Cancer Image Classification Explanations through Transfer Learning on ResNet and U-Net Architecture

Nate Hollenberg

May 12, 2019

1 Introduction

In the past few years, Convolutional Neural Networks (CNNs) have become the leading method of image classification in the field of artificial intelligence. This is because CNNs have relatively few parameters compared to standard feed forward neural networks of similar sizes. Further, CNNs have a large learning capacity necessary to perform well at a task with the complexity of image classification and have a performance level comparable to that of a human. Two developments that have mitigated obstacles towards using CNNs in recent years are the creation of labeled image datasets large enough to train CNNs and increased access to GPUs which provide the processing power necessary to facilitate the training of CNNs [8].

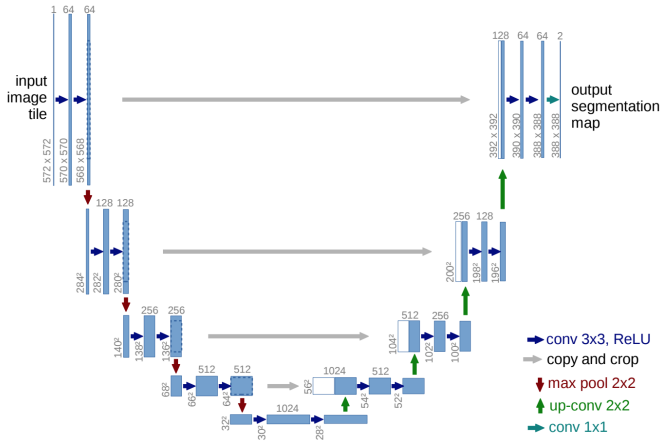

Figure 1: U-Net Architecture

In 2015, a team lead by Olaf Ronneberger at the University of Freiburg in Germany developed a CNN specifically designed for image segmentation. The team entered their neural network in the ISBI Cell Tracking Challenge of 2015. The neural network beat the second place network by

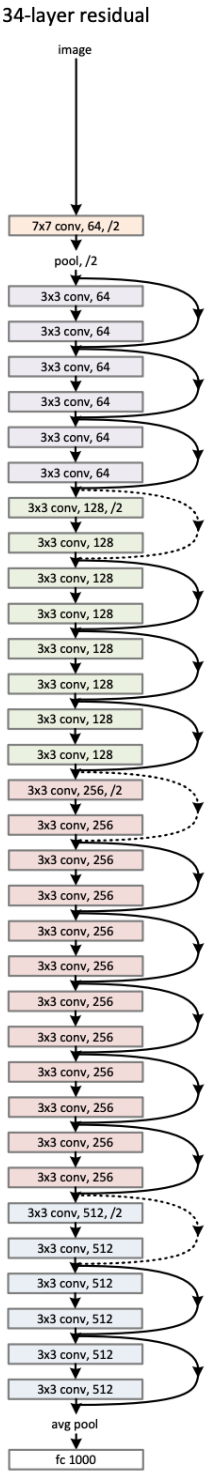

Figure 2: Res-Net Example Architecture

over 9% in terms of IOU (Intersection Over Union) a common evaluation metric for image segmentation. The team named their network the U-Net, because of the U-Shaped architecture of the model. The architecture starts with a contracting path to capture context before a symmetric expanding path that allows the network to perform localization. This unique architecture allows the U-Net to perform well on a small training sample with a relatively fast training time [3].

While the U-Net was winning image segmentation contests in 2015, a Microsoft Research team lead by Kaiming He was using a different type of neural network to win the ILSVRC 2015 classification task. The Residual Neural Network (ResNet) is different from the typical neural network as it not only feeds forward into the next layer, but it also jumps 2-3 layers and feeds directly into that as well. This jumping nature is displayed in the figure to the left (Figure 2), which is the architecture of the ResNet-34, a ResNet with 34 layers. This unique jumping nature of the ResNet allows it to better deal with degradation problems making the network easier to optimize. The ResNet is also unique because it has a lot of depth, many layers between the input and output, yet is relatively much easier to train than other deep neural networks because of its architecture. These factors make it able to perform exceptionally well on image recognition tasks [5].

There are many common ResNets in addition to the ResNet 34. The ResNet 50 is another popular network and is similar to the 34-layer network except each 2-layer block is replaced with a 3-layer bottleneck block [5]. This model can be accessed on the public open source neural network library Keras. In addition to the ResNet-50, Keras also carries pre-trained weights for the ResNet-101 and ResNet-152 [4].

In other academic papers, the ResNet has been applied to abnormal mammograms to classify breast cancer in patients as malignant or benign [6]. There have also been papers published applying the U-Net to mammograms to identify the location of abnormalities within the breast [9].

While previous research has been successful in image segmentation with U-Nets, our research hopes to introduce an new classification method into the U-Net. This classification strategy is outlined in the method sections. Further, we add new levels of interpretation to the U-

Net by visualizing predictions after every epoch to see how the model is learning with respect to its cost function. Overall, the aim of this research is to mimic a pipeline in classification cancer images. This pipeline will first understand the area of the abnormality through the U-Net, and then run the predicted ROI through the ResNet to fully comprehend the classification. However, with our new classification method applied to the U-Net, it is possible to add levels of interpretability to our classification.

2 Data Aggregation

To access the data, we decided to use a Curated Breast Imaging Subset (CBIS) of the USF DDSM data found on the Cancer Imaging Archive [1]. Here, it was necessary to download software from the Mac App Store called NBIA Data Retriver to fully access the data. We downloaded .tcia files from the Cancer Imaging Archive for just the Mass abnormalities as our research will just focus on this type of abnormality. We downloaded 776 images from the pre-split training set and 238 images from the pre-split testing set. We downloaded both the full mammography images as well as the Region of Interest (ROI) and cropped images that corresponded to each full mammography image. The images were originally downloaded in .dicom form, but we transformed them to .tif files, which took up much less disk space and were compatable with Keras Image Generators. We organized the data in different folders corresponding to full images, cropped, and ROI images. Further, we created a .csv file containing the cropped image path and its corresponding classification of malignant/benign to assist in using a keras image generator in our model building. Overall, for our U-Net models, we used the full mammography images as our input with the ROI images as our response. For the Res-Net we used the cropped images as our input, with the response as the class in the data frame. From here, our data was prepared for for the modeling analysis.

3 Preliminary Data Analysis

The data themselves were not very complicated. In Figure 3 we can see the class breakdown in both the training and test sets:

Cancer Status	Train	Test
Malignant	374	91
Benign	402	147

Figure 3: Class Breakdown in Train and Test Set

Further, it is essential to see each type of image before the analysis to see what the models are dealing with. Below are the full image, ROI image, and cropped image for a single patient in the training set.

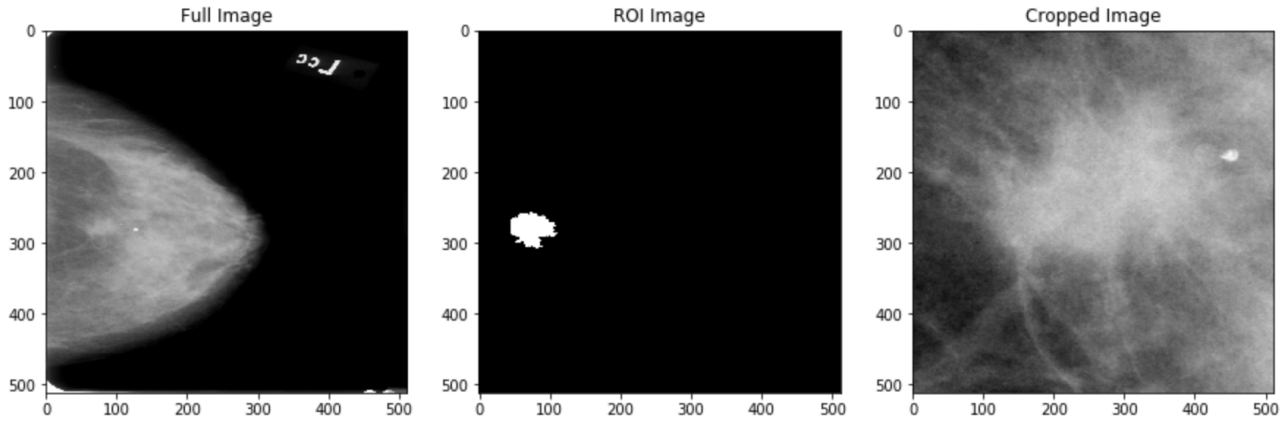

Figure 4: Example Images

After understanding the class breakdown and visualizing the types of images in our dataset, we can move onto to the methods of our analysis.

4 Methods

4.1 U-Net

To begin the model analysis, we downloaded a pre-existing U-Net model from github, titled the ZF-UNET-224 [13]. The U-Net model was influenced by the architecture in the 2015 University of Friedburg paper [3]. The network consists of a contracting path and an extracting path as visualized in Figure 1 with the two mirroring each other. The contracting path contains 6 double 3x3 convolutions with a ReLu activation in between. Batch normalization is included with a 2x2 max pooling operation for down-sampling following the double convolution. The extracting path is the opposite utilizing up-sampling and up-convolutions. A more detailed view of the network can be found in the repository found at this link ZF-UNET-224.

The model contained 31,466,753 parameters, but after accessing the initial pre-trained weights from the model, we needed to train 31,454,721 of the parameters. We utilized the strategy of **transfer learning** to apply our new training data to the initial pre-trained weights of the ZF-UNET-224. We set up a train generator and a validation generator with a 80/20 train/validation split. We then fit the model on the generator using various data augmentation techniques such as a vertical and horizontal flip, and different width and height shifts. The model used the Adam optimizer and the dice loss function [2]. We ran the model for 50 epochs and interpreted each epoch with visuals depicting the full image, ROI image, and predicted ROI image on the 4 images from the validation batch. This is a novel idea that assisted in our interpretation and training of the model as we were able to see the improvement of our model at each epoch, as well as how the visuals improved in relation to our cost function. Further, it is important to see how the model is training at each epoch to see how it arrives at its final classifications as way of explaining the model. Thus, these per epoch visualizations greatly assisted in the interpretabilitiy and tuning of our model.

The U-Net model was first trained on the full selection of ROI images that corresponded to each full mamography image. The model was trained and the weights were saved. The second U-Net model that was trained used the same weights from the first U-Net, but with a change in the data. Instead of using the full selection of ROI images regardless of classification, we chose to use the full ROI images for malignant abnormalities, and a "black-box" for a benign abnormalities – images with all black pixels. Thus the full images that were malignant were trained with their corresponding ROI image, while the full images that were benign were trained with an all black image. With this change, the second iteration of our U-Net attempted to simultaneously classify abnormalities while segmenting the ROI from a full image. The simplest way to perform classification on the output images would be to check that all the pixels in the image had value 0. We chose to instead take the sum of all pixels and compare this sum to some low threshold value. Images with pixel sums over the threshold would be classified as positive and those below negative. Prior work has attempted to classify masses by shape descriptions, though to our knowledge, U-Net's have not been used to perform classification in this way [11]. This approach offers a much more interpretable form of ANN classification of breast abnormalities than previous examples. In fact, the human eye could perform the same basic classification of determining the presence of white pixels. Further, it offers the opportunity to add context to uncertain decisions. In typical ANN classifications, uncertain decisions are returned with no additional information. In our approach, an uncertain decision would come with a corresponding mask showing a small region of white in a specific area of the mask. This would allow a human interpreter to check the corresponding area on the original image and check the result manually.

To see the effectiveness of each U-Net iteration, we predicted on a test generator from the testing set of full mammography images. We subsequently measured the performance of the U-Net using various accuracy metrics. The overall goal of the U-Net is to segment full mammography images to understand the location of the abnormality. The "black-box" extension helps to see whether the U-Net is able to classify correctly while understanding the location of mass abnormalities. However, if unsuccessful, further implementation is necessary to classify the ROI with the ResNet.

4.2 ResNet

Since our CBDIS included cropped images that zoomed in on the region of interest, we decided to build the ResNet on this data. Future research could potentially include an ensemble model to feed the ROI predicted by the U-Net into the ResNet model to deliver a classification. Yet, in the given time frame for the task, we chose to train the ResNet on the included cropped images from the Cancer Image Archive.

After much research, it became evident that training a ResNet from scratch was too time consuming given the available resources. Thus, it became necessary to implement **transfer learning** to build a ResNet suitable for classification given our resources and the given timeframe. Thus, we implemented a ResNet given the Keras pre-trained ResNet50 [7][13]. After uploading the pre-trained weights from Keras, we added one Dense layer to classify the images as malignant or benign. The Dense layer used a softmax activation, and an output dimension of 2 for the two classes of malignant and benign. The model only contained 4,098 trainable parameters, which could be a potential drawback in the robustness and accuracy of our model. The model was com-

piled using a binary cross-entropy loss function using a learning rate of 0.01, decay of 1e-6 and a momentum of 0.9. The model was fit again on a train and validation generator with an 80/20 train/validation split. The input were the cropped images of size (224,224), and the output was the classification. We fit the model on these generators and trained for 75 epochs.

To analyze the results of the model, we created a test generator from the test set of cropped images. From this test generator we predicted class, either malignant or benign. We analyzed our results using different accuracy based metrics such as confusion matrix and f1 score.

5 Results

5.1 U-Net

5.1.1 U-Net for Segmentation

First, we will examine the results of our U-Net trained on the standard image segmentation masks. The final model achieved a dice coefficient of 0.54. However, as we learned during training, the dice coefficient does not always capture the full notion of goodness of a segmentation mask. To this end, we will instead consider the performance on the test set qualitatively, examining actual predictions on test images.

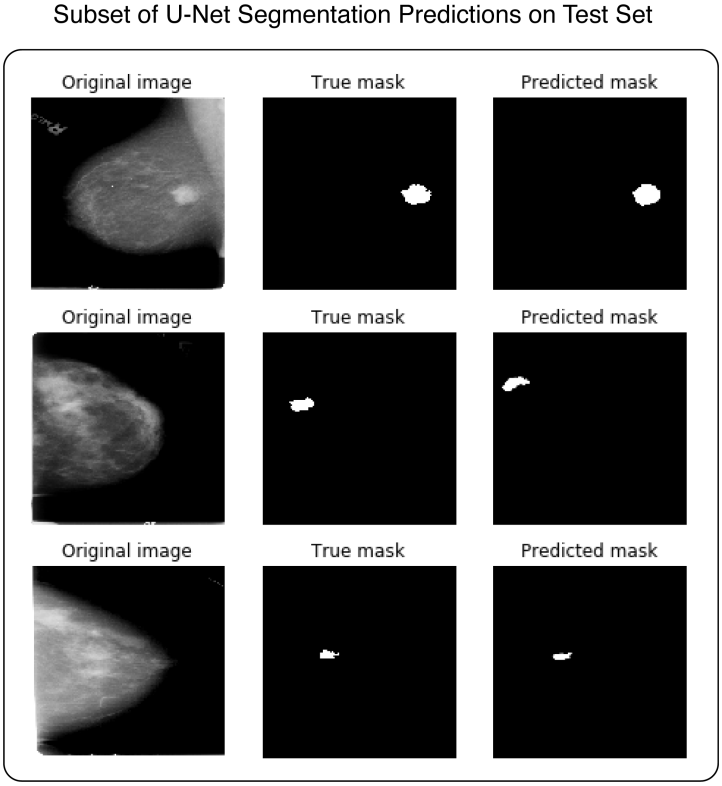

Figure 5: Example images from predicted image segmentation on test set.

We can see in figure 5 that our trained U-Net model performs well in predicting the ROI mask

given a mammography image. With further training and more available data, we expect this model to deliver even better results.

5.1.2 U-Net for Simultaneous Classification and Segmentation ("Black-Box" method)

Next, we will consider the performance of the U-Net trained on the "black-box" samples. In this case, we can consider both the effectiveness in classification as well as the performance of the masks.

The model achieved a testing accuracy of .58 on the test set using a threshold value of .0001. With higher thresholds, the model achieved higher scores, however, we wanted to minimize false negatives, as seen below.

	0	**1**
0	101	66
1	49	59

Figure 6: Confusion matrix of "black-box" U-Net with threshold of .0001

In Figure 6, we see the confusion matrix with the actual values on the y-axis and the predicted value on the x-axis. We can see that false negatives in the lower left corner are minimized. This is the case because we do not want to classify someone with cancer as benign, for the consequences of that misclassifcation are immense and severe. Next, we will examine the output images qualitatively.

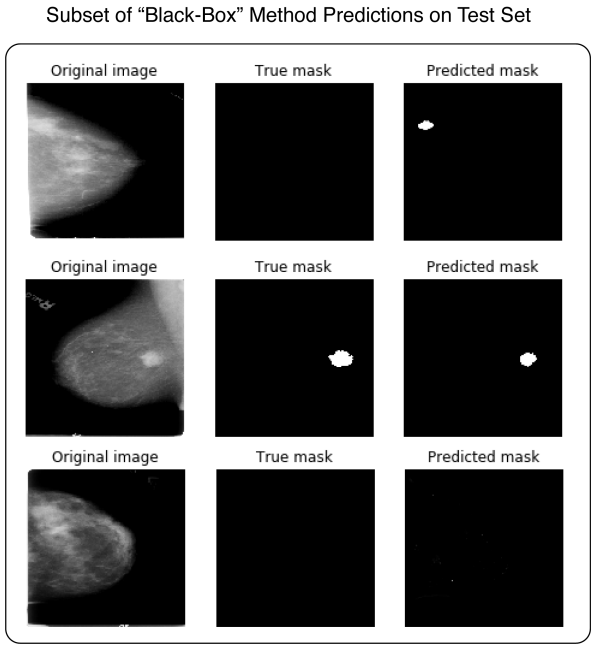

Figure 7: Example images from ”Black-Box” method on test set.

As we can see in Figure 7, the test images are similar in mask quality to the original U-Net. However, in addition there are some masks that have little or no white regions, which correspond to images classified as negative. While the accuracy score of our model was modest, as we mentioned above, this model is meant to be used as a tool for human interpreters, not just as a single number. For instance, if we consider images that fall between thresholds (say between .0001 and 100), we can find images that require furhter examination. For instance, in Figure 8 we can see an example where the model gave an intermediate score value.

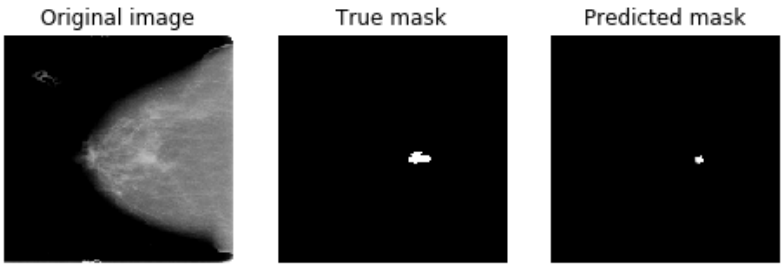

Figure 8: Example of uncertain classification from ”black-box” U-Net.

However, rather than just returning an unhelpful uncertain value, the model returns a partial

mask pointing to a possible area of concern. This would allow an interpreting doctor to examine the region manually and make the final classification.

5.2 ResNet

After analyzing the U-Net and its performance through both iterations, we wondered whether the ResNet achieved similar classification accuracy to the "black-box" U-Net. After predicting on the test generator with 238 images, we can get a sense of the accuracy results. After running the model, we can see that the model has difficultly with accuracy on the test set. The model achieves a 51% accuracy on the test set, with a F1 Score of 49%. It is evident that the implemented ResNet performs worse in classification tasks than the "black-box" U-Net. While the testing accuracy is poor, the model can still be somewhat informative to us. In our research, we are very concerned with not predicting false negatives. False negatives would be classifying individuals that have a malignant abnormality as benign, which would be very dangerous and inexplicable to the patient. To understand the breakdown of classifications, we can use a confusion matrix to see the distribution of true positives, false positives, true negatives and false negatives.

	0	1
0	67	80
1	37	54

Figure 9: Confusion Matrix

We can see in Figure 9 that we do not predict very many false negatives, which is a good sign. With this model, relatively few people will be walking around thinking that they are fine when they actually have cancer. However, in further research it would be advantageous to tune this model to achieve greater results in all aspects including accuracy and preventing false negatives. The ResNet was built through a transfer learning approach that took advantage of the weights of previously built image classification models. We then tweaked a few layers of the model to get it to train on our data and learn to classify between benign and malignant masses. To build an even better model, more parameter tuning or research into adjusting the weights of pre-trained models could be useful. Further, a more elaborate transfer learning approach might be advantageous. However, we limited our work on the ResNet in the given time frame in order to focus on the U-Net, which saw much better predictions that are more interpretable as they give not only whether or not there is cancer, but also where it is located in the image passed in. Further, the "black box" U-Net was able to achieve a higher accuracy score than the ResNet.

6 Conclusions

In this paper we explore existing approaches in transfer learning and propose a new approach to applying convolutional Neural Networks to the problem of breast mass segmentation and classification. Through transfer learning using the ResNet50 architecture, we were able to achieve moderate accuracy scores in mass classification. However, the lack of information provided along

with these results highlights the need for a more human-understandable output. To this end, the "black-box" method proposed in this paper offers a combination of image segmentation and classification in a single model. By tackling these problems simultaneously, we were able to create a more interpretable classification model. The main benefit of this approach is the ability for doctors to put the power in the hands of doctors and help support their decisions, rather than aim to create an oracle-like solution.

Future work could improve the classification accuracy of the model by training on more examples and using more resources. Further, it may be possible to apply a similar technique to number of different settings, such as other medical imaging classification problems or computer vision.

7 References

[1] University of South Florida. Digital Database for Screening Mammography. `http://www.eng.usf.edu/cvprg/Mammography/Database.html`

[2] Carole H. Sudre, Wenqi Li, Tom, Vercauteren, Sebastien Ourselin, M. Jorge Cardoso (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. arXiv:1707.03237 [cs.CV].

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV].

[4] Keras Documentation. Applications. https://keras.io/applications/

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2015). Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs.CV].

[6] Richa Agarwal, Oliver Diaz, Xavier Llad, Robert Mart (2018). Mass detection in mammograms using pre-trained deep learning models. Proceedings Volume 10718, 14th International Workshop on Breast Imaging. https://doi.org/10.1117/12.2317681.

[7] Sabyasachi Sahoo. Residual blocksBuilding blocks of ResNet. Medium. https://towardsdatascience.com/residual-blocks-building-blocks-of-resnet-fd90ca15d6ec.

[8] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM. Volume 60 Issue 6, June 2017. Pages 84-90.

[9] Timothy de Moor, Alejandro Rodriguez-Ruiz, Albert Gubern Mrida, Ritse Mann, Jonas Teuwen (2018). Automated soft tissue lesion detection and segmentation in digital mammography using a u-net deep learning network. arXiv:1802.06865 [cs.CV].

[10] Richa Agarwal, Oliver Diaz, Xavier Llad, Robert Mart, "Mass detection in mammograms using pre-trained deep learning models," Proc. SPIE 10718, 14th International Workshop on Breast Imaging (IWBI 2018), 107181F (6 July 2018)

[11] Vivek Kumar Singh, Hatem A. Rashwan, Santiago Romani, Farhan Akram, Nidhi Pandey, Md. Mostafa Kamal Sarker, Adel Saleh, Meritexell Arenas, Miguel Arquez, Domenec Puig, Jordina Torrents-Barrena (2018). Breast Tumor Segmentation and Shape Classification in Mammograms using Generative Adversarial and Convolutional Neural Network. arXiv:1809.01687 [cs.CV]

[12] Curated Breast Imaging Subset of DDSM. The Cancer Imaging Archive (TCIA) Public Access. `https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM`.

[13] Suni Kumar. Transfer Learning with ResNet50 for image classification on Cats Dogs dataset. Kaggle. https://www.kaggle.com/suniliitb96/tutorial-keras-transfer-learning-with-resnet50/data.