Image Segmentation of Breast Cancer Ultrasounds using U-Net Architecture

Michelle Cheng, Kandasamy Illanko and Naimul Khan
Department of Electrical, Computer and Biomedical Engineering
Toronto Metropolitan University
Toronto, Canada
michelle1.cheng@torontomu.ca, killanki@torontomu.ca, n77khan@torontomu.ca

***Abstract*—According to Canadian Cancer Statistics Advisory Committee, breast cancer is the second most common cancer in Canada and 1 in 8 Canadian women will develop breast cancer [12]. Ultrasound imaging is often used for breast cancer patients during the early diagnostic stages when anomalies are found. Analysis of these Ultrasounds can be a difficult task due to the low contrast, high speckle noise and shadows present in ultrasound Imaging. Image segmentation can aid healthcare professionals by extracting a region of interest (ROI) that could potentially be tumorous. A Convolutional Neural Networks (CNN) is a neural network used for image processing. The U-Net architecture is a fully convolutional neural network used in biomedical image segmentation. Automating the image segmentation of breast cancer ultrasounds can support healthcare professionals with image screening and improve early cancer diagnosis for patients. This report highlights the implementation of the U-Net architecture to perform image segmentation on a dataset of ultrasound images consisting of benign, malignant and normal breast ultrasounds. The final model and performance metrics show promising results with accurate image segmentation of tumors.**

***Index Terms*—Image Segmentation, Convolutional Neural Networks, U-Net, Ultrasound Imaging**

I. Introduction and Literature Review

According to Canadian Cancer Statistics Advisory Committee, breast cancer is the second most common cancer in Canada and 1 in 8 Canadian women will develop breast cancer [12]. Early detection of breast cancer cases is incredibly important and several imaging modalities may be used at various stages. One of these is ultrasound imaging, usually used alongside mammograms. Ultrasound imaging also known as a sonography uses high-frequency sound waves to produce an image of various structures in a human body. The transducer emits a high-frequency sound wave which travels through the body tissues. When encountered with different tissues, it is reflected back to the transducer producing the image seen. Ultrasounds are non-invasive and emit lower radiation compared to other imaging modalities. However, the limited image quality and high noise capture, is a downfall which can make early detection challenging. Success rate of detecting tumors with ultrasound can change depending on the size, depth, type, location of the tumor and skill of the operator and doctor. The rate of when cancerous cells are detected is highly correlated to the survival rate with early detection usually improving survival rates among patients.

Over the past decade, the addition of convolutional neural networks (CNNs) have proven to be a valuable tool in medical imaging. In particular, it is widely used in medical image segmentation where images can be partitioned into regions of interest (ROI) to aid healthcare professionals process various characteristics of an image. This may include information on lesions, organs, vascular diseased and tumor detection. The segmented image can help doctors with early detection, treatment and monitoring of tumors overtime. The U-Net architecture achieves this goal as different encoder and decoder layers separate and locate important features in the image. This gives the architecture its U-shape which corresponds to the down sampling and up sampling,

Continuous advancements in medical imaging have improved the accuracy and efficiency of image segmentation within ultrasound applications. Many research studies have been focused on refining segmentation using CNNs and U-Net architectures to solve challenges such as noise, quality and feature extraction capabilities. One study at Shandong University, automated tumor detection for breast cancer ultrasound by utilizing the fuzzy entropy principles and watershed segmentation to handle low-quality images and high speckle noise effectively [1]. Other preprocessing techniques such as edge detection have proven to also improve the results of a U-Net architecture by increasing the dice similarity coefficient [2]. Edge detection increases the feature extraction to help distinguish between regions.

Modifications to the U-Net model, as explored by Technical University of Cluj-Napoca, demonstrated that adjusting convolutional layers, normalization, and regularization could significantly boost accuracy in ultrasound image segmentation. Their enhanced model achieved high accuracy rates on both training and testing datasets [3]. INESC TEC's study on gynecological ultrasound images revealed that increasing U-Net's depth, optimizing filter sizes, and adjusting pooling operations could improve segmentation performance. Their findings underscored the importance of data augmentation and layer modifications for better accuracy in challenging ultrasound images.

The success of the U-net architecture can also be seen through studies that involve other medical imaging modalities such as X-ray and MRI. A study from Jiangsu University applied a U-Net-based approach to segment lung X-rays of COVID-19 patients, focusing on feature extraction and classification using SoftMax. The model effectively combined feature layers for precise segmentation [4]. In brain tumor detection, Romaiah Institute of Technology combined CNN with U-Net, achieving a remarkable 98% accuracy in MRI image segmentation. Their approach utilized CNN for initial tumor pattern detection and U-Net for refining segmentation results [5].

Alexandria University's research integrated a ConvMixer module with U-Net, reducing parameter count and improving segmentation accuracy. This hybrid model effectively managed spatial and channel information with fewer parameters, highlighting a more efficient approach to medical image segmentation [6].Further innovation was explored through CNN-Transformer hybrids, which balance local and global context capture. This approach, combining ConvMixer and Progressive Atrous Spatial Pyramid Pooling (PASPP) modules, enhanced feature extraction and segmentation accuracy by integrating CNN's local detail capture with Transformer's global context understanding [7].The ConvmixFormer-U-Net model combined local context capture from ConvMixer with global context from Transformers, resulting in improved performance and reduced parameter count. This model demonstrated effectiveness across multiple datasets, offering a robust solution for medical image segmentation [8].These studies collectively advance the field of medical image segmentation by addressing various challenges and proposing innovative solutions to enhance accuracy and efficiency in detecting and analyzing medical conditions.

II. Problem Statement and Dataset

The objective of this paper is to perform image segmentation on a series of ultrasound images including benign breast tumors, malignant breast tumors and normal breast ultrasounds. Ultrasound imaging has low contrast and high speckle noise which can make detection of the anomalies hard to analyze. Medical image segmentation can process the characteristics of an image by extracting the ROI from the rest of the image. This in return, can improve the diagnostic and screening process for patients. To achieve this, a CNN is used for finding classes and patterns within image data. The U-Net architecture is a CNN that is famous in biomedical imaging segmentation including, its ability to handle inputs with various size and ratios. During the down sampling, the encoder consists of a series of convolutional and pooling layers that reduce the spatial dimensions while increase feature maps. During the up sampling, a decoder consists of transposed convolutions to up sample the feature to the original image size. This produces the image segmentation map.

The Breast Cancer Ultrasound Image Dataset is a open source dataset created by Rudranarayan Baral. It contains a variety of ultrasound images of breast cancer patients which may come in various forms. This includes breast lesions such as ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), fibroadenomas, cysts and more. Each image in the dataset is labeled and is split between the following directories below and each tumor has a corresponding target mask. The mask separates the ROI indicated in white as depicted in *Fig.1 Benign tumor with Mask Target.* Some images may have more than one associated mask which then is indicated by the label ending with "mask_1". An example is provided in *Fig.2 Benign tumor image with multiple target mask from the dataset.*

- Benign - 437 benign tumor patients
- Malignant - 210 malignant tumor patients
- Normal -133 total normal patients images

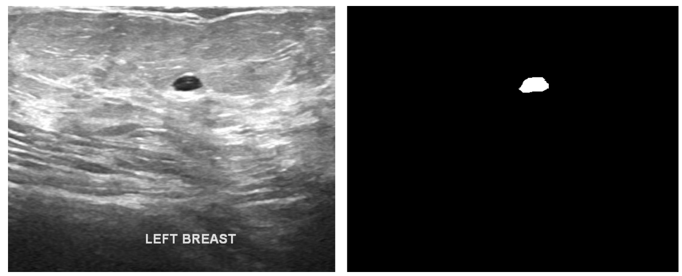

Fig. 1. Benign Tumor Image (1) with Corresponding Target Mask from the Dataset.

Fig. 2. Benign Tumor (4) Image with multiple target mask from the Dataset.

III. Model Description

A. Data Preprocessing

The initial step for preprocessing the data includes separating the images with a single mask from those with multiple masks. In the dataset, each image has a corresponding mask labelled with "mask.png" at the end of its root path. Those with two masks will have a second labelled with "mask_1.png" at the end of its root path. While iterating through all of the single mask paths, the corresponding image path is appended onto a separate list. This results in two lists which holds the path to both image and mask in matching order. For example, the image at index 50 would correspond with the mask at index 50. In order to ensure each image has one corresponding mask, those with 2 masks must be merged together. This was done by iterating through the list of double mask paths and initializing a empty image with the same input dimensions. The image for both the first and second mask are then loaded through a defined function load_image which returns the image at a specified file path resized to a given size and normalized between 0 and 1. The pixel values are ensured to be within the range [0,1] using .clip before the numPy array is converted back to an image object and saved in the path of mask 1. The merged mask as seen previously for the benign tumor (4) is shown in *Fig.4 Benign Tumor (4) Image with merged mask.* All images from the dataset are also converted to a numPy array and resized to 256x256 pixels and normalized between 0 and 1 with decimal place 4.

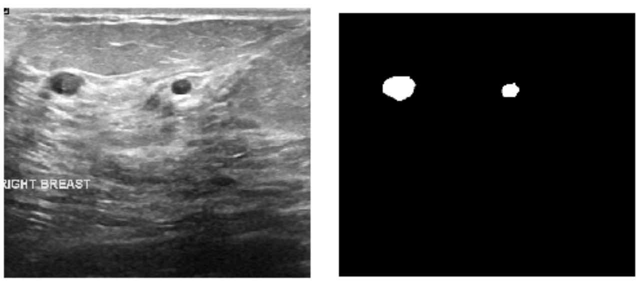

Fig. 3. Benign Tumor (4) Image with merged mask

B. U-Net Architecture

The U-Net model is implemented to perform image segmentation on the breast cancer dataset. It is a fully convolutional neural network that is famous for its applications in localization and feature capturing. Its ability to do pixel-level localization is what makes it a popular model for biomedical imaging.

The model has 4 encoders with the first having 32 filters and a dropout rate of 0.1, second with 64 filters and 0.1 drop out rate, third with 128 filters and 0.2 drop out rate and fourth with 256 filters with 0.2 drop out rate. Each convolutional kernel has a size of 3x3. The ReLU activation function is applied after the first and batch normalization layers. The spatial dimensions is not changed during the convolutions and batch normalization as the padding is kept same. Max poling layer with a pool size of 2x2 is used to down-sample the feature maps by reducing both the height and weight by a factor of 2. After encoder block 1 the spatial size will be (128,128), (64,64) after encoder block 2, (32,32) after encoder block 3 and (16,16) after encoder block 4. These dimensions correspond to the decoder blocks for up sampling.

A single encoder block with 512 filters is used at the bottleneck of this model, which is the bridge between the encoder and decoder. At this layer, the spatial dimension remains the same but has a higher feature channels to allow for the object extraction from the image. The output from the last encoder block is passed to the bottle neck.

The filters are then reversed for the up-sampling to refine the features. The model has 4 decoders with the first having a filter of 256 and drop out rate of 0.2, second decoder with 128 filters and drop out rate 0.2, third decoder with 64 filters and 0.1 drop out rate and fourth decoder with 32 filters and drop out rate 0.1. The spatial dimensions of the feature map are up-sampled using the up convolution. The concatenate layer combines the feature maps from the up sampled output with the skip connections from the encoder. Two convolutional layers are then used to further process the concatenated feature map with ReLU activation function. The final output layer has 1 filter which generates one output channel for each pixel to range from 0 and 1. A sigmoid activation function converts the output score from the final convolutional layer to an output between 0 and 1. It helps to determine the foreground and background as pixels with values above a threshold are classified as the object and those under will be classified as background This generates the final segmentation image.

IV. Results

The data frame which includes both the image and mask data were split with 10% being allotted to testing and 90% to training. The training data was further divided into validation which allotted for 20%. Several evaluation metrics were used including loss based on binary cross entropy (BEC), accuracy and intersection over union (IoU).

BEC is a loss function that provides a score between 0 and 1 to indicate the dissimilarity between predicted probabilities and actual binary labels in a classification problem. In the context of image segmentation, it is the measure on how well the predicted mask matches with a ground truth mask, A lower value indicates that the prediction is closer to the actual binary value. The loss is plotted below for each epoch for both training and validation as seen in *Fig 4.Loss Metric for Training and Validation data.* The data shows a decreasing loss for each epoch, which further confirms the ability of the model to product a mask prediction that is close to the ground truth mask.

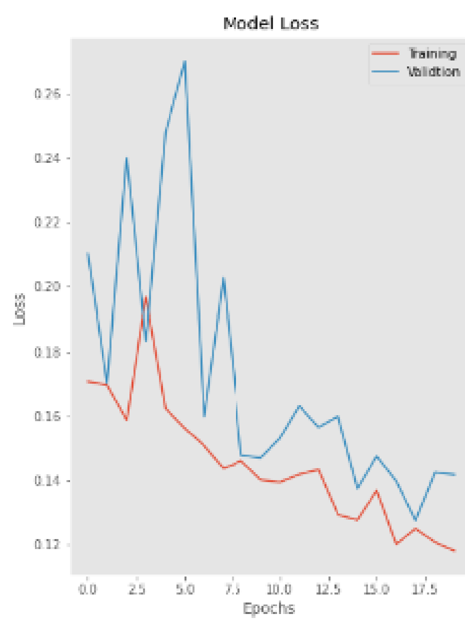

Fig. 4. Loss Metric for Training and Validation data.

The accuracy is based on how often the predicted masks match the actual mask. The results for the training and validation data is seen in *Fig 5. Accuracy Metric for Training* *and Validation data.*. Although the results for accuracy is fluctuating, we can see that it has a general increasing slope after each epoch. IoU is used to measure the accuracy of the object detection in image segmentation models. It is the overlap between the predicted and truth segmentation. A high IoU indicates a model's ability to accurately identify and localize objects in an image. The IoU for training and validation data can be seen in *Fig 6. IoU Metric for Training and Validation data.*

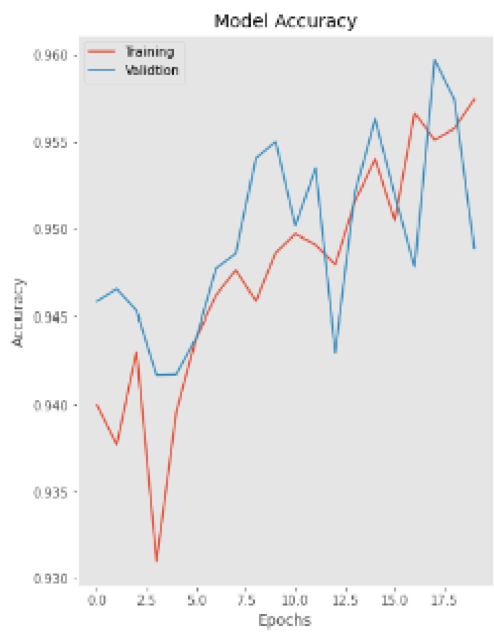

Fig. 5. Accuracy Metric for Training and Validation data

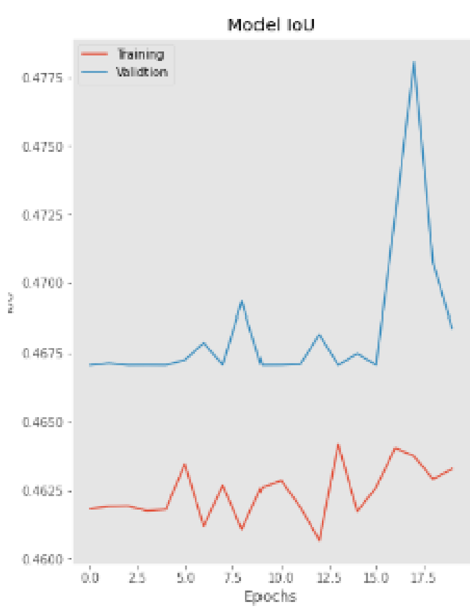

Fig. 6. IoU Metric for Training and Validation data.

The final model was evaluated using the test data and the performance metrics can be found in *Table 2.Average Metric Results for Test Data.* Although the results show promising results, it is also important to note that over-segmentation is present as seen in *Fig 7.Over-segmentation Prediction.*. The model is able to accurately segment the original mask area but also incorrectly predicted additional areas. In medical imaging, having false positive may be preferred compared to false negative as it can pose a bigger risk if detection's are missed. That being said, some factors worth looking into include balancing the dataset, adjusting the thresholds or experimenting with other regularization techniques. Additional next steps could include experimentation with variations of U-Net architecture including adding more layers or attention mechanisms to catch more features.

Metric	Value
Testing loss	0.2423
Testing accuracy	0.9163
Testing IoU	0.4543

TABLE I
AVERAGE METRIC RESULTS FOR TEST DATA

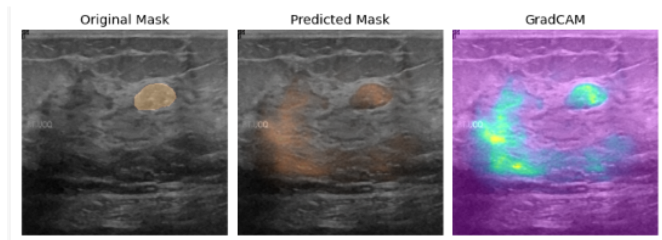

Fig. 7. Over-segmentation Prediction

V. CONCLUSION

In Conclusion, this report highlights the application of U-Net architecture for image segmentation of breast cancer ultrasounds. The encoder-decoder structure of U-Net has shown significant promise in accurately identifying the targeted tumor areas but should be refined further in order to improve the prediction results. The use of binary cross-entropy as the loss function proved effective in guiding the model towards accurate segmentation , but efforts should focus on reducing the loss and improving the intersection over Union(IoU) scores. The model's capacity to capture and reconstruct complex features from low resolution ultrasound images is evident but challenges such as high speckle noise and image quality remain. Future work may involve exploring hybrid models, incorporating attention mechanisms and looking into the over-segmentation cases. The application of CNNs to perform image segmentation has the potential to greatly increase early diagnosis and detection, thereby supporting healthcare professionals with their practice and improving patient care.

References

[1] L. Zhang, Y. Ren, C. Huang and F. Liu, ”A novel automatic tumor detection for breast cancer ultrasound Images,” 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Shanghai, China, 2011, pp. 401-404

[2] S. Cai, Y. Zhu, J. Zhang and T. Liu, ”A Study on the Combination of Image Preprocessing Method Based on Texture Feature and Segmentation Algorithm for Breast Ultrasound Images,” 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2022, pp. 760-764

[3] A. Ciobotaru, D. Gota, T. Covrig and L. Miclea, ”Reliable Breast Cancer Detection from Ultrasound Images using Image Segmentation,” 2024 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), Cluj-Napoca, Romania, 2024, pp. 1-5

[4] M. Amiri, R. Brooks and H. Rivaz, ”Fine-Tuning U-Net for Ultrasound Image Segmentation: Different Layers, Different Outcomes,” in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 67, no. 12, pp. 2510-2518

[5] S. Marques et al., ”Segmentation of gynaecological ultrasound images using different U-Net based approaches,” 2019 IEEE International Ultrasonics Symposium (IUS), Glasgow, UK, 2019, pp. 1485-1488

[6] C. Wang, J. Zhu, K. Snu, D. Li, Z. Wang and H. Yuan, ”A COVID-19 medical image Segmentation method based on U-NET,” 2021 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE), Shanghai, China, 2021, pp. 1-4

[7] S. R. Totad, Meeradevi, B. J. Sowmya, P. Boppudi, N. Yaswanth and I. Nivedith, ”Brain Tumour Detection and Segmentation using CNN Architecture and U-Net Architecture,” 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), Kalaburagi, India, 2023, pp. 1-5

[8] S. AbdElhakem, S. Basiony and M. Torki, ”ConvMixer-UNet: A Lightweight Network for Breast Lesion Segmentation in Ultrasound Images,” 2023 20th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA), Giza, Egypt, 2023, pp. 1-5

[9] T. -T. Tran, D. -T. Vu, T. -H. Nguyen and V. -T. Pham, ”A CNN-Transformer-based Approach for Medical Image Segmentation,” 2023 International Conference on System Science and Engineering (ICSSE), Ho Chi Minh, Vietnam, 2023, pp. 22-27

[10] V. -Q. Nguyen, Q. -H. Nguyen and T. -T. Tran, ”ConvmixFormer-Unet: A New Approach for Medical Image Segmentation based on Convmixer and Transformer,” 2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS), Hanoi, Vietnam, 2023, pp. 662-667

[11] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015.

[12] Canadian Cancer Society / Société canadienne du cancer, “Breast cancer statistics,” Canadian Cancer Society, https://cancer.ca/en/cancer-information/cancer-types/breast/statistics