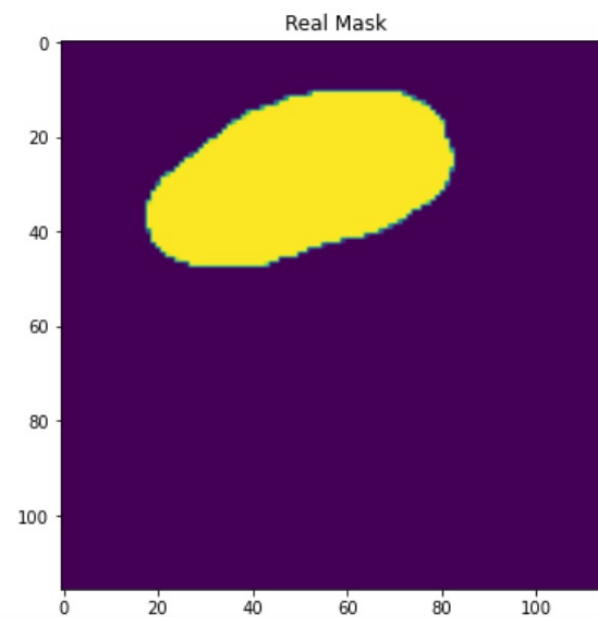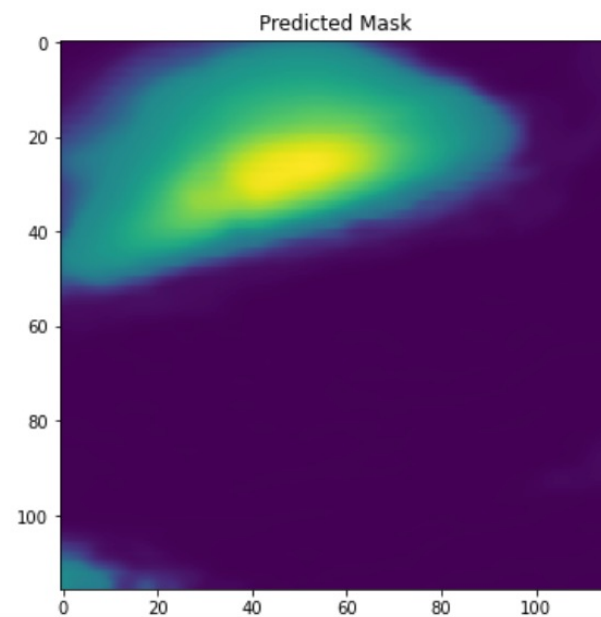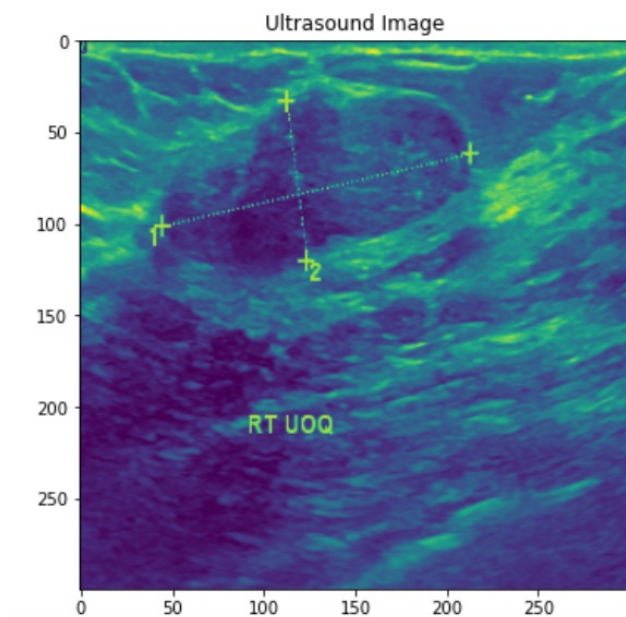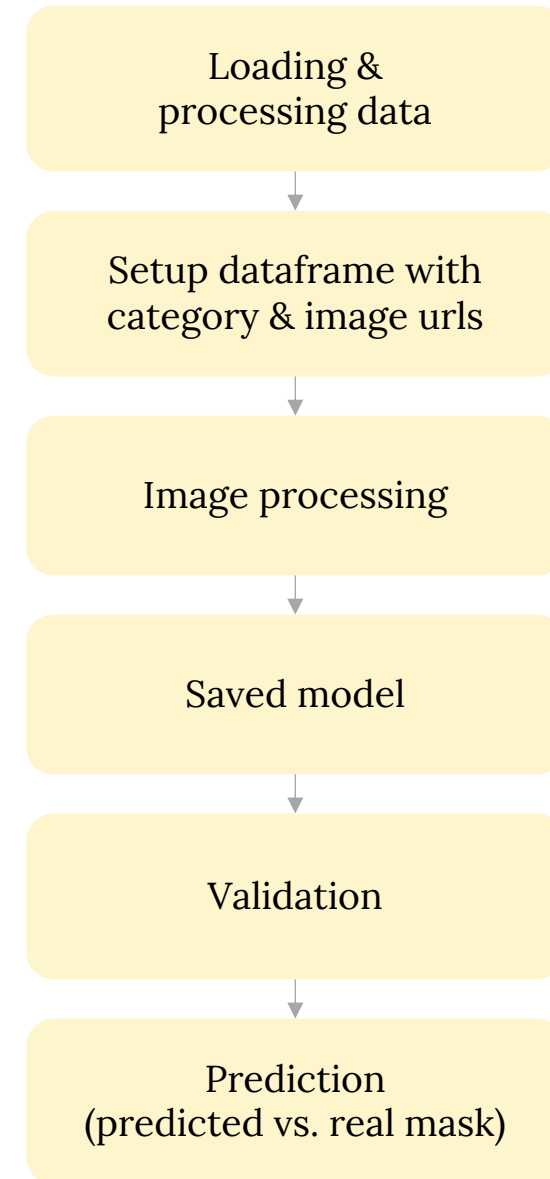
```
In [108]: new_model.evaluate(files_ds)

82/82 [==============================] - 33s 397ms/step - loss: 0.3815 - accuracy: 0.8628 - auc_1: 0.9045

Out[108]: [0.38149309158325195, 0.8628021478652954, 0.9044783115386963]
```
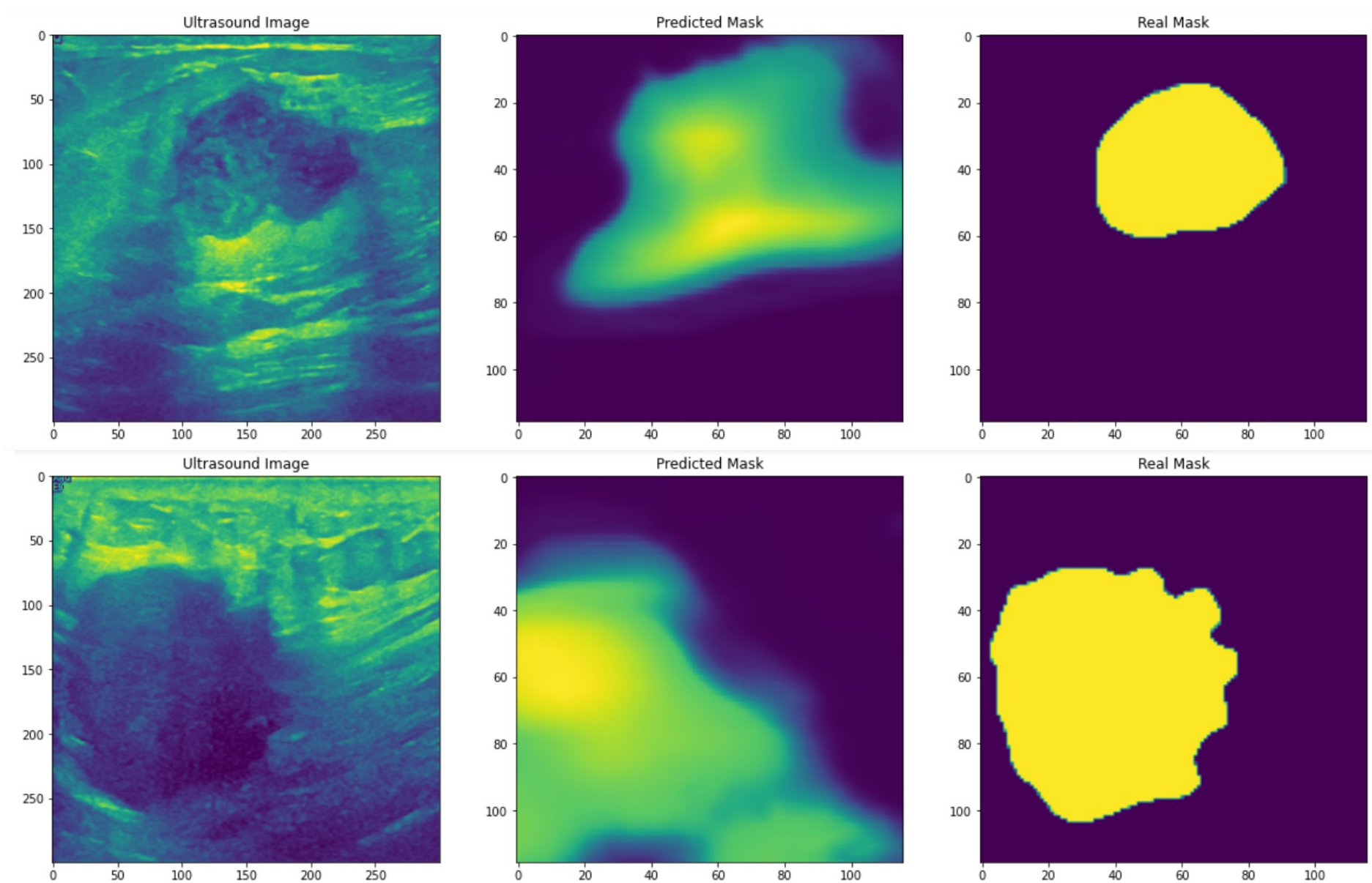
References

⊙ Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief. 2020 Feb;28:104863. DOI: 10.1016/j.dib.2019.104863.

⊙ *Breast cancer*. Who.int. (2021). Retrieved 12 February 2022, from https://www.who.int/news-room/fact-sheets/detail/breastcancer#:~:text=In%202020%2C%20there%20were%202.3, the%20world's%20most%20prevalent%20cancer.

⊙ *U-Net*. Papers With Code. Retrieved 23 April 2022, from https://paperswithcode.com/method/u-net.