Optimizing Ultrasound dB Thresholding for Breast Tumor Classification and Segmentation Tasks

Sarah Putney **Sabrina Qi** **Teresa Mao** **Adrienne Hawkes**

Abstract

Using raw radio-frequency ultrasonic echoes from malignant and benign breast lesions, we performed classification and segmentation tasks. The classification task of identifying benign and malignant tumors was performed using a custom CNN model. The segmentation task of identifying lesions within the breast tissue was performed using MobileNetV2, a pre-designed U-Net auto-encoder. An additional parameters was also trained as a representation of adjustable physical parameters: decibel cutoff for image compression. Combined with the physical layers, the classification model was able to achieve 93% training accuracy, while the segmentation model was able to achieve 71% training accuracy. These results indicate promising future directions in using radio-frequency data for identifying and classifying lesions through use of machine learning algorithms.

1 Introduction

Breast cancer was the 5th leading cause of mortality in 2018. Approximately 2.1 million new diagnoses were made in 2018, and about 1 in 4 cancer diagnoses in women are breast cancer, making it the most commonly diagnosed cancer in women. Additionally, breast cancer is the leading cause of cancer death in women at 15% [1].

These statistics make early diagnosis extremely important for prevention of breast cancer development. The current gold standard for diagnosing malignant breast lesions is biopsy, which is a physically invasive and time-consuming process. To investigate a quicker and non-invasive method of diagnosis, we used radio-frequency data obtained from ultrasound (US) data in two machine learning tasks. The first task was benign-malignant classification to identify whether a lesion is cancerous. This task has large implications in bypassing the invasive and time-consuming biopsy procedure. The second task of segmentation was performed in hopes of easily identifying tumor lesions from US data. The OASBUD data [2] set included masks identifying the region of interest, the lesion, in each image, motivating the interest in attempting segmentation on these images.

In addition to the optimization of the CNN models used in these tasks, we attempt to optimize a physical parameter representative of the imaging system. Processing the raw US data into US images is dependent on a decibel threshold that is set to manage applied compression [3]. The resultant compression is known to affect the B-mode images, which can affect classification and segmentation. Common thresholds used in the medical setting include 40, 50, and 60 dB as showing in Figure 1 [3].

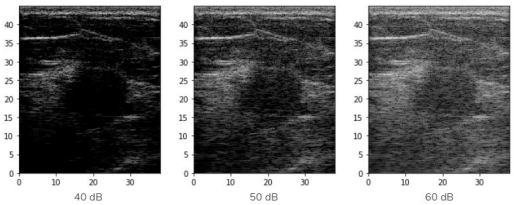


Figure 1: B-mode Ultrasound for Varying dB Thresholds

2 Related Works

There have been numerous studies done on the the OASBUD raw US data set. In a paper studying the affects of ultrasound image reconstruction methods on classification of breast lesions, it was discovered that image reconstruction using varying compression levels resulted in varying levels of neural network accuracy. In this study, three different threshold levels were chosen for testing: 40, 50, and 60 dB [3]. Using the pre-trained deep CNNs InceptionV3 and VGG19, the study made several interesting observations about impact of image reconstruction. It was found that when the test and train data sets are reconstructed with dissimilar thresholds, the accuracy in classification decreases greatly [3]. This study's investigation of the effects of variable image reconstruction methods on classification performance inspired the inclusion of a physical layer in our classification model that optimizes for compression level.

The OASBUD data set has also been used to train deep convolutional neural networks on just raw radio-frequency data (without US image reconstruction) [4]. This proved to have good classification performance, thereby indicating that not only can preprocessed US images be classified through CNNs, but raw RF data can also be directly used to train CNN's successfully [4].

3 Methods

3.1 Pre-Processing Data

For all tasks performed in the Deep Neural Networks designed by our team, preprocessing was performed on the data to convert it from raw RF signals to B-mode ultrasound images. The reconstruction performed to compress the raw US signals into images was performed following the methods outlined in a the OASBUD data publication [3]. To compress the data, first the Hilbert Model was used to find the data envelope. The following equation was applied to perform a logarithmic compression of the data:

$$A_{log} = 20 * log_{10}(A/A_{max})$$

After pre-processing, we applied a threshold value simulating how many decibels can be applied via the US transducer. This is described in greater depth in Section 3.4.

3.2 Classification

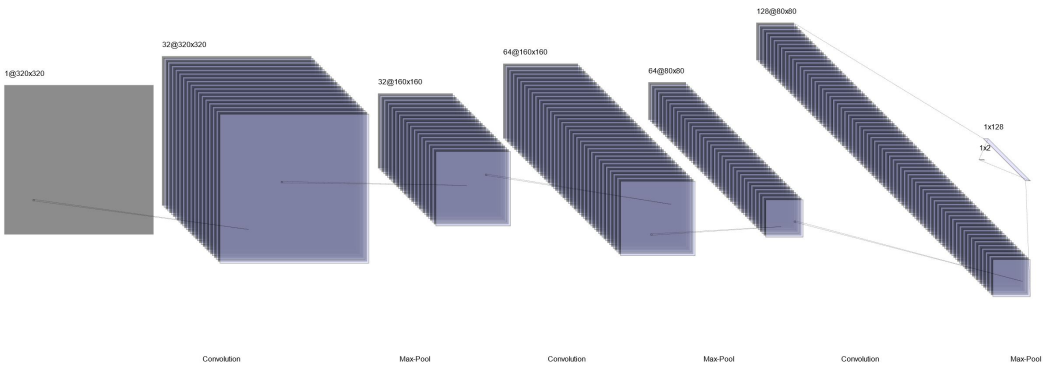


Figure 2: CNN Architecture for Classification

In our Classification task, we began by augmenting our data set of 200 B-mode images in order produce more data for training and testing. To do this, we first identified the corners of the lesion in each of the B-mode images, then added a buffer of 20 pixels of data on each side of the lesion. We randomly selected the locations of the 20 pixel border to shift the location of the tumor within the resultant images. Lastly we flipped the images horizontally and vertically. Overall we performed this sequence of augmentation 10 times for each image giving 6,000 images in our final data set for classification.

The model we used in this task was formatted after the model used in the machine learning article [4]. In this model we included three repeating layers including a convolution, max pooling, and batch normalization. In each convolutional layer, we increased the number of filters by 2, maintained the stride, and used a kernel size of 3x3. We also used a leaky reLu activation function to correct for the

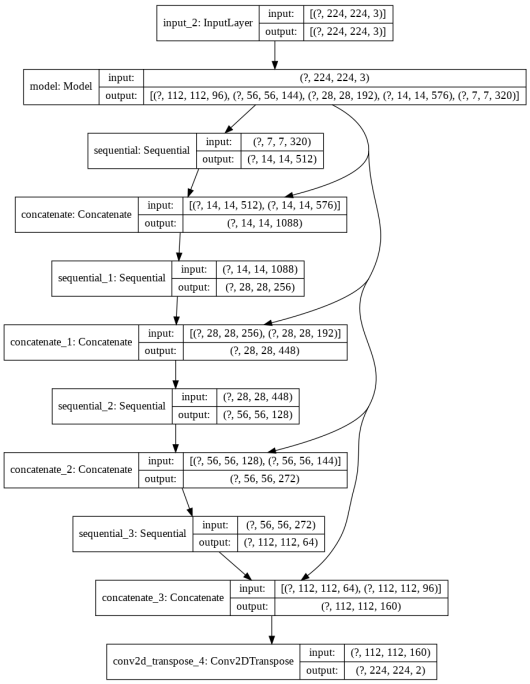


Figure 3: CNN Architecture for Segmentation

error found in using a reLu. Our fully connected layer consisted of two Dense layers and a dropout of 0.1. The complete architecture can be seen in Figure 2.

3.3 Segmentation

In our segmentation task, we also began by augmenting our data. To augment the data, we began by randomly cropping and resizing all the data to [224,224] images. To ensure there was sufficient amounts of the lesion in each image (such that there would be notable contrast between the breast tissue and the tumor), after each cropping iteration we ran a check that mandated that each cropped image obtained more than 30 percent and less than 70 percent lesion data. To ensure satisfactory outputs, we discarded all invalid crops, and lowered the minimum percent if too many augmented images were discarded. Lastly, we flipped each cropped image in three different ways. We performed random cropping on each image 8 times, then flipped each image horizontally and physically to give a final data set of 4,800 different images.

To complete the segmentation, a U-Net autoencoder, the MobileNetV2, [9] was utilized. The decoding half of the U-Net utilized the upsample function found in the Pix2Pix generator [8]. The model also included a 2D convolutional layer as the last layer of the model. The full architecture can be seen in Figure 3. For training and testing, we passed in our batch data with a batch size of 10. We then fit our model over 10 epochs.

3.4 Trainable Physical Parameters

We considered several potential Ultrasound imaging parameters when determining which to train in our physical layer. And ultimately chose to optimize the level of dB thresholding performed prior to running the images through the CNN. Applying a decibel threshold is a common tool used by ultrasound technicians to enhance the contrast of the B-Mode image generated by the acquired rf data. We created a physical layer that would identify all values below some variable, trainable, threshold in the ultrasound image, and replace those values with the threshold itself. This allowed for the optimization of the image contrast via the training of the dB Threshold tensorflow variable.

4 Results

4.1 Full Numeric Results

Table 1 below shows all final testing accuracy and testing loss produced for the Classification, Segmentation, and Classification and Segmentation with a trained physical parameter designed to predict the optimized dB threshold for ultrasound image contrast. For both the raw Classification and raw Segmentation tasks, the model was fit using a set threshold of 50dB. The dB thresholds used in the physical layer training for the last two tasks are found in Table 2.

CNN Task Name	Train Acc	Test Acc	Train Loss	Test Loss
Classification	93.10	60.75	0.5837	7.299
Segmentation	80.17	63.05	0.4290	0.6412
Classification including Physical Layer	71.42	61.10	3.73	3.78
Segmentation including Physical Layer	71.42	61.08	3.726	3.782

Table 1: Accuracy and Loss for all CNN

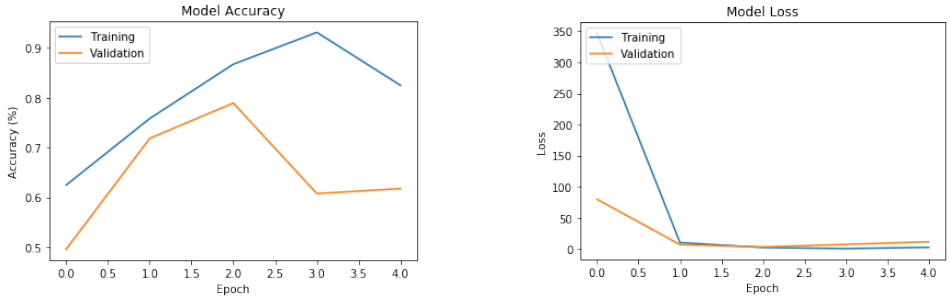
The following decibel thresholds were optimized as trainable physical parameters. These are the compression values that a doctor should implement when imaging to acquire the best results for classification and segmentation.

Classification	Segmentation
53.970 dB	50.0 dB

Table 2: Optimized dB Threshold as Physical Layer in CNN

4.2 Classification

The two following plots were produced using a classification CNN with fixed physical parameters. Both training and test accuracy increase over epoch until there is a drop at around the same point in epoch 3. This indicates potential over fitting of the model after epoch 3. Loss decreases dramatically for both training and testing.



(a) Accuracy (b) Loss

Figure 4: Varying dB Threshold Results over Epoch for Segmentation

4.3 Segmentation

The two figures below show the results of the raw segmentation task and different points in the training process. Figure 5 shows the results of segmentation prediction earlier in training, during epoch 2. It has a low accuracy and high loss, producing a lower quality segmented image. Figure 6 depicts a higher quality segmented lesion after 10 epochs of training, with higher training accuracy and loss.

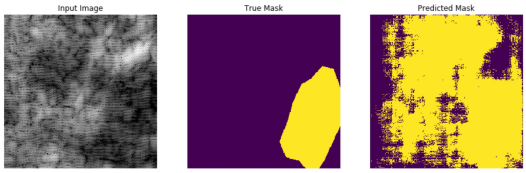


Figure 5: B-Mode US, Mask, Segmentation at Epoch 2

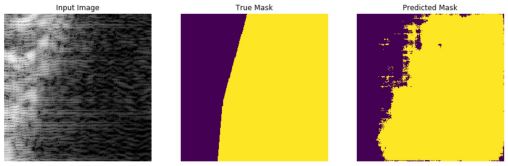


Figure 6: B-Mode US, Mask, Segmentation at Epoch 10

4.4 Trainable Physical Parameters

The following four figures were generated while training the physical parameter of decibel threshold along with the classification and segmentation networks. Figure 7 shows the train and test accuracy against the dB threshold level. The model optimized the dB threshold to be approximately 54 dB at the highest training and validation accuracy.

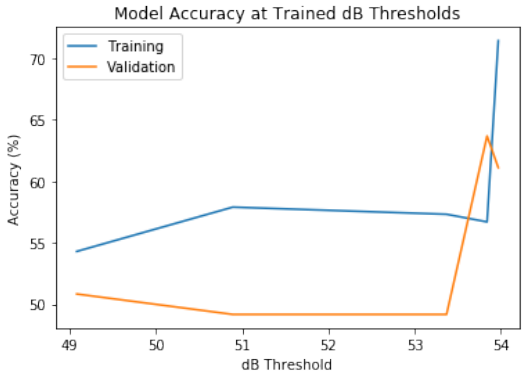
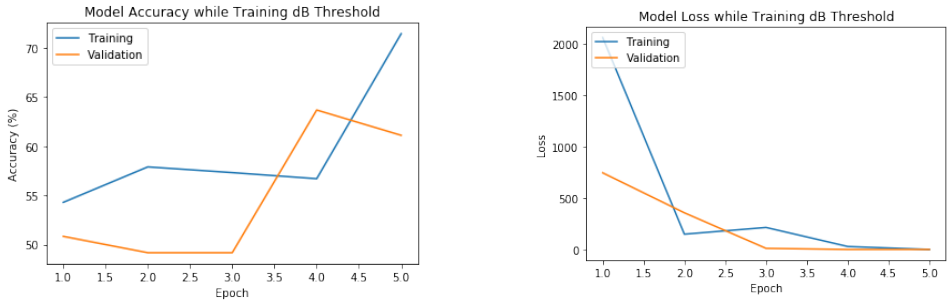


Figure 7: Accuracy over Varying dB Threshold for Segmentation

Figure 8 shows the train and test accuracy and loss for classification when training the physical parameter of dB threshold. Both training and test accuracy increase over epoch until there is a drop at approximately epoch 4, indicating potential over fitting after epoch 4. Loss decreases dramatically for both training and testing after the second epoch.



(a) Accuracy

(b) Loss

Figure 8: Varying dB Threshold Results over Epochs for Classification

Figure 9 shows the test and train accuracy against the dB threshold level for segmentation. The highest accuracy occurs around 50dB for the validation data set, similar to that of the classification task with physical parameter. Testing remains at a high accuracy throughout the process of training the network.

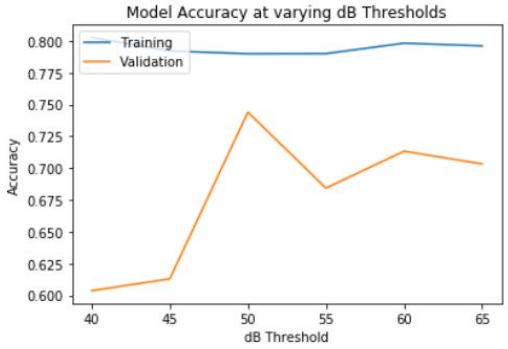
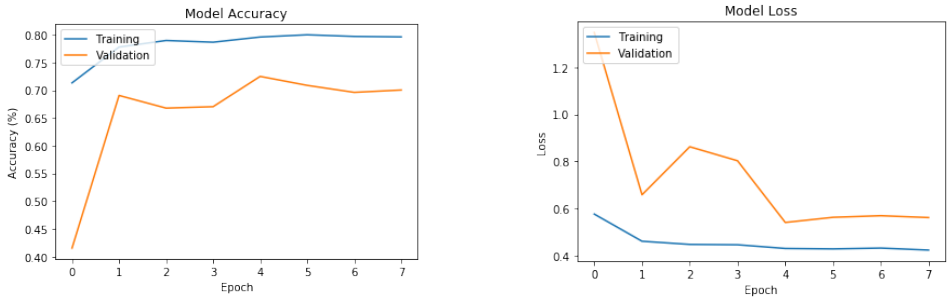


Figure 9: Accuracy over Varying dB Threshold for Segmentation

Figure 10 shows the train and test accuracy and loss for segmentation when training with the physical layer of the dB threshold over 10 epochs. The loss decreases steadily excluding the brief increase in the middle of training the model. Accuracy increases rather rapidly with and remains at approximately the same value.



(a) Accuracy

(b) Loss

Figure 10: Varying dB Threshold Results over Epoch for Segmentation

5 Discussion

From the results shown above, we can make several observations about our the classification and segmentation tasks in light of the physical layer. Looking at Table 1, we can see that our test accuracy consistently falls significantly below training accuracy, indicating that the model we train may not generalize well. This may be due to larger batch sizes that were used to increase accuracy. Additionally, the classification model had the highest training accuracy, but its test accuracy was relatively similar to that of other tasks. This implies that inability to generalize affects the classification model more than it affects other models. Segmentation has relatively high training accuracy at 80.17%, has the smallest loss, and has the highest test accuracy. This indicates that the segmentation-augmented data works well to train the model, but may not generalize well due to a small batch size.

For the classification model, the test accuracy increased from the first to the third or fourth epoch, then decreased after that as the loss also became smaller and the train accuracy continued to increase, as seen in Figure 4. This indicated overfitting of the model, which in part may be due to the lack of data that we had. The OASBUD data included 100 unique lesions, each of which was imaged from two directions to create 200 unique images. These images were then augmented through random cropping and flipping to a total of 6000 images - still of the same 100 lesions - that were passed to

the model - 80% of which were training data and the rest test data. Because of the RAM constraints of our machine, we were not able to increase the size of the data set beyond these values.

Another possible explanation for the low accuracy is the augmentation method. The rf data starts as a rectangular array of approximately (3x, x) dimension, that when converted to a B mode ultrasound image, looks like a square. So when the rectangular rf matrix is cropped to a square, the resulting B mode image looks like a squashed rectangle, and the tumor is hard to distinguish by eye. The resizing also may have had an impact, as in most cases it lowered the resolution of the cropped images, which may make it more difficult for the model to pick out finer features that distinguish between a benign and malignant lesion. It also makes it difficult to compare the sizes of the lesions against each other. However, the resizing was necessary to first ensure that the images passed to the model are all the same size, and then keep the images small enough so as to not use up all the RAM.

The segmentation model shows some promising results, as pictured in Figure 6. Both Figure 5 and 6 show a B-Mode US image, the corresponding true mask that indicates where the tumor is, and the predicted mask, which is the result of our U-Net autoencoder. It is evident that the segmentation task is optimized over the course of the model training because in Figure 6 fewer of the pixels that are not tumor are highlighted yellow and most of the area that is tumor is highlighted yellow.

The results of training the physical parameters provide interesting insight on the affect of thresholded compression on ability of our algorithm to train. Figures 7 and 8 show that the classification system approaches an optimized dB threshold as the number of epochs nears 5. The rise in accuracy coinciding with reduction of loss across epochs indicates that changes in the dB threshold variable affect the calculated gradient that is used in gradient descent towards an optimized model. Figures 9 and 10 show the impact of including the dB threshold as a trainable variable in the segmentation task. The plots show that a different optimal dB threshold was found at 50 dB for this task.

References

[1] "Latest Global Cancer Data: Cancer Burden Rises to 18.1 Million New Cases and 9.6 Million Cancer Deaths in 2018." IARC. https://www.iarc.fr/featured-news/latest-global-cancer-data-cancer-burden-rises-to-18-1-million-new-cases-and-9-6-million-cancer-deaths-in-2018/.

[2] Piotrzkowska-Wróblewska, Hanna, Katarzyna Dobruch-Sobczak, Michał Byra, and Andrzej Nowicki. "Open Access Database of Raw Ultrasonic Signals Acquired from Malignant and Benign Breast Lesions." Medical Physics 44, no. 11 (2017): 6105–9. https://doi.org/10.1002/mp.12538.

[3] Byra, Michal, Tomasz Sznajder, Danijel Korzinek, Hanna Piotrzkowska-Wroblewska, Katarzyna Dobruch-Sobczak, Andrzej Nowicki, and Krzysztof Marasek. "Impact of Ultrasound Image Reconstruction Method on Breast Lesion Classification with Deep Learning." Pattern Recognition and Image Analysis Lecture Notes in Computer Science, 2019, 41–52. https://doi.org/10.1007/978-3-030-31332-6_4.

[4] Jarosik, Piotr, Ziemowit Klimonda, Marcin Lewandowski, and Michal Byra. "Feasbility of Using Radio-Frequency Ultrasound Signals for Breast Lesion Classification with Deep Learning." *Proceedings of Machine Learning Research* (Under Review).

[5] L. Arbach, A. Stolpen and J. M. Reinhardt, "Classification of breast MRI lesions using a back-propagation neural network (BNN)," 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821), Arlington, VA, USA, 2004, pp. 253-256 Vol. 1. doi: 10.1109/ISBI.2004.1398522

[6] Hairong Qi and J. F. Head, "Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms," 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Istanbul, Turkey, 2001, pp. 2866-2869 vol.3. doi: 10.1109/IEMBS.2001.1017386

[7] "TensorFlow 2 Quickstart for Experts : TensorFlow Core." TensorFlow. Accessed December 11, 2019. https://www.tensorflow.org/tutorials/quickstart/advanced.

[8] "Pix2Pix:TensorFlowCore."TensorFlow.AccessedDecember11,2019. https://www.tensorflow.org/tutorials/generative/pix2pix.

[9] "Image Segmentation : TensorFlow Core." TensorFlow. Accessed December 13, 2019. https://www.tensorflow.org/tutorials/images/segmentation.