

B.Sc. in Electrical and Electronic Engineering Thesis

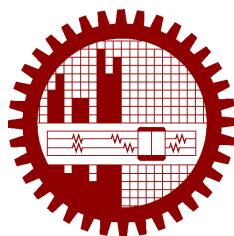
DeepSnake: A Deep Convolutional-Snake Model Combination for Breast Ultrasound Image Segmentation

Submitted by

Sadi Mohammad Siddiquee
201406162

Supervised by

Dr. Md. Kamrul Hasan



**Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology**

Dhaka, Bangladesh

February 2020

CANDIDATES' DECLARATION

This is to certify that the work presented in this thesis, titled, “DeepSnake: A Deep Convolutional-Snake Model Combination for Breast Ultrasound Image Segmentation”, is the outcome of the investigation and research carried out by us under the supervision of Dr. Md. Kamrul Hasan.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Sadi Mohammad Siddiquee

201406162

CERTIFICATION

This thesis titled, “**DeepSnake: A Deep Convolutional-Snake Model Combination for Breast Ultrasound Image Segmentation**”, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in February 2020.

Group Member:

Sadi Mohammad Siddiquee

Supervisor:

Dr. Md. Kamrul Hasan
Professor
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology

ACKNOWLEDGEMENT

I would like to take the opportunity to thank the people whose contribution made it possible for me to complete this thesis. Firstly to my supervisor Dr. Md. Kamrul Hasan for his excellent guidance and overwhelming support. He associated me with other individuals whose advice was vital for this thesis. He also provided me with devices which alleviated the performance of the experiments. His constant encouragement and directive instructions guided us towards the results achieved in this thesis.

Lastly I would like to thank Department of Electrical and Electronic Engineering, BUET for giving me the suitable environment and equipment necessary for the completion of this thesis.

Dhaka

Sadi Mohammad Siddiquee

February 2020

Contents

<i>CANDIDATES' DECLARATION</i>	i
<i>CERTIFICATION</i>	ii
<i>ACKNOWLEDGEMENT</i>	iii
<i>List of Figures</i>	vi
<i>List of Tables</i>	vii
<i>ABSTRACT</i>	viii
1 Introduction	1
1.1 Problem Statement	1
1.2 Objective	2
1.3 Contributions	2
1.4 Organization of the Thesis	3
2 Related Works and Literature Review	4
2.1 Active Contour Model	4
2.1.1 Morphological Snake Model	5
2.2 CNN-based Segmentation Model	6
2.3 Loss Functions	7
2.3.1 Dice Coefficient (DC) Loss	8
2.3.2 Cross-Entropy (CE) Loss	8
2.3.3 Active Contour (AC) Loss	9
3 DeepSnake: A Deep Convolutional-Snake Model Combination	10
3.1 Ensembled CNN model	10
3.1.1 DeepLabV3	11
3.1.2 UNet	13
3.2 Morphological Snake Model	14
3.3 Weighted BCE-Dice-AC loss	14

4 Experimental Results	16
4.1 Dataset	16
4.2 Training Protocols	17
4.2.1 Evalutaion Matrics	17
4.2.2 Baseline Models	18
4.3 Quantitative Evaluation	19
5 Conclusion	21
5.1 Concluding Remarks	21
5.2 Limitations and Future Directions	21
References	23
A Performance Visualization	27

List of Figures

2.1 Segmentation error due to initialization.	5
2.2 Lack of spatial resolution in CNN-based model prediction.	7
3.1 DeepSnake architecture.	11
3.2 Atrous Convolution with Different Rates r	12
3.3 Atrous Spatial Pyramid Pooling (ASPP)	12
4.1 Deeplabv3 and Unet convergence	18
4.2 Performance Comparison	20
A.1 Test Image 1	27
A.2 Test Image 2	28
A.3 Test Image 3	28
A.4 Test Image 4	29
A.5 Test Image 5	29
A.6 Test Image 6	30
A.7 Test Image 7	30
A.8 Test Image 8	31
A.9 Test Image 9	31
A.10 Test Image 10	32
A.11 Test Image 11	32
A.12 Test Image 12	33
A.13 Test Image 13	33

List of Tables

4.1 Comparison of CNN-based models' performance on BUSI test set	19
4.2 Comparison of performance on both BUSI and BUSIS test set	19

ABSTRACT

Segmentation of breast tumor cell is needed for patients aid. It is helpful for radiologist's aid in quick scoping and lessening the variability of different radiologists. For breast image segmentation, two widely used conventional segmentation method are - Active Contour Model and Convolutional Neural Network(CNN)-based model. In this work, we revisit this long established approach for segmentation - active contour model. The problem with former active contour models is that they are image gradient dependent and strongly depend on initialization. That is why first we employ cnn-based models(ensembled baseline) to locate and roughly segment the hot-portion. CNN-based model are good at localization but they lack spatial resolution and shape information in there prediction. So this output from cnn-model is then fed into morphological snake model for better contour shaped segmentation. This erases active contour model's initialization and localization problem, also exports a high spatial information based and fine boundary-wise prediction. Our experiments show that this leads to a highly effective architecture that produces sharper predictions around object boundaries and significantly boosts performance on thinner and smaller objects. Also this retrieves more spatial resolution than the baseline models. Our method achieves state-of-the-art performance on the BUSI and BUSIS benchmark, in terms of mask quality (mIoU), improving by 6% and 11% over strong baselines.

Chapter 1

Introduction

Breast Cancer occurs in the highest frequency in women among all cancers, and is also one of the leading cause of cancer death worldwide [1]. South Asian countries like Bangladesh, India are facing a hidden breast cancer epidemic [2]. So the key to reduce the mortality is to find the signs and symptoms of breast cancer at its early stage. In current clinic practice, breast ultrasound (BUS) imaging with computer-aided diagnosis (CAD) system has become one of the most important and effective approaches for breast cancer detection due to its noninvasive, painless, non-radioactive and cost-effective nature. In addition, it is the most suitable approach for large-scale breast cancer screening and diagnosis in low-resource countries and regions.

Thus segmentation of breast tumor cell is needed for patients aid. These days automatic segmentation and classification are on par with that of average, but dedicated, breast radiologists on heterogeneous datasets of mammograms [3]. So it has been quite a journey from deep learning based first state of the art segmentation research [4] to AI aided tele-radiology company [5].

1.1 Problem Statement

Ultrasound Imaging has a large part in medical diagnostics. It has an excellent safety record. It is based on non-ionizing radiation, so it does not have the same risks as X-rays or other types of imaging systems that use ionizing radiation. Breast Ultrasound request by doctors for breast lump, cancer, breast implants, breast pain, redness, and swelling etc. Mostly it is prescribed in case of breast lump which is the preliminary symptom of breast cancer.

This work is on Ultrasound Breast Segmentation. Breast ultrasound segmentation is needed for radiologist's aid in quick scoping and lessening the variability of different radiologists. Segmentation work is also amplify exact classification of cancer cell as this can help in pooling

correct feature map needed for classification. Two different dataset named as BUSIS [6] and BUSI [7] were used in this work. This datasets carry both pre-processed Ultrasound B mode image and a ground truth showing the region to segment.

1.2 Objective

Contour models were one of first methods used in medical image segmentation. It was first introduced by Kass *et al.* [8]. It uses a method to minimize the energy function from an external and internal force. Through the first decade of twenty first century, Researchers has modified this algorithm which has improved segmentation. But as the computation power begin rise, Machine learning and Deep learning methods became more popular tool.

These days, most of the related works use deep learning models. But a large portion of these models algorithm are generated based on common object in context detection or segmentation. The ImageNet, Pascel VOC or COCO datasets [9–11] are the objective for these works. But breast ultrasound segmentation is a medical imaging problem. A medical image is different in sense that it bears more data variability due to data acquisition noise and object’s variable nature [12]. And medical image analysis need to be more accurate then all other field. So the aim was -

- Reconstruction of the state of the art deep learning models’ result.
- Propose a optimized algorithm/method which will generate better spatial resolution based segmentation and will improve the above mentioned result.
- Comparison of state of the art models and proposed method’s result.

1.3 Contributions

We propose *DeepSnake*, a *novel* ultrasound image segmentation method which integrates classical active contour model and training based deep learning model. As deep learning model is shaky at preserving shape information, spatial resolution in segmentation task, a segment with stiff contour is generated. Many proposed specialized CNN modules [13–15] that help restore the spatial resolution of the network output.

We argue here that there is an inherent inefficacy in the architecture design since color, shape and texture, spatial resolution information are all processed together inside one deep CNN. So we introduce a cascaded *deep learning*(pre-trained on BUSI dataset) and *morphological snake*

model(modified active contour model) to reconstruct spatial resolution.

Our main contributions include-

- State of the art models' result analysis.
- Cascading Morphological Snake Model with the baseline pre-trained CNN-based model for better contour enriched segmentation.
- Applying ensembled CNN models as baseline model. So that it takes more spatial information to morphological snake model.
- Comparison of the state of the art models and proposed method's result.

Integrating a deep learning model and snake model for better spatial resolution based segmentation was the main theme of the work.

1.4 Organization of the Thesis

The remainder of the book is organized as follow, In chapter [2](#) we review related recent literature. We describe the *DeepSnake* architecture and its analysis in chapter [3](#). In chapter [4](#) we evaluate the performance of *DeepSnake* on BUSI and BUSIS breast ultrasound datasets. We conclude in chapter [5](#).

Chapter 2

Related Works and Literature Review

In the following sections, we review semantic segmentation methods used in past works. Firstly widely used conventional segmentation method - *Active Contour Model* is surveyed. Then we explore *CNN-based* state of the art segmentation methods. Finally different *loss functions* used in segmentation are also reviewed.

2.1 Active Contour Model

Active contours, or snakes, are computer-generated curves that move within images to find object boundaries. The active contour model is a method to minimize the energy function,

$$E = E_{internal} + E_{external} \quad (2.1)$$

where $E_{internal}$ is the internal energy of the active contour that quantifies the contour smoothness:

$$E_{internal} = \frac{1}{2} \int_0^1 \left\{ \alpha \left[\left(\frac{dX(s)}{ds} \right)^2 + \left(\frac{dY(s)}{ds} \right)^2 \right] + \beta \left[\left(\frac{d^2X(s)}{d^2s} \right)^2 + \left(\frac{d^2Y(s)}{d^2s} \right)^2 \right] \right\} \quad (2.2)$$

Here, α and β are two nonnegative weighting parameters expressing, respectively, the degree of the resistance to stretching and bending of the contour. The external energy $E_{external}$ is typically defined such that the contour seeks the edges in the image, I :

$$E_{external} = - \int_0^1 f(X(s), Y(s)) ds \quad (2.3)$$

where $f(x, y) = |\nabla I(x, y)|^2$, to minimize E , different types of algorithm like *pde*, *morphological operator* are used.

It was first introduced by Kass *et al.* [8]. An external force is measured from image data, and an internal force is defined as curves or surfaces. The external force is a force that enables initial contours (refer to those curves or surfaces) to change their shapes automatically into the shapes of objects in the images.

Snake or active contours are extensively used in medical image processing, and particularly to locate object boundaries. However, there are still limitations related to initialization and poor convergences. Choi and Kim (2010) [16] presented an automatic initialization value of the snake algorithm for the segmentation of the chest wall. Their method of external force for active contours, which is called gradient vector flow (GVF), was introduced to address the problems. However, the use of active contour model in this research is limited in the chest wall by the chest CT.

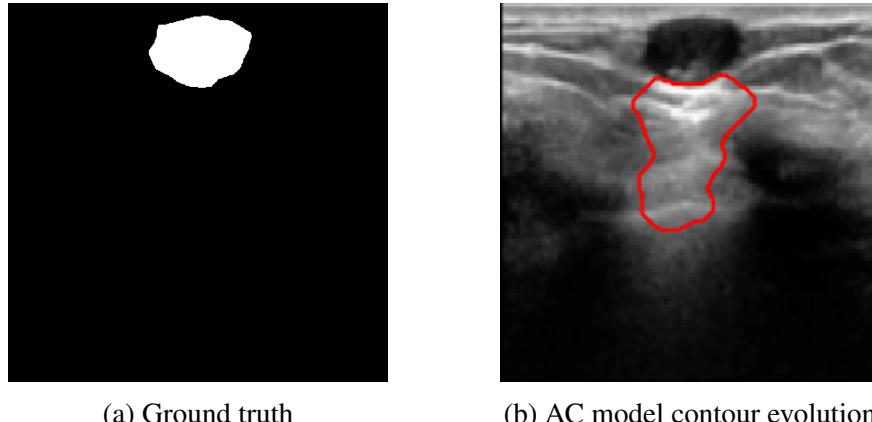


Figure 2.1: Segmentation error due to initialization.

2.1.1 Morphological Snake Model

P Márquez-Neila (2013) [17] introduced new results connecting differential and morphological operators that provide a formal and theoretically grounded approach for stable and fast contour evolution. The standard solution based on partial differential equations and level-sets requires the use of numerical methods of integration that are costly computationally and may have stability issues. Here a morphological approach was demonstrated to contour evolution based on a new curvature morphological operator valid for surfaces of any dimension. The numerical solution of the curve evolution PDE was approximated by the successive application of a set of morphological operators defined on a binary level-set and with equivalent infinitesimal behavior. These operators are very fast, do not suffer numerical stability issues and do not degrade the level set function, so there is no need of re-initializing it. Moreover, their implementation is much easier since they do not require the use of sophisticated numerical algorithms.

2.2 CNN-based Segmentation Model

As a class of deep neural networks, CNNs show remarkable performance in many computer vision tasks, such as classification, segmentation and registration. One of the particular strength of CNN based models is that they work in an end-to-end fashion, which can extract hierarchical and multi-resolution features during the learning process. CNN architectures like Alex-Net, VGG-Net, GoogleNet and Dense-Net [18–21] have been developed and introduced into various image recognition tasks. Broadly speaking, CNN based segmentation models can be classified into pixel-based or images-based approaches. The pixel-based approaches will classify each pixel into different objects as a classification problem. A patch is often produced for each pixel (or super-pixel) and the patch is used as input to CNN models for classification with the label of the pixel used as the target to train the model [22].

Earlier approaches [4, 23] convert classification networks into fully convolutional networks (FCNs) for efficient end-to-end training for semantic segmentation. The FCN network pipeline is an extension of the classical CNN. The main idea is to make the classical CNN take as input arbitrary-sized images. The restriction of CNNs to accept and produce labels only for specific sized inputs comes from the fully-connected layers which are fixed. Contrary to them, FCNs only have convolutional and pooling layers which give them the ability to make predictions on arbitrary-sized inputs. One issue in this specific FCN is that by propagating through several alternated convolutional and pooling layers, the resolution of the output feature maps is down sampled. Therefore, the direct predictions of FCN are typically in low resolution, resulting in relatively fuzzy object boundaries. A variety of more advanced FCN-based approaches have been proposed to address this issue, including SegNet, UNet, DeepLab, and Dilated Convolutions [13, 24–26].

The image-based approaches, such as U-Net [25], uses an image as input which outputs the segmentation of the input image (the size will be the same). U-Net like models have become popular in medical imaging because of its good performance and simplicity when compared to pixel-wise approaches [27, 28]. However, due to subsequent use of pooling layer, information get lost. So Deeplab [26] tried to solve this by using *Atrous Spatial Pyramid Pooling* and *Atrous Convolution*.

Consider two-dimensional signals, for each location i on the output y and a filter w , atrous convolution is applied over the input feature map x :

$$y[i] = \sum_k x[i + r \cdot k]w[k] \quad (2.4)$$

where the atrous rate r corresponds to the stride with which we sample the input signal, which is equivalent to convolving the input x with upsampled filters produced by inserting $r - 1$ zeros between two consecutive filter values along each spatial dimension.

Deeplab also implemented the Atrous Spatial Pyramid Pooling (ASPP), revised version of spatial pyramid pooling proposed in [29]. Here four parallel atrous convolutions with different atrous rates are applied on top of the feature map. ASPP is inspired by the success of spatial pyramid pooling which showed that it is effective to resample features at different scales for accurately and efficiently classifying regions of an arbitrary scale. But still it's shaky at preserving shape information, spatial resolution in segmentation task, so a rigid shaped mask is generated.

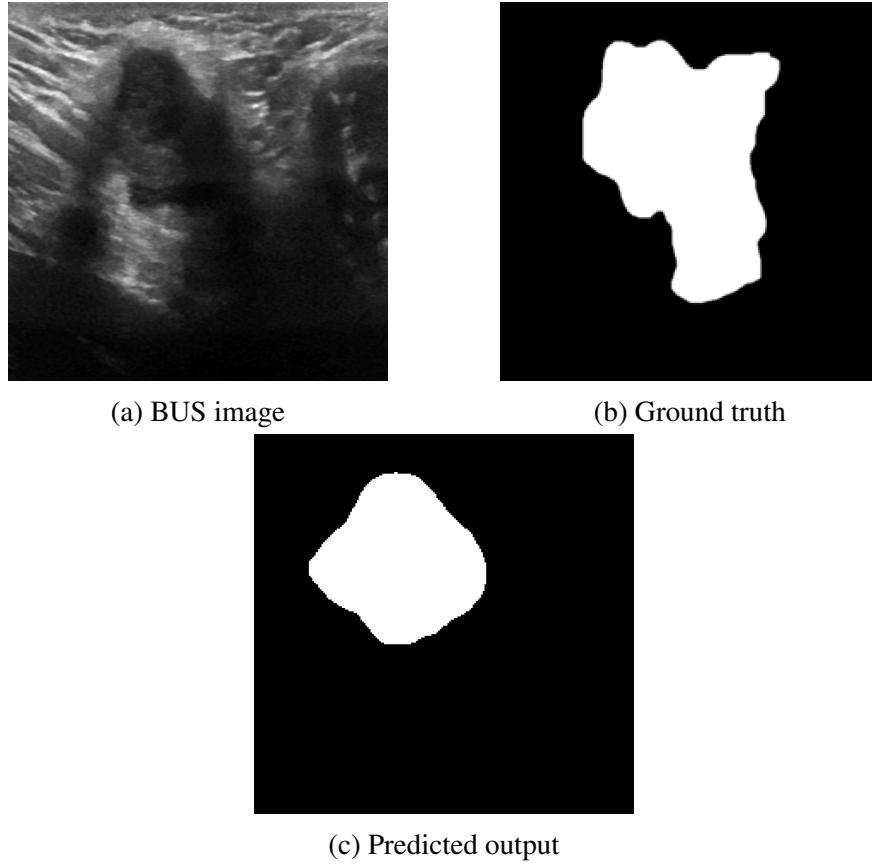


Figure 2.2: Lack of spatial resolution in CNN-based model prediction.

2.3 Loss Functions

Loss function (i.e. cost function) is a function to measure the error of prediction or segmentation which can be back propagated to previous layers in order to update/optimize the weights. Here, we briefly review the commonly used loss functions. In the following equations, ground truth

image (or expert annotation) and the prediction (or segmentation) is denoted as $T, P \in [0, 1]^{m \times n}$ respectively; n indexes each pixel value in image spatial space N ; the label of each class is written as l in C classes.

2.3.1 Dice Coefficient (DC) Loss

DC along with IOU is traditionally used as a metric for the evaluation of the segmentation performance [30] and now also demonstrated a good performance as a loss function. DC measure the degree of overlapping between the reference and segmentation. This element-wise measure ranges from 0 to 1 where a DC of 1 denotes perfect and complete overlap. DC can be defined as:

$$DC(T, P) = 2 \cdot \frac{\sum_{n=1}^N (T_n \times P_n)}{\sum_{n=1}^N (T_n + P_n)} \quad (2.5)$$

DC loss is defined in Eq. (2.6) as it tends to the best segmentation.

$$Loss_{DC}(T, P) = 1 - DC(T, P) \quad (2.6)$$

2.3.2 Cross-Entropy (CE) Loss

CE is a widely used pixel-wise measure [24] to evaluate the performance of classification or segmentation model. For two-class problems, CE loss function can be expressed as Binary-CE (BCE) loss function as follows:

$$Loss_{BCE}(T, P) = -\frac{1}{N} \sum_{n=1}^N [T_n \cdot \log(P_n) + (1 - T_n) \cdot \log(1 - P_n)] \quad (2.7)$$

BCE loss functions treat the output from sigmoid layer as a pixel classification problem to evaluate each pixel. Ronneberger *et al.* [25] pointed out that in order to improve the performance in cells' border segmentation from biomedical images, BCE loss function with weighting scheme can be as one of the solutions to help U-Net model segments cells border as accurately as possible. Moreover, there are numerous studies on CE-based loss functions but merely a few functions consider the geometric detail of objects [31].

Even though CE and DC loss functions have achieved a success in image segmentation, there are two main limitations: they are pixel-wised loss functions to measure the similarity between T and P , but the geometrical information are not taken into consideration.

2.3.3 Active Contour (AC) Loss

The idea of AC loss [32] was behind the minimization problem of ACWE model [33] to efficiently find an active contour which is a global minimization of active contour energy for automated image segmentation. A 2-dimensional example of AC loss function is defined as follows:

$$Loss_{AC} = Length + \lambda Region \quad (2.8)$$

where

$$Length = \sum_{\Omega}^{i=1, j=1} \sqrt{|(\nabla P_{x_{i,j}})^2 + (\nabla P_{y_{i,j}})^2|} \quad (2.9)$$

and

$$Region = \left| \sum_{\Omega}^{i=1, j=1} P_{i,y} (c_1 - T_{i,j})^2 \right| + \left| \sum_{\Omega}^{i=1, j=1} (1 - P_{i,y}) (c_2 - T_{i,j})^2 \right| \quad (2.10)$$

c_1 and c_2 are represented as the energy of inside (foreground) and outside (background) and simplified as constants in advance as $c_1 = 1$ and $c_2 = 0$.

Chapter 3

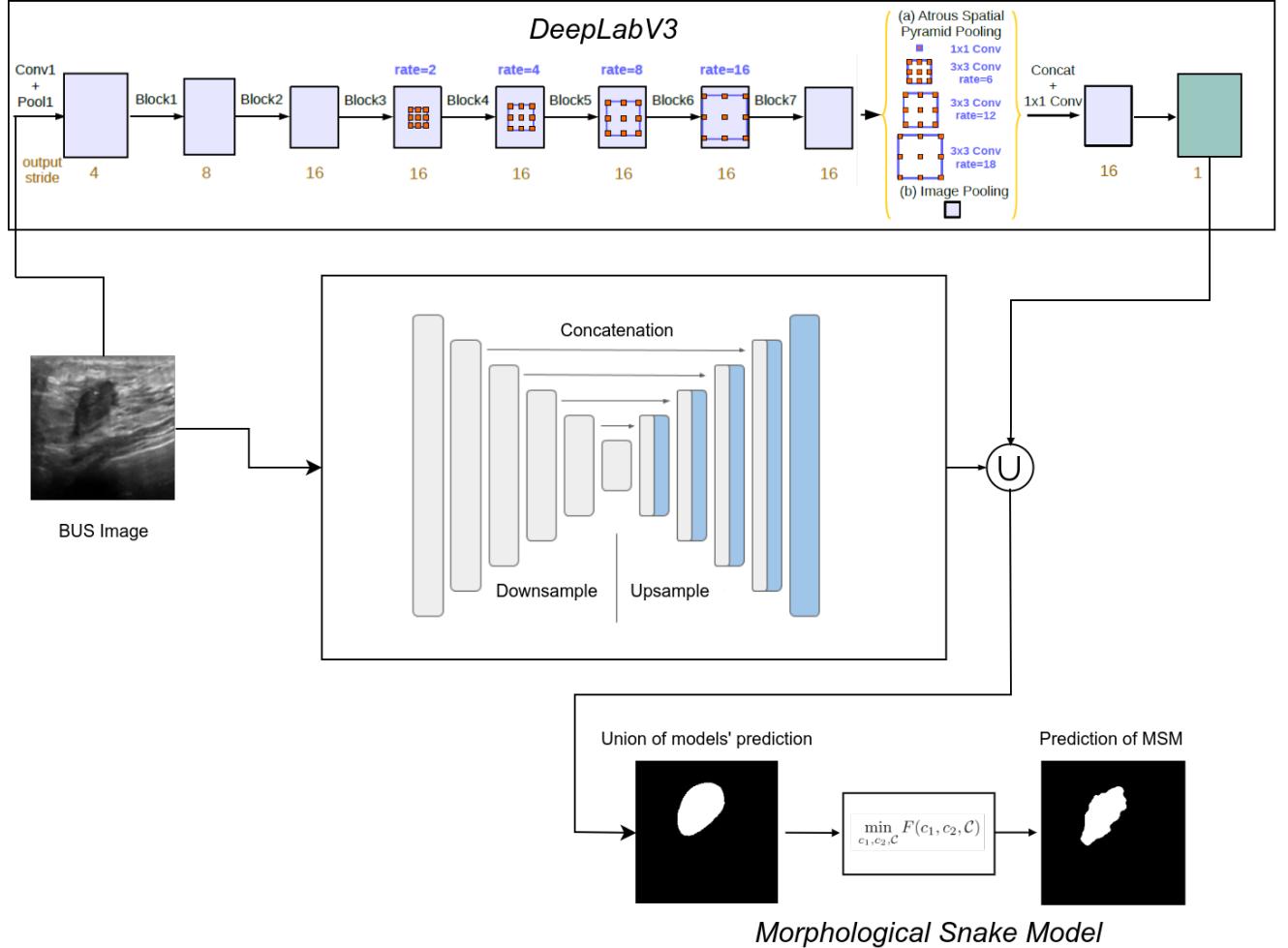
DeepSnake: A Deep Convolutional-Snake Model Combination

In this chapter, we discuss the details of our novel proposed method *DeepSnake*. In section 3.1, we discuss our Ensembled Model used as baseline model, which helps to localize the segment in BUSI image. Subsection 3.1.1 and 3.1.2 are used to describe two of the models used for ensemble. We then discuss the proposed modules with atrous convolution modules employed in cascade or in parallel. Then in section 3.2, we narrate Morphological Snake Model(MSM) used for better spatial resolution. Finally in section 3.3, we report our loss scheme for best case solution.

Our architecture constitutes of two main parts. Ensembled CNN model(Deeplabv3 & UNet) and morphological snake model. The ensembled model can be any cnn-based model, here best two models are used because of constraint in computation power. A fusion module later combines information from the two cnn models. And about contour model, morphological snake model are used morphological operator which is quite fast. So this serves our purpose.

3.1 Ensembled CNN model

As cnn-based models are well known for localizing, two cnn-based models were ensembled here. Different cnn-based models like SegNet [24], FCN [4], Gated-SCNN [34], Deeplab v1-v3 [26, 35], UNet [25] were used. From these best two performing models were ensembled. These two models were first trained individually on BUSI [7] dataset, then tested on both BUSI and BUSIS [6] datasets. In the following subsections, we will discuss the best performing DeeplabV3 and UNet models.

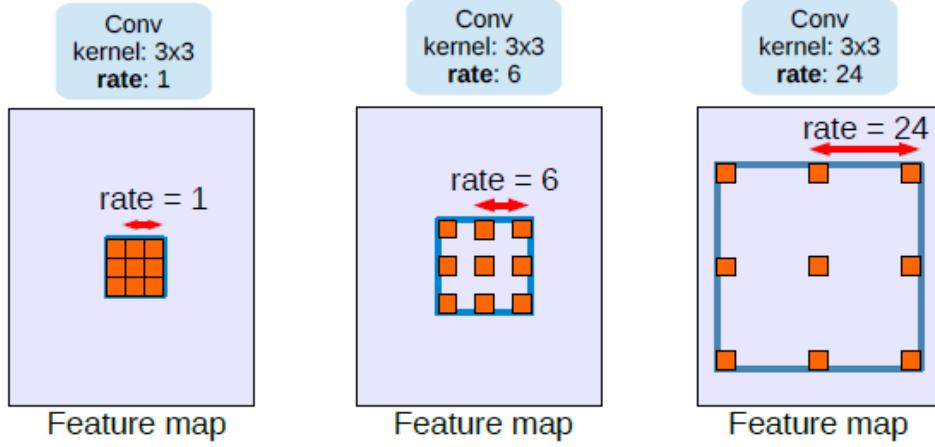
Figure 3.1: **DeepSnake** architecture.

3.1.1 DeepLabV3

In this work, we revisit DeeplabV3. Main two idea of deeplab was Atrous Convoultion and Atrous Spatial Pyramid Pooling. Atrous convolution allows us to effectively enlarge the field of view of filters to incorporate multi-scale context, in the framework of both cascaded modules and spatial pyramid pooling. In particular, it consists of atrous convolution with various rates and batch normalization layers which we found important to be trained as well. It also visit spatial pyramid pooling [14] and modified it as Atrous spatial pyramid pooling which incorporate with information loss because of pooling.

Atrous Convolution

Atrous convolution, originally developed for the efficient computation of the undecimated wavelet transform in the “algorithme à trous” scheme of [36] and used before in the DCNN context by [37].

Figure 3.2: Atrous Convolution with Different Rates r

Atrous convolutions introduce another parameter to convolutional layers called the atrous rate. This defines a spacing between the values in a kernel. A 3x3 kernel with a dilation rate of 2 will have the same field of view as a 5x5 kernel, while only using 9 parameters. Imagine taking a 5x5 kernel and deleting every second column and row. This delivers a wider field of view at the same computational cost.

Atrous Spatial Pyramid Pooling

ASPP has been introduced in DeepLabv2 [35]. In v3, batch normalization (BN) from Inception-v2 is included into ASPP. The reason of using ASPP is that it is discovered as the sampling rate becomes larger, the number of valid filter weights (i.e., the weights that are applied to the valid feature region, instead of padded zeros) becomes smaller. Also, image pooling, or image-level feature, by ParseNet [38], is also included for global context.

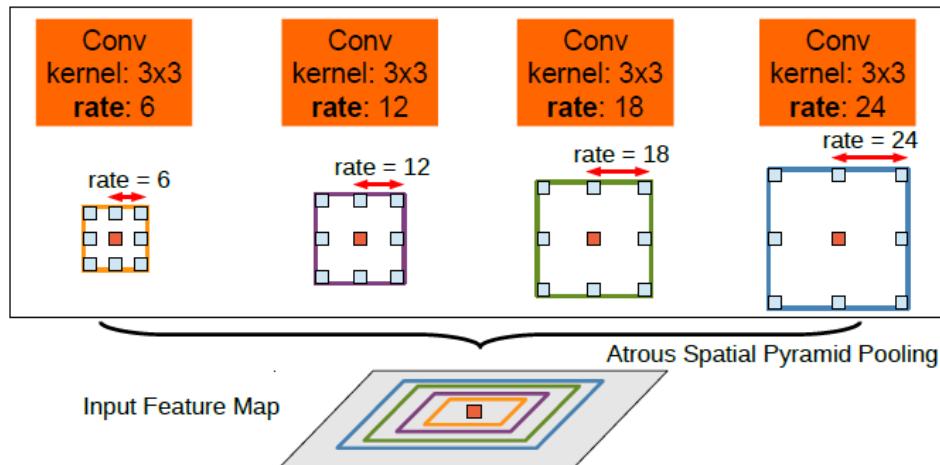


Figure 3.3: Atrous Spatial Pyramid Pooling (ASPP)

The resulting features from all the branches are then concatenated and pass through another 1×1 convolution (also with 256 filters and batch normalization) before the final 1×1 convolution which generates the final logits.

3.1.2 UNet

UNet uses the classical idea of autoencoders [39] and extends it by using encoding information at the decoding or upsampling side. In a classical autoencoder architecture, the size of the input information is initially reduced, along with the following layers. At this point, the encoder part of the architecture is completed and the decoder part begins. Linear feature representation is learned in this section, and the size gradually increases. At the end of the architecture, the output size is equal to the input size. This architecture is ideal in preserving the output size, but one problem is that it compresses the input linearly, which results in a bottleneck in which all features cannot be transmitted.

This is where U-Net differs. U-Net performs deconvolution on the decoder side (i.e. in the second half) and, in addition, can overcome this bottleneck problem, which results in the loss of features through connections from the encoder side of the architecture. We also tried using noisy activation function described in Gulcehre *et al.* [40]. But it did not work in our case.

Encoder

The contracting path is composed of 4 blocks. Each block is composed of

- 3x3 Convolution Layer + activation function (with batch normalization)
- 3x3 Convolution Layer + activation function (with batch normalization)
- 2x2 Max Pooling

Note that the number of feature maps doubles at each pooling, starting with 64 feature maps for the first block, 128 for the second, and so on. The purpose of this contracting path is to capture the context of the input image in order to be able to do segmentation. This coarse contextual information will then be transferred to the upsampling path by means of skip connections.

Bottleneck

This part of the network is between the contracting and expanding paths. The bottleneck is built from simply 2 convolutional layers (with batch normalization), with dropout.

Decoder

The expanding path is also composed of 4 blocks. Each of these blocks is composed of

- Deconvolution layer with stride 2
- Concatenation with the corresponding cropped feature map from the contracting path
- 3x3 Convolution layer + activation function (with batch normalization)
- 3x3 Convolution layer + activation function (with batch normalization)

The purpose of this expanding path is to enable precise localization combined with contextual information from the contracting path.

3.2 Morphological Snake Model

Chan and Vese [33] define an energy functional for image segmentation which takes into account the content of the interior and exterior regions of the curve (or surface) in contrast to the GAC(geodesic active contour) [41], which only take into account the places where the curve (or surface) passes. That is why ACWE can be used to segment objects in images and volumes without well defined borders. On the other hand GAC can be used to segment objects with visible but noisy, cluttered, broken borders. The ACWE functional of a curve C is,

$$\begin{aligned} F(c_1, c_2, C) = & \mu \cdot \text{length}(C) + \nu \cdot \text{area}(\text{inside}(c)) + \lambda_1 \int_{\text{inside}(C)} \|I(x) - c_1\| dx \\ & + \lambda_2 \int_{\text{outside}(C)} \|I(x) - c_2\| dx \end{aligned} \quad (3.1)$$

where the non-negative parameters μ , ν , λ_1 and λ_2 control the strength of each term. The minimization of function is

$$\min_{(c_1, c_2, C)} F(c_1, c_2, C) \quad (3.2)$$

Here c_1 and c_2 are initialized as 1 and 0.

3.3 Weighted BCE-Dice-AC loss

We used weighted binary cross entropy(bce), dice and active contour(ac) loss [32]. BCE loss is good for shape optimization but it does not contain how spatial information changes that is

why ac loss was used. And dice loss is needed for localization and mapping the segment. But in our experiment, we found that ac loss does not improve in our case. This may be because of it worsen the localization.

Chapter 4

Experimental Results

In this chapter, we provide our training scheme and an extensive evaluation of our framework on the challenging BUSI [7] and BUSIS [6] dataset. In section 4.1, we narrate two used BUS image dataset. Then in following section 4.2, we provide our full training scheme. We finally evaluate our model in section 4.3, where we compare our approach with state of the art benchmarking models.

4.1 Dataset

All of our experiments are conducted on the BUSI [7] and BUSIS [6] dataset. BUSIS dataset is from 2018. This dataset contains 562 images which are separately annotated by four radiologists and finally curated by majority voting count. But only a portion of this dataset are public [42]. This public portion of dataset contains 261 images with annotated contour ground-truth mask, from where we generated ground-truth mask using python PIL package. These images were collected by the Second Affiliated Hospital of Harbin Medical University, the Affiliated Hospital of Qingdao University, and the Second Hospital of Hebei Medical University using multiple ultrasound devices: GE VIVID 7 and LOGIQ E9, Hitachi EUB-6500, Philips iU22, and Siemens ACUSON S2000.

The other dataset, BUSI [7] came in late 2019. This dataset contains total 780 images, where no of benign case is 437, no of malignant case is 210, no of normal case is 133. The original images contain unimportant information not used for mass classification, that is why the cropped and published pre-processed version of images. The instruments used in the scanning process are LOGIQ E9 ultrasound system and LOGIQ E9 Agile ultrasound system.

As BUSI dataset contains more annotated images, we used 95% images of BUSI dataset(only

benign and malignant, so total 600 images) for training purpose. Remaining 5% was used for testing and visualization in section 4.3. We further examine our approach by testing it on BUSIS dataset. These results are also shown on previously mentioned section.

4.2 Training Protocols

We used BUSI dataset's 95% data as training set and 5% for testing. Of this 95% training set 15% were used for validation. Training procedure was simple. Hyper-parameter's were-

- BUS images were cropped to 224×224 .
- Learning rate was varied between 10^{-3} to 10^{-6} .
- Weighted bce-dice loss was used during training process.
- Around 200 to 400 epoch was run for each model convergence.
- Batch Size was around 20-40 depending on the model size and computational power.
- Both SGD and Adam optimizer was used depending on model(momentum = 0.9)
- Image pre-processing like brightness-contrast adjustment was used, but they did not make any difference as cnn-based model are mostly invariant to them. Also 2d *Cepstrum* [43] was used for pre-processing. But it did not make any improvement.

4.2.1 Evalutaion Metrics

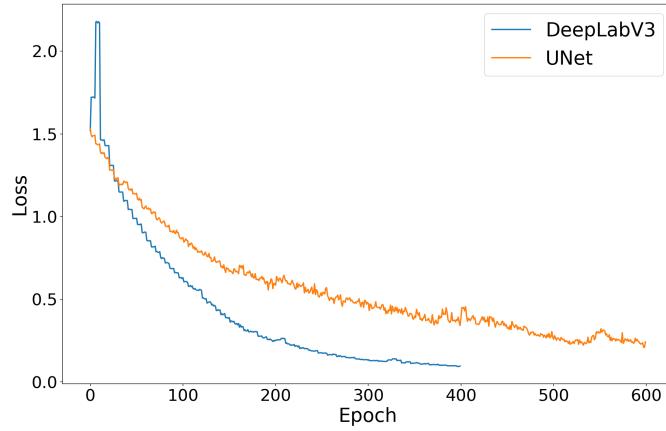
We use two quantitative measures to evaluate the performance of our approach. **1)** We use the widely used intersection over union (IoU) to evaluate whether the network accurately predicts regions. The IoU, also known as the Jaccard Index, is one of the most commonly used metrics in semantic segmentation... and for good reason. The IoU is a very straightforward metric that's extremely effective. IoU is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth, as shown on the image to the left. This metric ranges from 0–1 (0–100%) with 0 signifying no overlap (garbage) and 1 signifying perfectly overlapping segmentation (fat dub).

$$IoU = \frac{\text{target} \cap \text{prediction}}{\text{target} \cup \text{prediction}} \quad (4.1)$$

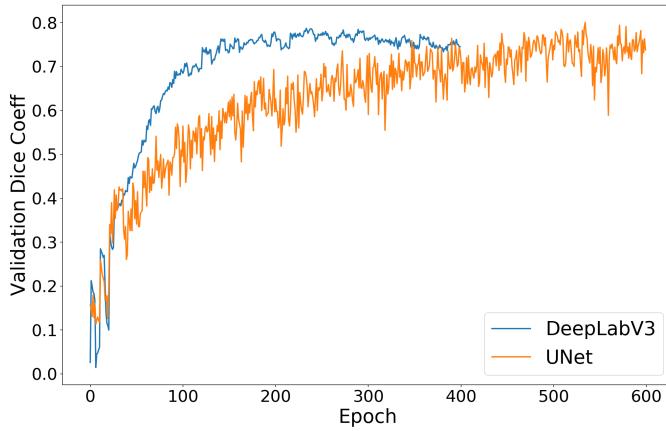
2) Second evaluation metric is Dice Coefficient [44]. The Dice coefficient is very similar to the IoU. They are positively correlated, meaning if one says model A is better than model B at

segmenting an image, then the other will say the same. Like the IoU, they both range from 0 to 1, with 1 signifying the greatest similarity between predicted and truth.

$$Dice = 2 \cdot \frac{|target \cap prediction|}{|target| + |prediction|} \quad (4.2)$$



(a) Loss convergence



(b) Dice coef convergence

Figure 4.1: Deeplabv3 and Unet convergence

4.2.2 Baseline Models

We experimented different cnn-based model like SegNet [24], FCN [4], Gated-SCNN(GSCNN) [34], Deeplab v1-v3 [26, 35], UNet [25]. Here is our comparison of dice score are shown.

Table 4.1: Comparison of CNN-based models' performance on BUSI test set

<i>Model Name</i>	<i>mDice</i>
SegNet	56.6
DeepLabV3	60.1
FCN	46.2
UNet	62.7
GSCNN	59.1

From table 4.1 it is established that DeepLabV3 and UNet gives the best dice score in our case. So we ensembled these two model and used it as our baseline model.

In Figure 4.1, convergence of loss and dice coefficient is shown.

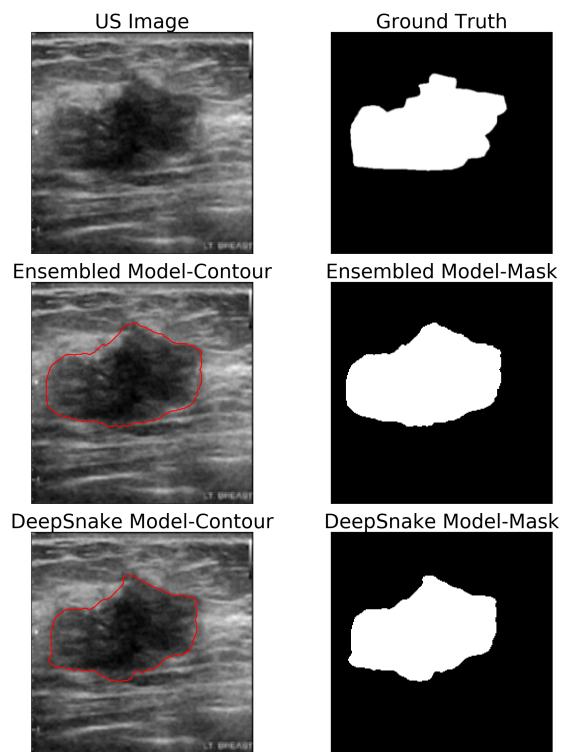
4.3 Quantitative Evaluation

In Table 4.2, we compare the performance of our proposed *DeepSnake* against the baselines in terms of mean IoU(mIoU) and mean Dice score(mDice). The numbers are reported on the testing set, and computed on the full image (no cropping). In this metric, we achieve a 6% mIoU and 2% mDice improvement on BUSI dataset, and on BUSIS dataset we got 11% mIoU and 6% mDice improvements, which is a significant result at this level of performance. We also showed Pixel accuracy(pixel) and Boundary F1(bf) score comparison in the table.

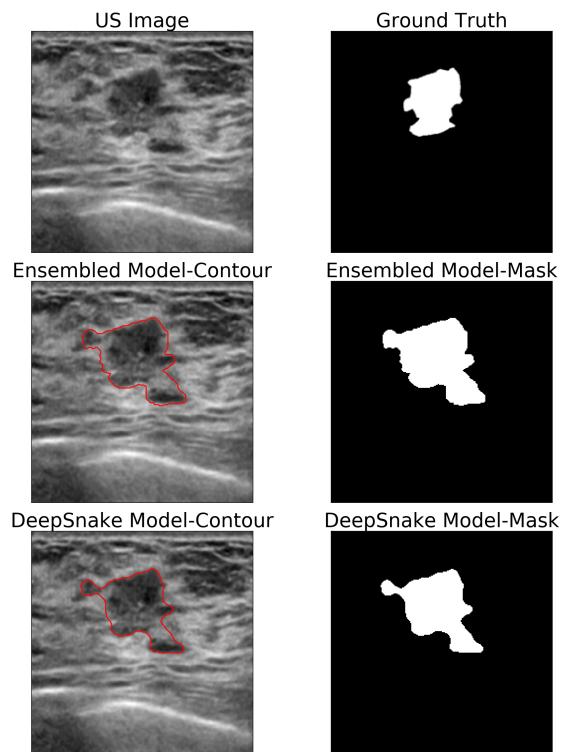
Table 4.2: Comparison of performance on both BUSI and BUSIS test set

Model Name	BUSI test set				BUSIS test set			
	mDice	mIoU	pixel	bf	mDice	mIoU	pixel	bf
SegNet[Badrinarayanan et al.]	56.6	10.2	68.3	22.1	66.9	13.1	66.7	8.7
UNet[Ronneberger et al.]	62.7	11.1	67.1	30.2	66.8	12.1	65.1	9.1
DeepLabV3[Chen et al.]	60.1	9.8	72.4	26.9	64.3	12.4	68.3	5.4
Ensembled model	61.7	46.6	67.1	28.6	68.0	51.8	70.3	7.5
Ensembled model with MSM	63.5	53.1	68.1	22.3	72.7	63.3	75.3	9.2

In following pictures, performance comparison of our proposed method is shown. More is added in Appendix A.



(a) Dice Coeff improved by 8%



(b) Dice Coeff improved by 15%

Figure 4.2: Performance Comparison

Chapter 5

Conclusion

We developed a *novel* model *DeepSnake* which can give a better contour based segmentation result. This work can be used to further bettering the result.

5.1 Concluding Remarks

We developed a DeepSnake model for better segmentation.

- We analysed state of the art models' result analysis.
- We cascaded Morphological Snake Model with the baseline pre-trained CNN-based model for better contour enriched segmentation.
- We applied ensembled CNN models as baseline model. So that it takes more spatial information to morphological snake model.
- We compared the state of the art models and proposed method's result. Section 4.3 shows the final result comparison.

5.2 Limitations and Future Directions

Better contour enriched segmentation can be done by using our proposed method. But there is some limitations in which our method did not score as high as others. So some future work can be done regarding this limitation.

- BUS image with high shadow got our proposed method confused sometimes(e.g. A.13). So we can use another model or model stream to locate shadow and infuse it to our method. Meng *et al.* [45] uses GAN to automatic detect shadow from US image.

- Morphological Snake Model's hyper-parameter can be made learnable.
- This proposed method can be used in other medical image, like CT, micro CT, MRI etc.

We hope that, our work will motivate and assist more analysts in the field of medical image segmentation.

References

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2015,” *CA: a cancer journal for clinicians*, vol. 65, no. 1, pp. 5–29, 2015.
- [2] M. S. Hossain, S. Ferdous, and H. E. Karim-Kos, “Breast cancer in south asia: a bangladeshi perspective,” *Cancer epidemiology*, vol. 38, no. 5, pp. 465–470, 2014.
- [3] “Ai applications in breast imaging.” <https://healthmanagement.org/c/hospital/issuearticle/ai-applications-in-breast-imaging>. Issue: 2019.
- [4] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.
- [5] “Teleradiology solutions.” <https://telradsol.com/>. Last Accessed: 2020.
- [6] M. Xian, Y. Zhang, H. Cheng, F. Xu, K. Huang, B. Zhang, J. Ding, C. Ning, and Y. Wang, “A benchmark for breast ultrasound image segmentation (busis) 2018,” 1801.
- [7] W. Al-Dhabayani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” *Data in brief*, vol. 28, p. 104863, 2020.
- [8] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.
- [10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results.” <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>.
- [11] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *European conference on computer vision*, pp. 740–755, Springer, 2014.

- [12] J. Beutel, H. L. Kundel, and R. L. Van Metter, *Handbook of medical imaging*, vol. 1. Spie Press, 2000.
- [13] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” *arXiv preprint arXiv:1511.07122*, 2015.
- [14] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, 2017.
- [15] R. Gadde, V. Jampani, M. Kiefel, D. Kappler, and P. V. Gehler, “Superpixel convolutional networks using bilateral inceptions,” in *European Conference on Computer Vision*, pp. 597–613, Springer, 2016.
- [16] S. Choi and C. Kim, “Automatic initialization active contour model for the segmentation of the chest wall on chest ct,” *Healthcare informatics research*, vol. 16, no. 1, pp. 36–45, 2010.
- [17] P. Marquez-Neila, L. Baumela, and L. Alvarez, “A morphological approach to curvature-based evolution of curves and surfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 2–17, 2013.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [19] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
- [22] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, “Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation,” *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.
- [23] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected crfs,” *arXiv preprint arXiv:1412.7062*, 2014.

- [24] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [25] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [26] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.
- [27] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *International conference on medical image computing and computer-assisted intervention*, pp. 424–432, Springer, 2016.
- [28] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation,” *arXiv preprint arXiv:1802.06955*, 2018.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [30] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells III, F. A. Jolesz, and R. Kikinis, “Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports,” *Academic radiology*, vol. 11, no. 2, pp. 178–189, 2004.
- [31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.
- [32] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, and Y. Zheng, “Learning active contour models for medical image segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11632–11640, 2019.
- [33] T. F. Chan and L. A. Vese, “Active contours without edges,” *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [34] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, “Gated-scnn: Gated shape cnns for semantic segmentation,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5229–5238, 2019.

- [35] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [36] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, “A real-time algorithm for signal analysis with the help of the wavelet transform,” in *Wavelets*, pp. 286–297, Springer, 1990.
- [37] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella, and J. Schmidhuber, “Fast image scanning with deep max-pooling convolutional neural networks,” in *2013 IEEE International Conference on Image Processing*, pp. 4034–4038, IEEE, 2013.
- [38] W. Liu, A. Rabinovich, and A. C. Berg, “Parsenet: Looking wider to see better,” *arXiv preprint arXiv:1506.04579*, 2015.
- [39] P. Baldi, “Autoencoders, unsupervised learning, and deep architectures,” in *Proceedings of ICML workshop on unsupervised and transfer learning*, pp. 37–49, 2012.
- [40] C. Gulcehre, M. Moczulski, M. Denil, and Y. Bengio, “Noisy activation functions,” in *International conference on machine learning*, pp. 3059–3068, 2016.
- [41] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” *International journal of computer vision*, vol. 22, no. 1, pp. 61–79, 1997.
- [42] “Basis open-sourced dataset.” <http://www.onlinemedicalimages.com/index.php/en/>. Last Edited: 2020.
- [43] D. G. Chidders, D. P. Skinner, and R. C. Kemerait, “The cepstrum: A guide to processing,” *Proceedings of the IEEE*, vol. 65, no. 10, pp. 1428–1443, 1977.
- [44] “Sørensen–dice coefficient.” https://en.wikipedia.org/wiki/S\OT1\orensen\OT1\textendashDice_coefficient. Last Edited: 2020.
- [45] Q. Meng, C. Baumgartner, M. Sinclair, J. Housden, M. Rajchl, A. Gomez, B. Hou, N. Toussaint, J. Tan, J. Matthew, *et al.*, “Automatic shadow detection in 2d ultrasound,” 2018.

Appendix A

Performance Visualization

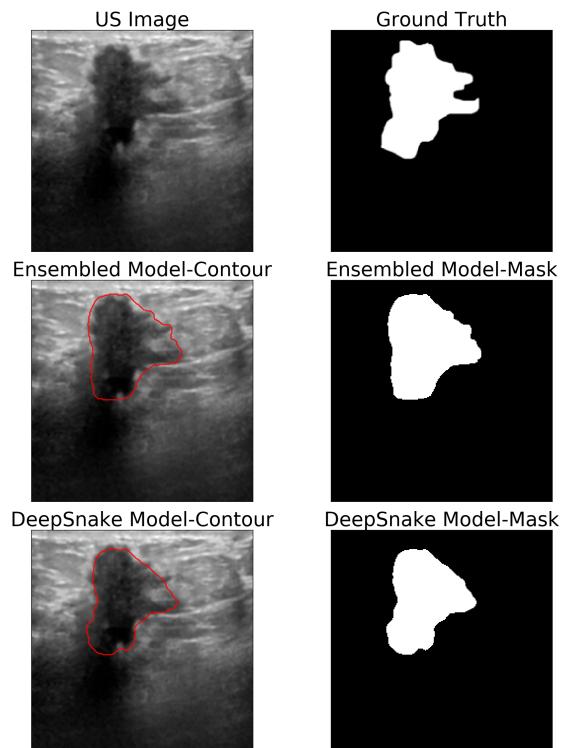


Figure A.1: Test Image 1

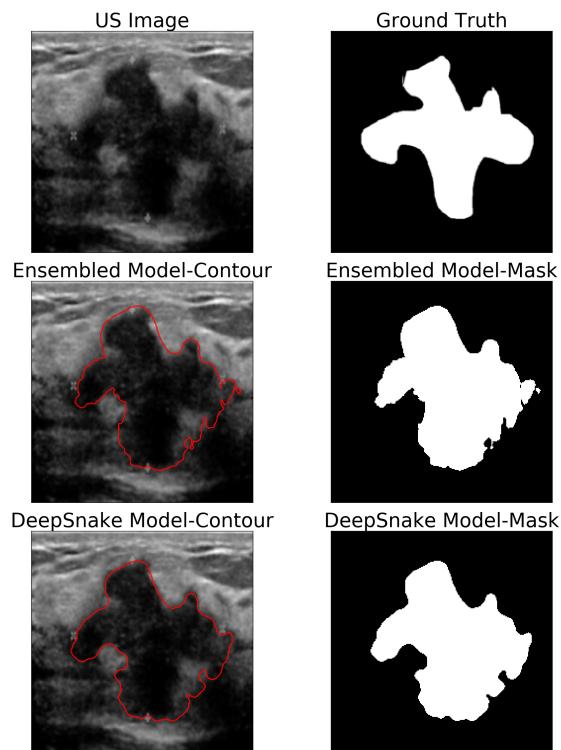


Figure A.2: Test Image 2

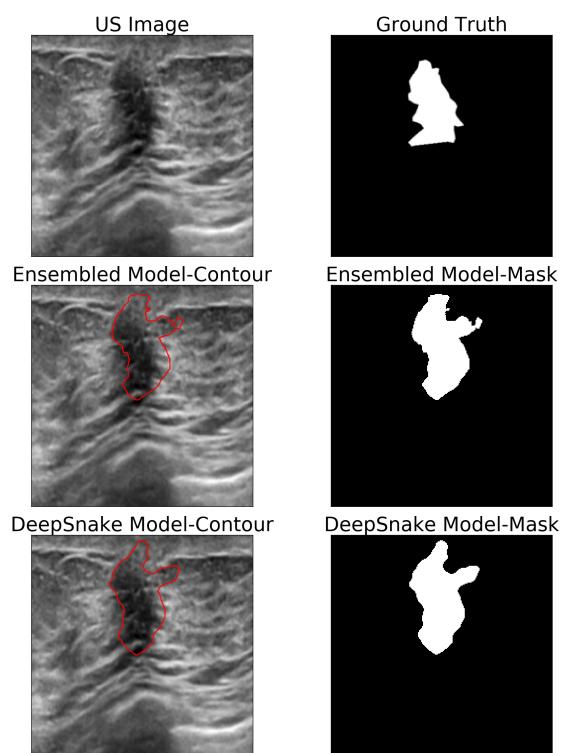


Figure A.3: Test Image 3

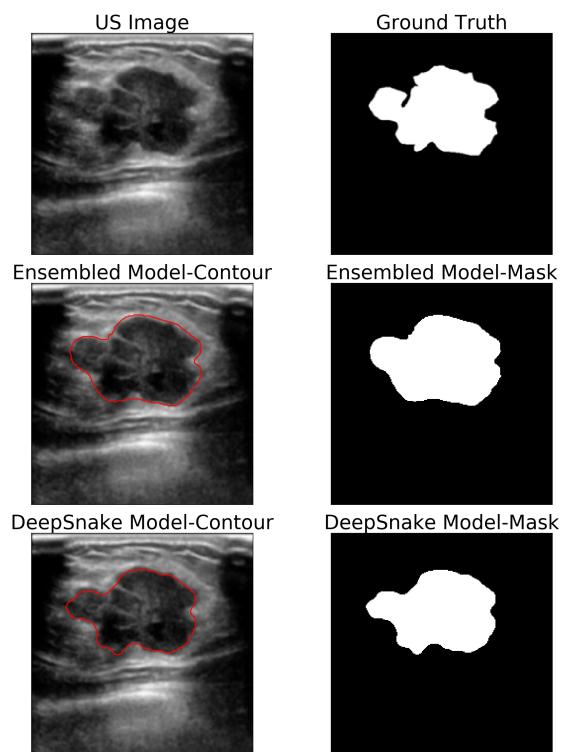


Figure A.4: Test Image 4

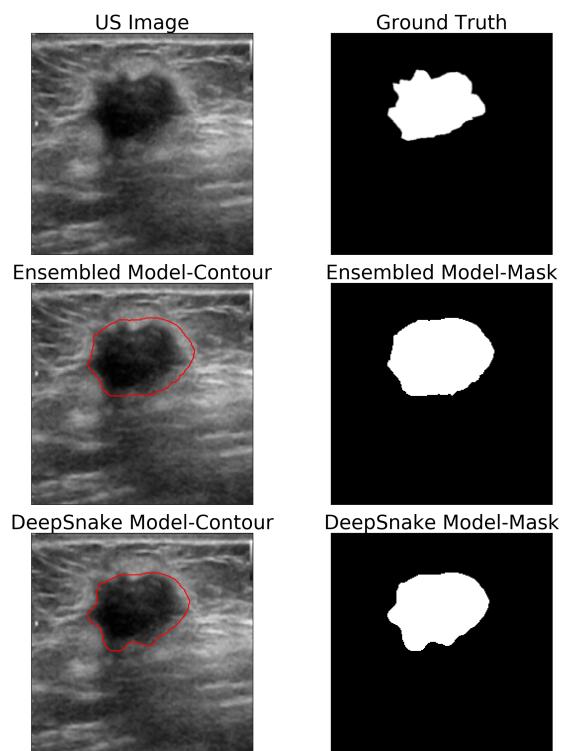


Figure A.5: Test Image 5

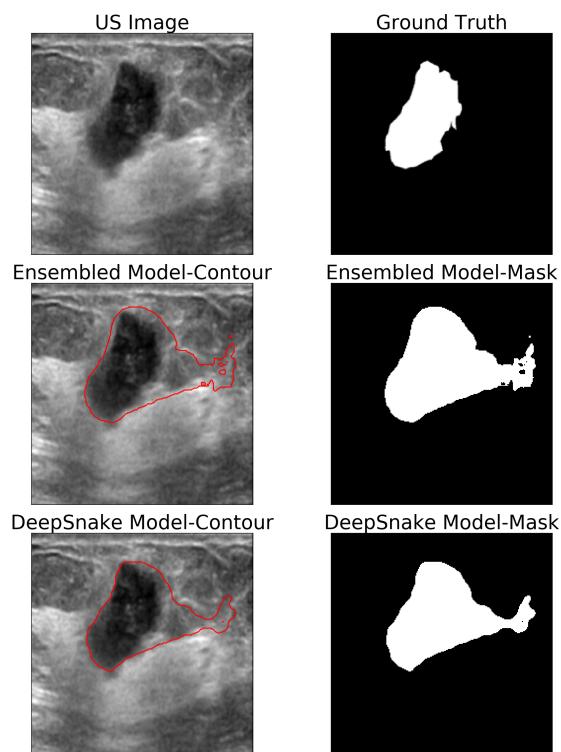


Figure A.6: Test Image 6

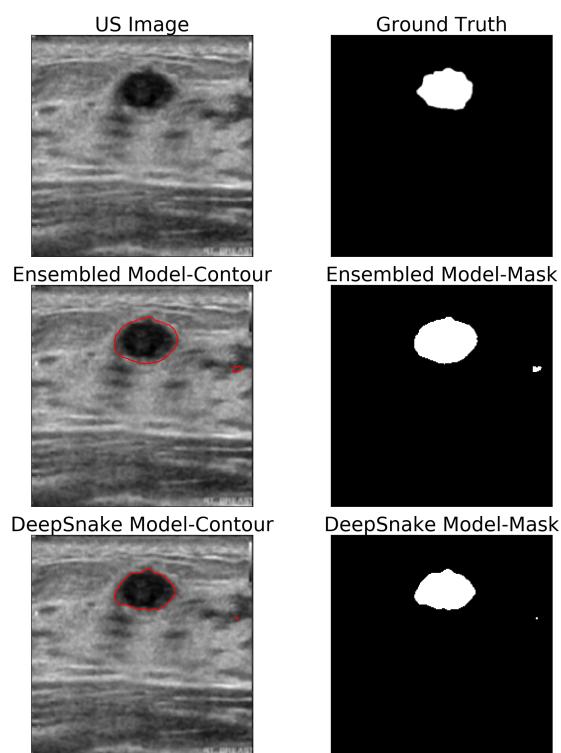


Figure A.7: Test Image 7

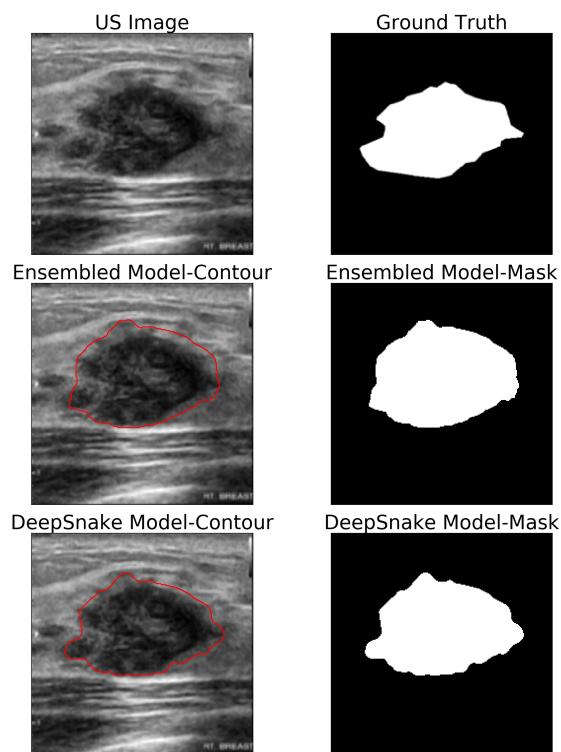


Figure A.8: Test Image 8

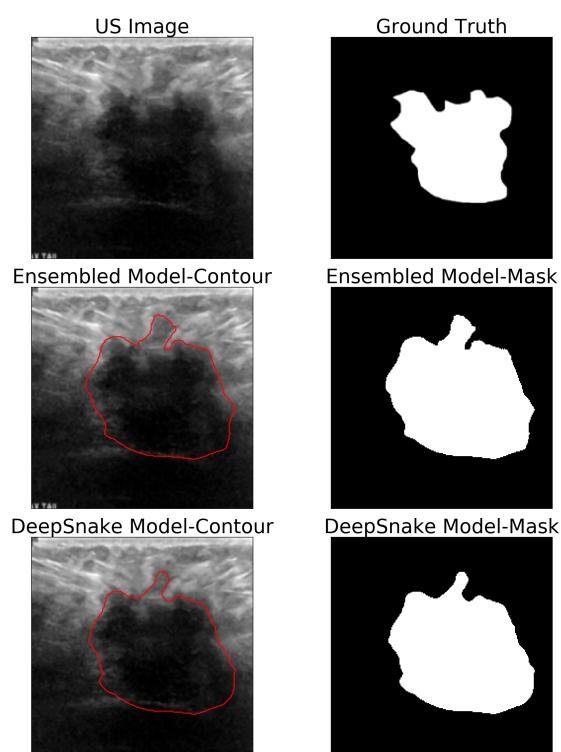


Figure A.9: Test Image 9

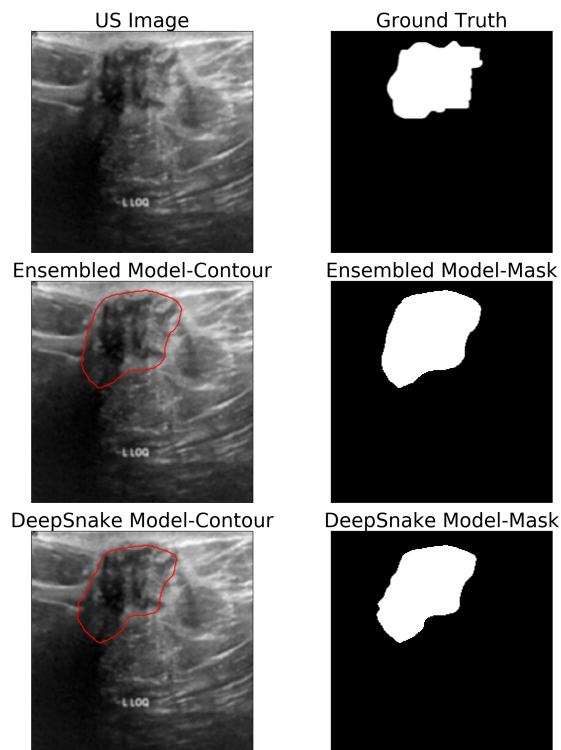


Figure A.10: Test Image 10

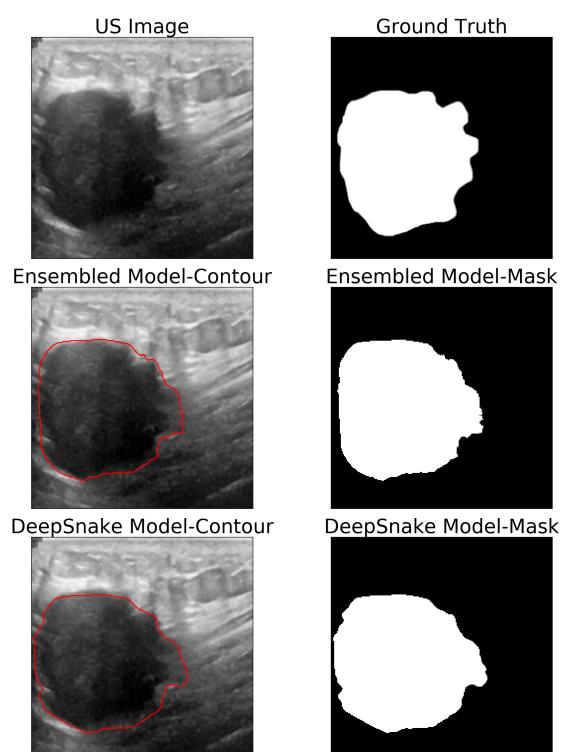


Figure A.11: Test Image 11

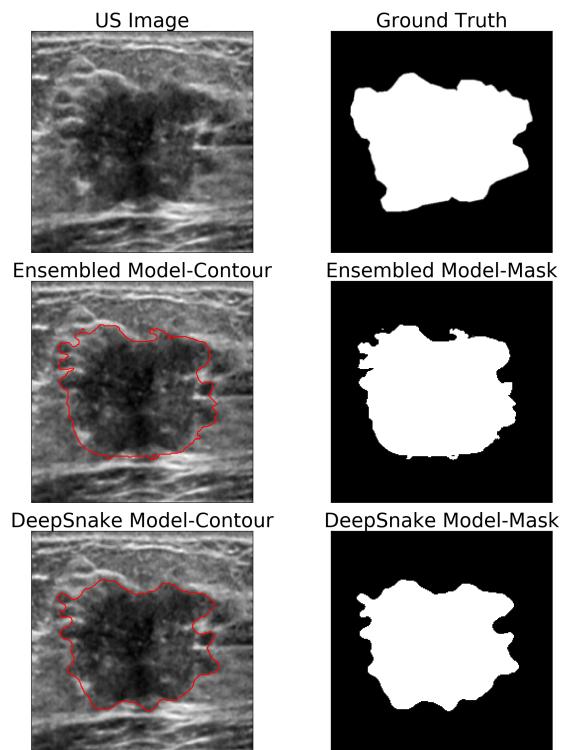


Figure A.12: Test Image 12

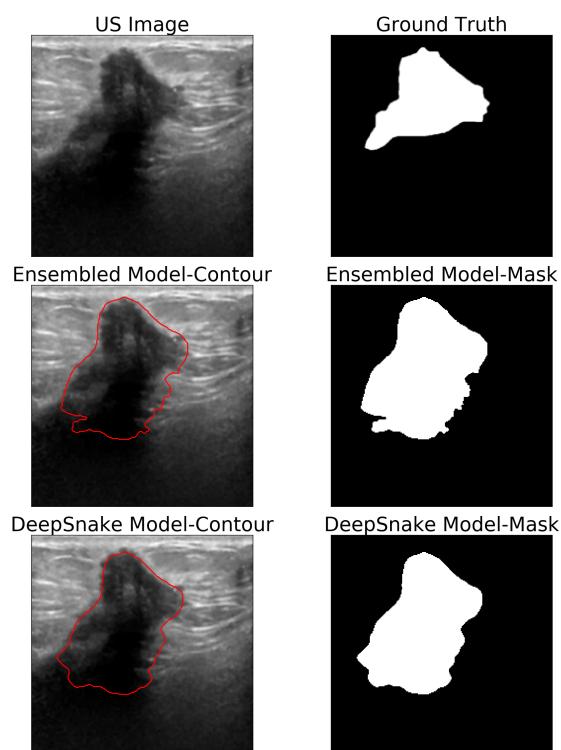


Figure A.13: Test Image 13

Generated using Undegraduate Thesis L^AT_EX Template, Version 1.4. Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

This thesis was generated on Saturday 15th February, 2020 at 4:18am.