Deep Learning for Classification and Segmentation of Breast Ultrasound Images

By:

K. K. Kartik
(202SP012)

Under the guidance of:

Dr Shyam Lal

1) Introduction:

Breast cancer is one of the most common forms of cancer effecting woman more than men. According to the statistics reported by ICMR for the year 2020 out of the total 7,12,158 cases of cancer among women, breast cancer accounted for 2,38,908 cases and this is only set to increase massively by 2025. Carol Milgard Breast Centre reports that with early detection of breast cancer the chances of survival are at 93% or higher for the first five years. Hence the breast cancer detection and identifying whether the tumour is malignant or benign is a highly growing research field.

The breast cancer dataset that we have used for our research is a novel dataset made of ultrasound breast images consisting of 780 images along with their ground truths. All the images belong to three classes which malignant, benign and normal. With this dataset my project aims to contribute to the ongoing research for early cancer detection

2) Motivation:

In the year 2017 researchers at Stanford's AI intelligence lab created a dataset consisting of various images of skin lesions and the challenge was to identify and classify the cases which are skin cancer from the other type of skin related ailments. A deep learning model was developed from the same purpose which gave an accuracy of 72%. This deep learning model had outperformed board certified dermatologists at Stanford which led to the revolution of using deep learning models for biomedical imaging processing.

Drawing motivation from this deep learning revolution for biomedical imaging, our aim is to evaluate the performance of existing benchmark deep learning models and to propose a novel algorithm for the classification and segmentation of breast ultrasound cancer images.

3) Problem Statement and Objectives:

Simulation and performance evaluation of existing deep learning models for segmentation and classification of breast ultrasound images.

Design and implementation of robust deep learning model for segmentation of breast ultrasound images

Design and implementation of robust deep learning model for classification of breast ultrasound images

4) Pre-processing and Augmentation of dataset:

All the images in the dataset were pre-processed, the initial stage of pre-processing consists of resizing all the images to the size of 512x512 by using inter-cubic interpolation technique. Finally, we use a medial blur filter with a 5x5 kernel size to filter all the images. Median Blur filter replaces each pixel with its median value. Dataset augmentation was also done to triple the size of the training images in the dataset from 625 images to 1875 images. Rotation by 45 degrees and horizontal flipping were the augmentation techniques used.

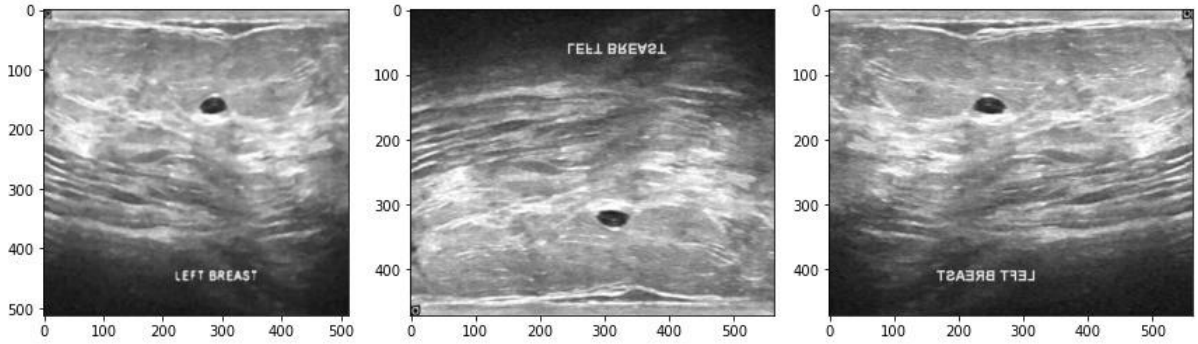

Figure: Augmented and pre-processed images in the dataset

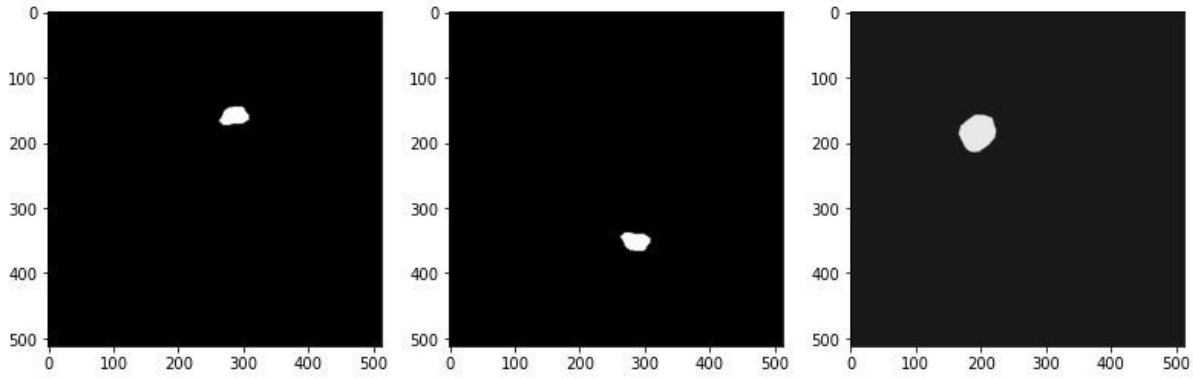

Figure: Corresponding masks for augmented and pre-processed images in the dataset

5) Literature Review:

I have gone through various deep learning models developed over the recent years for my project and I have also gone through the paper on Batch Normalization for improving the performance of the models. I have decided to focus on 3 main deep learning models which are the benchmark models and I will briefly summarise all of them below:

- U-Net
- Segnet
- High Resolution Encoder Decoder Network

- Atrous Spatial Pyramid Pooling

U-Net model for biomedical imaging was first proposed in the year 2015 by scientists from Germany as an improved version of FCN, it had long skip connections to get the spatial information required for segmentation. Segnet was proposed during same time as U-net, its aim was also to improve upon the existing FCN model. Segnet consisted of 13 encoder and corresponding decoder layers which were similar to the VGG16 architecture. In Segnet unlike in U-net, maxpooling indices were stored and these indices were subsequently used while up sampling. In 2019 High Resolution Encoder Decoder model was introduced which was proposed to improve segmentation accuracy of low contrast images by using dilated convolutions instead of normal convolutions and by introducing high resolution pathway as opposed to only long skip connections between the encoder side and the decoder side. Atrous Spatial Pyramid model is another block which improves the performance of the models drastically. Here all the feature maps obtained from the encoder are passes through 3 different convolutional layers and then these feature maps are then combined and fed as an input to the encoder decoder model. We have used integrated Atrous spatial pyramid pooling with the U-net model to obtain better performance metrics

The next section will focus on these benchmark models and I will try to explain all these models briefly.

6) Benchmark Models:

a) U-Net: The first major deep learning model developed for segmentation was the Fully Convolutional Network for segmentation which was proposed in the year 2014. This was a breakthrough paper as it had surpassed the performance of all the models by 20% with little pre-processing steps. The proposed model had encoder decoder type structure and was based on a pre-trained network. This model hence consisted of convolutional operation followed by max-pooling at the encoder side and transpose convolution at the decoder side. Transpose convolution had given segmentation maps at the decoder and it was found that FCN 8 had given the best performance, however there was one flaw that this model which was that the model was segmenting images only based on feature level information and that spatial information was missing. Hence to address this problem in 2015 U-Net model was proposed which introduced long skip connections at various stages of convolution and transpose convolution which provided not only the feature level information but also the spatial information. This model then emerged as the ISBI cell challenge winner.

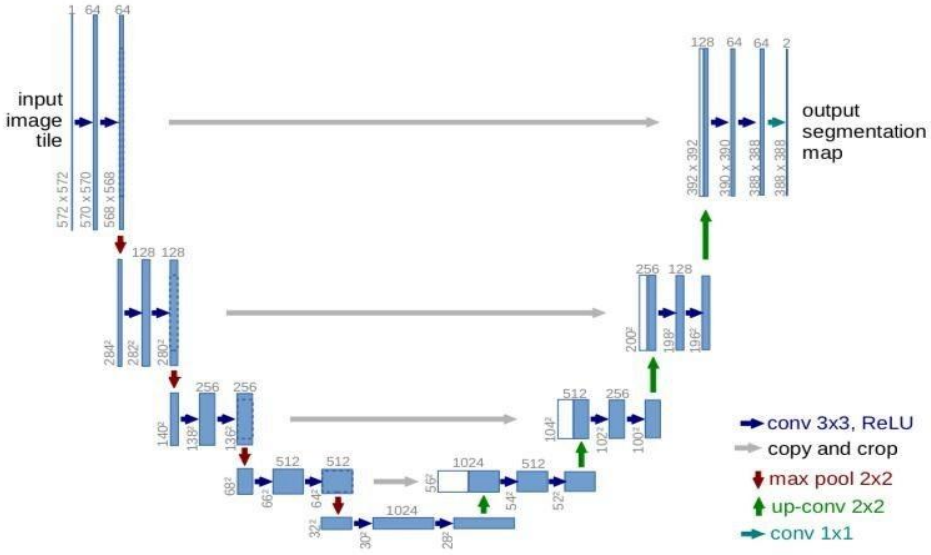

Fig: U-Net architecture, the blue boxes are the multi-channel features and the arrows indicate different operations being performed

b) Segnet: Segnet model was introduced in 2016 around the same time when U-net model came into picture. In Segnet model, there is an encoder network and a corresponding decoder network. The encoder network was like VGG16 model but all the fully connected layers were replaced with fully convolutional layers to reduce the number of parameters. Unlike in U-net where feature maps from encoder were used in the decoder to obtain contextual information. In Segnet the maxpooling indices at the encoder side was saved and was subsequently used in the decoder model to obtain better accuracy.

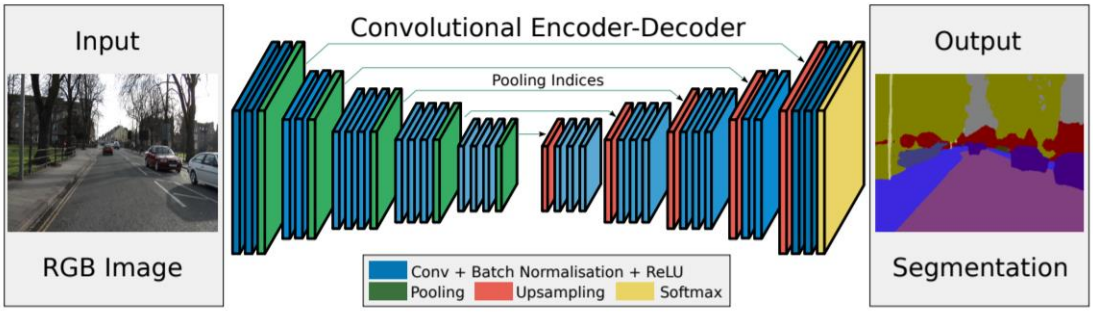

Fig: Block diagram of Segnet architecture

c) High Resolution Encoder Decoder Model: This model was introduced in the year 2019, apart from having various number of skip connections two key things that were present in this model was the high-resolution pathway and the use of dilated convolutions. Dilated convolutions have been derived from the atrous convolution used first in wavelet decomposition. It is a technique that expands the kernel by putting holes between its consecutive elements but it involves pixel skipping. It is like the regular convolution operation but has larger receptive size which results in more efficient

convolution without the loss of information. For example, if we have 2x2 kernel and a 9x9 image then a dilation factor of 2 would change the kernel size to 4x4 and it would result in the output being a 6x6 image. If one were to go by the traditional convolution operation it would require 3 convolution operations to get to the same size. The high-resolution pathway which is used also provides a mechanism for getting the high-resolution contextual information based on the above stated dilated convolution method.

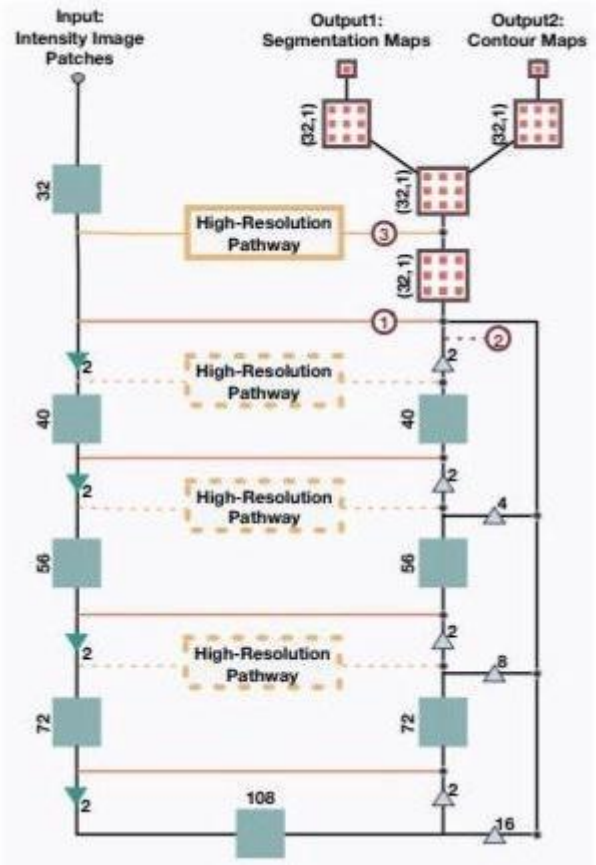

Fig: Schematic block diagram of High-Resolution Encoder Decoder Model

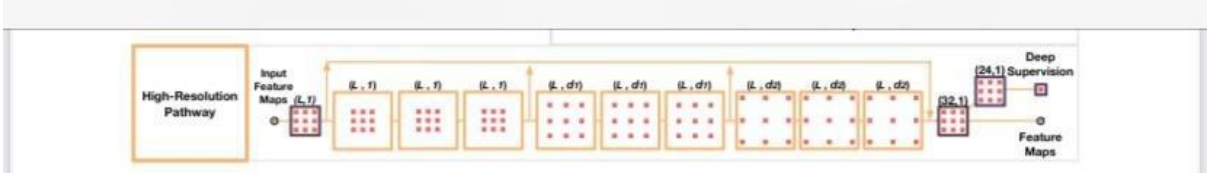

Fig: Schematic Block diagram of the High-Resolution Pathway

d) Atrous Spatial Pyramid Pooling: ASPP consists of several parallel atrous convolutions with different rates. It is a combination of atrous convolution and spatial pyramid pooling, and it can capture the contextual information at multiple scales for a more accurate classification. The ASPP block consists of one 1x1 convolution and three

parallel 3x3 convolutions with different dilation rates. The resulting features from all of the branches are bilinearly up sampled to the input size and then concatenated and passed through another 1 × 1 convolution. This is the ASPP block which is integrated with U-net to improve the accuracy and the performance metrics.

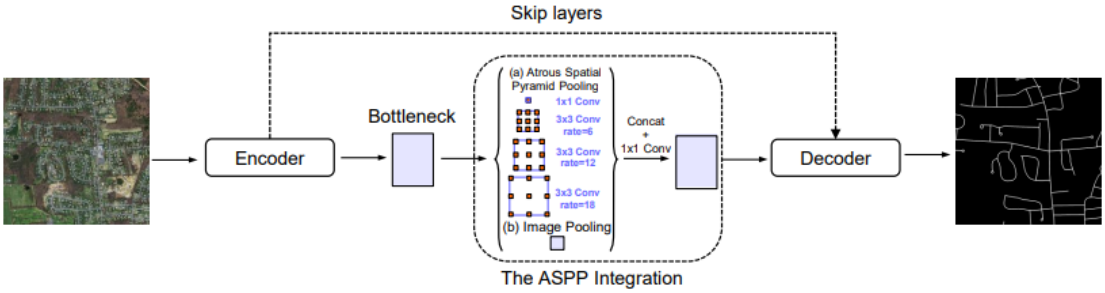

Figure: Schematic Block Diagram of Integration of ASPP model with U-net.

7) Training and Implementation Details:

All the models have been trained for 50 epochs using Adam optimizer and the loss function that was used is the dice loss whose mathematical equation is given by: $1 - \frac{2*X*Y+1}{X+Y+1}$. The metric used for training was accuracy. Relu activation was used for all except the last layer and the final layer used the sigmoid activation function. Weights of all the kernels used in convolution were taken from a gaussian distribution.

8) Simulation Results:

Training vs Validation accuracy, loss plots:

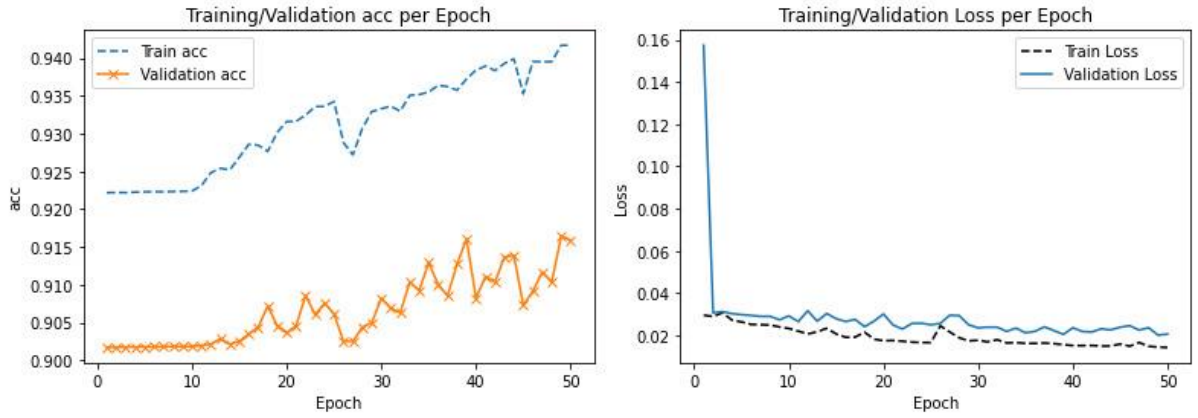

Fig: Training vs Validation accuracy and loss plot for U-Net

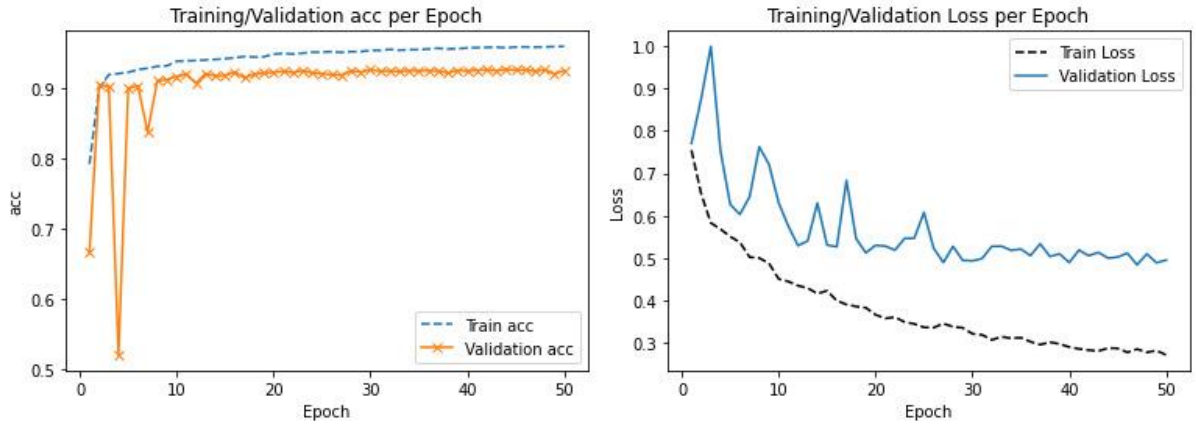

Fig: Training vs Validation accuracy and loss plots for Segnet

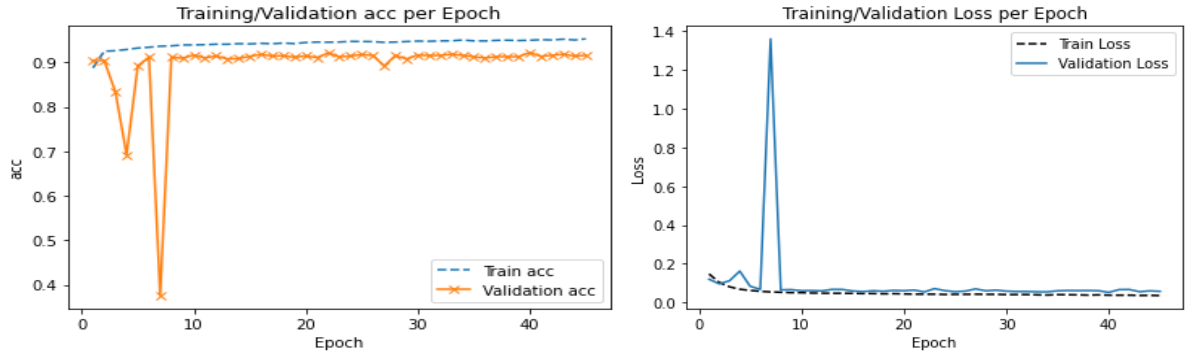

Fig: Training vs Validation Accuracy and loss plots for High Resolution Encoder Decoder Network

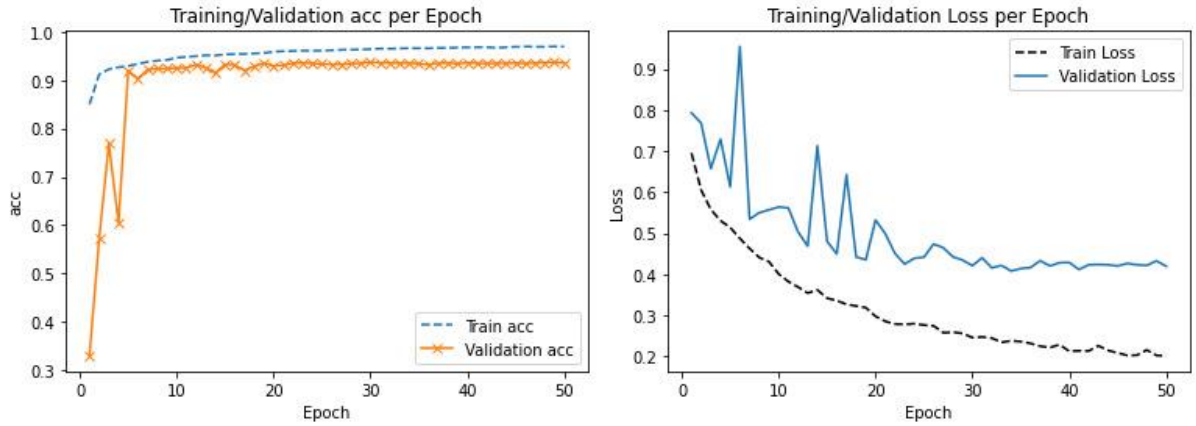

Fig: Training vs Validation Accuracy and loss plots for U-net with ASPP integration

Predicted vs Ground Truth Images for all models:

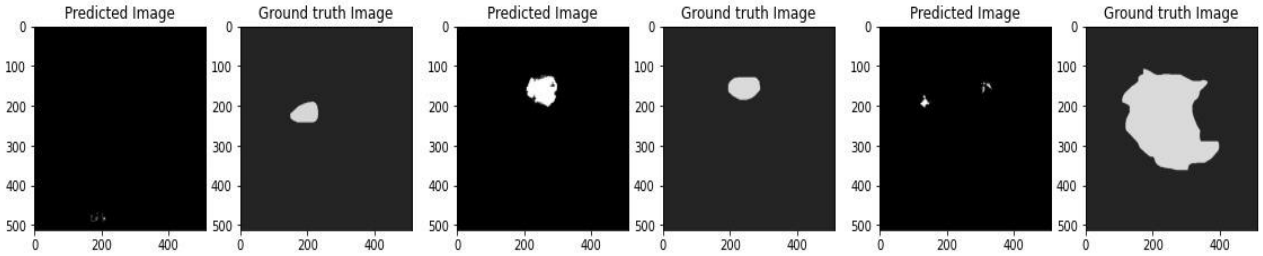

Fig: Predicted Image vs Ground Truth image for U-Net

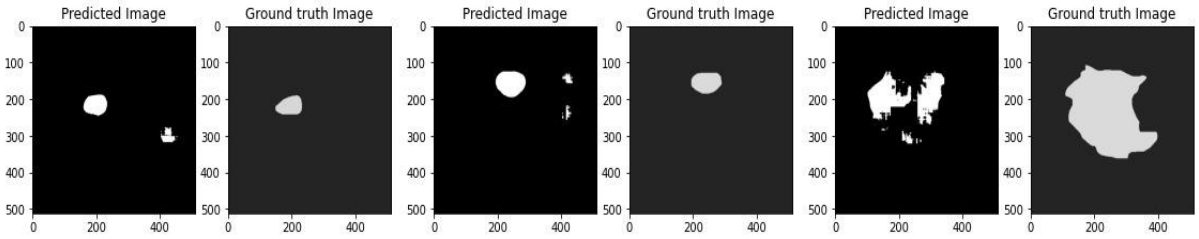

Fig: Predicted Image vs Ground Truth Image for Segnet

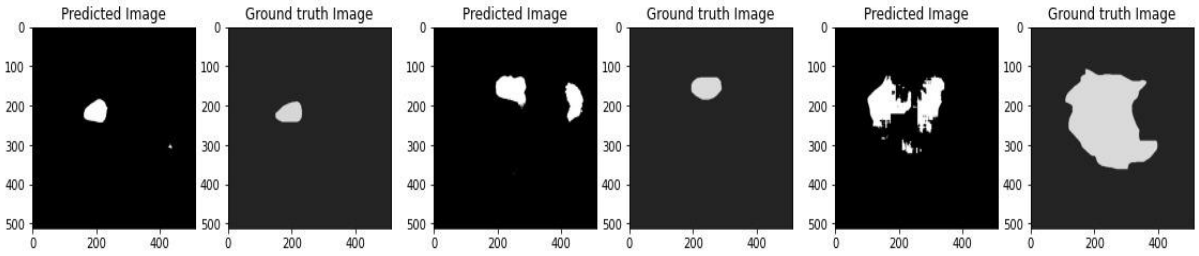

Fig: Predicted Image Vs Ground Truth for High Resolution Encoder Decoder Model

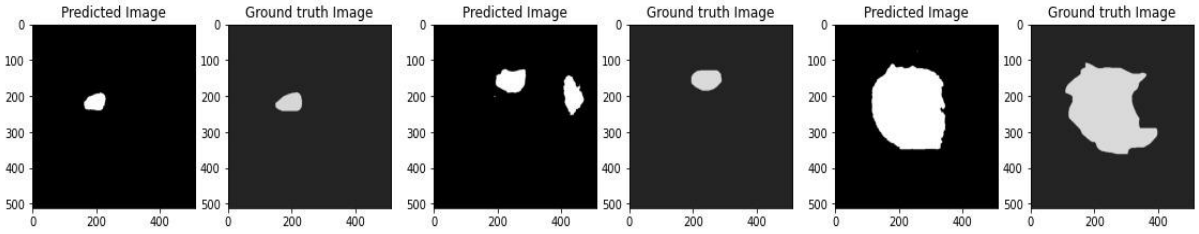

Fig: Predicted Image Vs Ground Truth for Resnet type encoder for Unet with ASPP integration

9) Performance metrics of all the models in percentages:

Model	Dice score	Jaccard Index	Recall	Precision	Final F1	Overall Accuracy
U-Net	31.34	32.83	27.38	68.88	68.88	90.50
Segnet	45.25	44.50	45.52	68.63	60.33	92.54
HRED net	41.74	39.31	40.045	73.74	53.38	91.07
U-Net with ASPP integration	50.12	53.52	49.15	72.46	61.22	93.54

10) Conclusions And Future work to do:

- Overall, the best performing model was U-Net with ASPP integration but it required the highest number of parameters for training
- High Resolution Encoder Decoder model worked had the least number of parameters and has achieved good results
- Implement and explore other models to achieve better performance.
- Already augmentation techniques were used to increase the size of the dataset, will explore more augmentation techniques to increase the size of the dataset further
- Design a custom loss function for this specific problem so that the metrics can be further improved

11) References:

- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
- Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- S. Zhou, D. Nie, E. Adeli, J. Yin, J. Lian and D. Shen, "High-Resolution Encoder–Decoder Networks for Low-Contrast Medical Image Segmentation," in IEEE Transactions on Image Processing, vol. 29, pp. 461-475, 2020, doi: 10.1109/TIP.2019.2919937.

- Drozdzal, Michal, et al. "The importance of skip connections in biomedical image segmentation." *Deep learning and data labeling for medical applications*. Springer, Cham, 2016. 179-187.
- Yu, Fisher, and Vladlen Koltun. "Multi-scale context aggregation by dilated convolutions." *arXiv preprint arXiv:1511.07122* (2015).
- Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *nature* 542.7639 (2017): 115-118.
- Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv 2015." *arXiv preprint arXiv:1511.00561*.
- He H, Yang D, Wang S, Wang S, Li Y. Road Extraction by Using Atrous Spatial Pyramid Pooling Integrated Encoder-Decoder Network and Structural Similarity Loss. *Remote Sensing*. 2019; 11(9):1015. https://doi.org/10.3390/rs11091