Active Learning for Biomedical Image Segmentation

Keshav Jethaliya
The State University of New York at Buffalo
Buffalo, New York
keshavje@buffalo.edu

Abstract

Image Segmentation has been a fundamental problem in the field of Medicine. From detecting skin cancer to brain tumor, everything needs the segmentation of the MRI images. Though, the recent advances in deep learning have shown promising results for image segmentation, all those methods highly rely on a large number of training data, and in the case of biomedical images, a professional is needed to label these images, which can cost a lot of money and time. So, we propose an Active Learning approach, which can train a model with a limited number of labeled data, saving the cost and invaluable time of the professional. The approach has been applied to two different datasets, "Skin Lesion Analysis Toward Melanoma Detection" dataset from International Skin Imaging Collaboration (ISIC) and "Lower Grade Glioma Detection" dataset from The Cancer Imaging Archive (TCIA).

1. Introduction

One of the major challenges in biomedical imaging is the dependence on a specialist, who is quite knowledgeable about the problem statement. Such a specialist is hard to find, and if you can find one, it will be really expensive in terms of money as well as time. With active learning, we help a model become less dependent on, a human annotator or an Oracle, for labelling purposes. Many studies related to active learning have been around in the recent years, most of them select a number of informative samples from the data, based on the uncertainty of the model predictions, to be labelled by an Oracle. In this paper, we implement a more advance and more cost-effective solution than traditional active learning methods. The Cost-Effective Active Learning[10, 4, 9, 3], also selects the most informative samples to be labeled by an Oracle, but also selects a number of high confidence samples for pseudo-labelling. This helps the model to learn faster with even less data.

1.1. Monte Carlo Dropout

According to Yarin Gal *et al*. [2], the dropout regularization techniques for neural networks can be interpreted as a Bayesian approximation to Gaussian processes[7]. We implement Monte Carlo dropout in our model, which, in addition to applying dropout whiling training (normal dropout technique), it is also applied while testing. This helps in uncertainty estimation of the predictions.

1.2. U-Net Architecture

The baseline architecture used here is one of the most popular CNN model for image segmentation: U-Net[8]. U-Net was created to solve the image segmentation task for biomedical images, as the paper states[8]. The network first makes a CNN with downsampling layers and combines it with upsampling layers while providing the output of downsampling layers to the input of upsampling layers in order to obtain desired results. Our model consists of 5 pairs of downsampling convolutions (3×3). After each pair a max-pooling layer is applied and between the convolution layers, dropout is applied. The upsampling part consists of 4 pairs of convolutions (3×3) with Conv2DTranspose layer after each pair. Dropout is applied between the convolution layers. The network is 20 convolutional layers deep and has around 31 million trainable parameters.

2. Dataset Definition

There are two datasets used in this paper: "Skin Lesion Analysis Toward Melanoma Detection" dataset from International Skin Imaging Collaboration (ISIC) [1] and "Lower Grade Glioma Detection" dataset from The Cancer Imaging Archive (TCIA)[6, 5].

1. Skin Lesion Analysis Toward Melanoma Detection dataset consists of 2000 lesion images in JPEG format and 2000 binary mask images in PNG format for training and 600 lesion images in JPEG format and 600 binary mask images in PNG format for testing.(Table 1)

2. Lower Grade Glioma Detection dataset consists of

Data	Samples
Train	1600
Validation	400
Test	600

Initial Data	Samples
Labeled	600
Unlabeled	1000

Table 1. ISIC Dataset

Data	Samples
Train	878
Validation	220
Test	275

Initial Data	Samples
Labeled	329
Unlabeled	549

Table 2. TCIA Dataset

1373 Brain MRI images in tif format and 1373 binary masks in tif format. 20% of this data is kept for testing.(Table 2)

3. Task Statement

Our model aims to provide an Active Learning approach, which can help reduce the effort of labeling data in case of image segmentation. The task of manually annotating image masks is very time consuming and requires a lot of effort. In case of Biomedical images, it also takes up a lot of money. Our model helps annotate the data and get trained on very little labeled data. It does not mask the images as detailed as a human annotator, but still enough to obtain desired information from it. It explores the possibility of pixel-wise uncertainty using Monte Carlo Dropout at the testing phase.

4. Methodology

Initially we divide our training data into two sets: labeled set(D^L), which is only 37.5% of the training data and unlabeled set(D^U), which is 62.5% of training data.

There are 2 parts of this approach. First, the initial training part, which trains the baseline U-Net model on the labeled data. The second part is the Active Learning loop, which iteratively makes predictions on the unlabeled data using initially trained model, computes the uncertainty of the predictions and selects 2 kinds of samples, the ones which needs to be annotated by an oracle and the others that will be pseudo-annotated by the model. The selected samples are then added to the labeled set of data and again the model is trained.(Figure 1)

4.1. Computing Uncertainty

In the Active Learning loop, the weights from the trained model are used to predict on D^U and we do this iteratively for t steps so that we have t different predictions, say P, because of Monte Carlo Dropout[2]. Now we calculate the variance on each pixel for these t predictions, which gives us the **Uncertainty Map**(Figure 2). Using the uncertainty map, we divide the predictions, to be manually annotated and that to be pseudo-annotated. Now, to compare the uncertainty we need a numerical value, so we add all the pixel level variances and use that value for comparing uncertainties.

4.2. Selecting Samples

The samples are selected based on the uncertainty score as shown in the graph. These 4 types of samples are selected based on the graphs as in Figures 5 and 6:

1. The samples which have a very low uncertainty score is because they are mostly missed by the model and left undetected. (10 samples)
2. The samples which have a high uncertainty score and a very low confidence. (10 samples)
3. These samples have a low uncertainty score but also have a low dice coefficient, which reduces the efficiency of model. They get mixed up with the highly confident samples selected for pseudo labeling, So, some of these samples are selected randomly to be labeled by oracle. (15 samples)
4. These samples are highly confident, which are selected for pseudo labeling. ($20 \times$ *number of iterations* samples)

The Uncertainty vs Dice Coefficient (Figure 5) is just to understand, but actually we make use of the Uncertainty histogram (Figure 6) to distinguish labels, as in real world, the ground truth of the unlabeled data is not available.

4.3. Labeling selected samples

The selected sample are then either labeled by the oracle (1,2,3) or predicted by the model (4). These samples are removed from the Unlabeled pool (D^U) and added to the labeled pool (D^L). The model is again trained on D^L and the process repeats for the specified number of iterations.

5. Results

We trained the model on two different datasets, ISIC Skin lesion towards Melanoma Detection dataset and TCIA Lower Grade Glioma detection database. The metrics used to asses the quality of predictions was Dice Coefficient.

$$DICE(A,B) = \frac{2|A\hat{B}|}{|A| + |B|}$$

We obtain a Dice coefficient 77% while testing on Melanoma Segmentation dataset and A Dice Coefficient of 58% while testing on LGG Segmentation dataset.

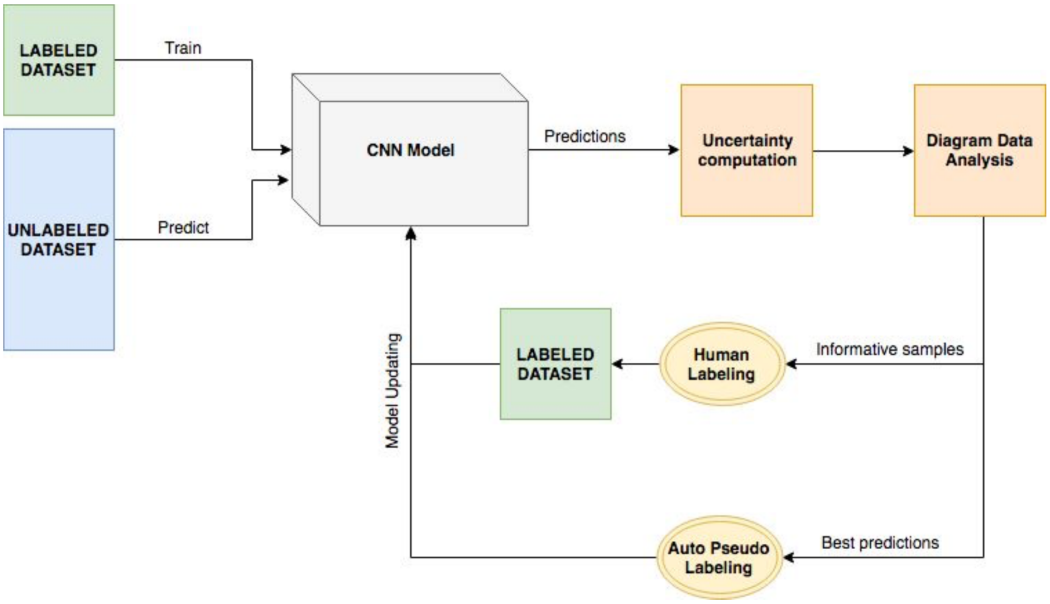

Figure 1. Active Learning Methodology.

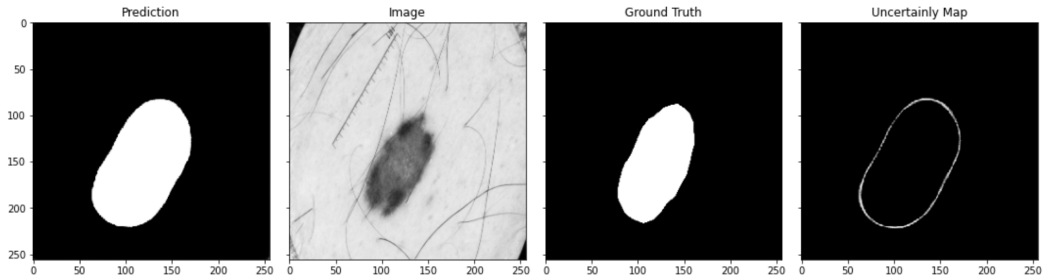

Figure 2. Prediction of a Brain MRI on LGG Segmentation dataet by TCIA.

As we can see from the graphs(Figures 7, 8, 9 and 10), the model learns and achieves 90% of its accuracy in the initial training phase and improves by upto 10% in the active learning iterations. The model trains for 20 epochs in initial learning and for 5 epochs in each iteration of active learning. It took almost 3 hours to train the model with Melanoma segmentation dataset[1] and almost 2 hours to train LGG segmentation dataset[6, 5].

6. Conclusion

From above results, we can conclude that Active Learning approach can produce results, almost as good as some model trained completely on labeled data, while training on only 50% of labeled data. It. is therefore a cost effective solution for labeling and segmenting the biomedical images.

References

[1] Celebi ME Helba B Marchetti MA Dusza S Kalloo A Liopyris K Mishra N Kittler H Halpern A Codella N, Gutman D. ”skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)”. 1, 3

[2] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2016. 1, 2

[3] Marc Gorriz, Axel Carlier, Emmanuel Faure, and Xavier Giró-i Nieto. Cost-effective active learning for melanoma segmentation. 11 2017. 1

[4] Y. Huang, Z. Liu, M. Jiang, X. Yu, and X. Ding. Cost-effective vehicle type recognition in surveillance images with deep active learning and web data. *IEEE Transactions on Intelligent Transportation Systems*, 21(1):79–86, 2020. 1

[5] Nicholas M. Czarnek Parisa Shamsesfandabadi Katherine B. Peters Ashirbani Saha Maciej A. Mazurowski, Kal Clark. ”radiogenomics of lower-grade glioma: algorithmically-

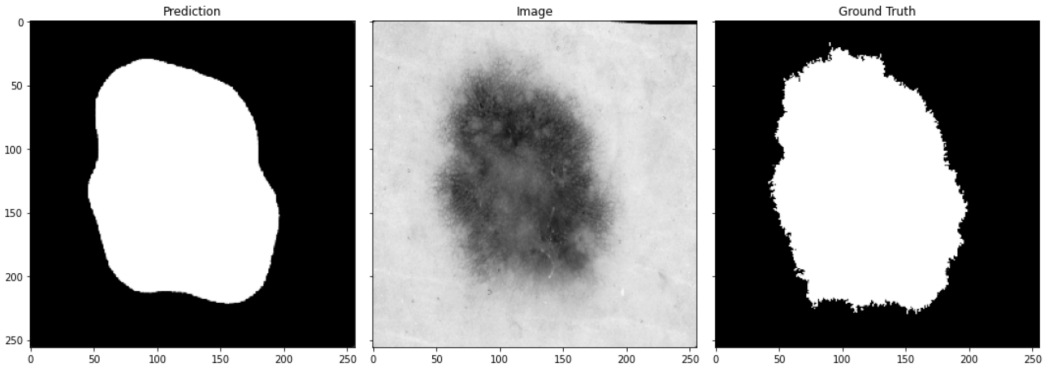

Figure 3. Prediction of a Skin Lesion on Melanoma Segmentation dataet by ISIC.

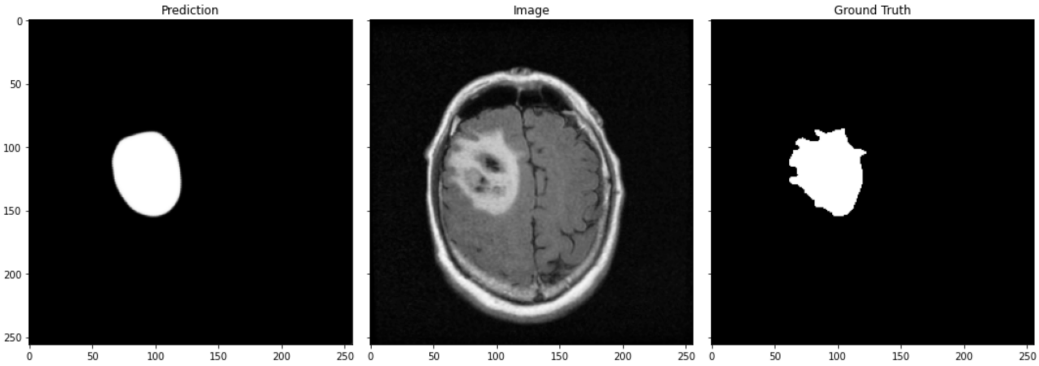

Figure 4. Prediction of a Brain MRI on LGG Segmentation dataet by TCIA.

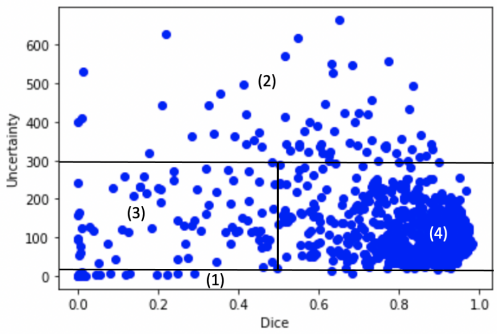

Figure 5. A graph Uncertainty vs Dice Coefficient of the predictions.

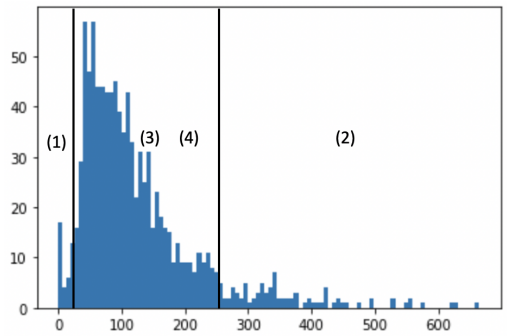

Figure 6. A histogram of Uncertainty.

assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with the cancer genome atlas data."., 2017. 1, 3

[6] Maciej A. Mazurowski Mateusz Buda, AshirbaniSaha. "association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm." computers in biology and medicine., 2019. 1, 3

[7] C. Rasmussen and C. K. Williams. Gaussian processes for machine learning. In *Adaptive computation and machine learning*, 2009. 1

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. 1

[9] D. Wang and Y. Shang. A new active labeling method for deep learning. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 112–119, 2014. 1

[10] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin. Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology*,

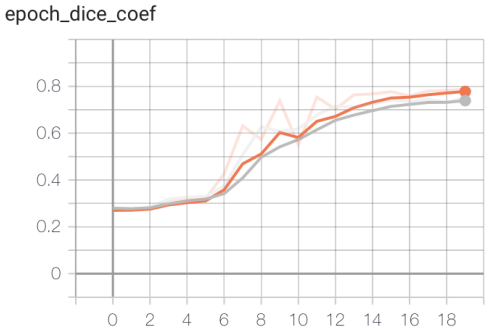

Figure 7. Training and Validation Dice Coefficient in initial learning phase for Melanoma Segmentation.

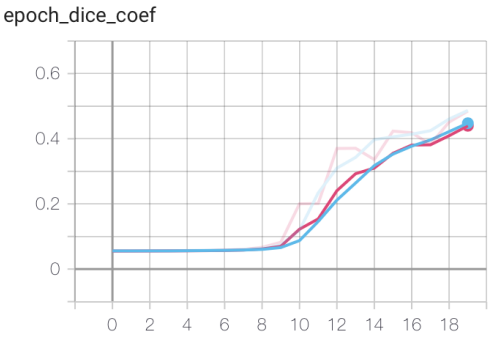

Figure 8. Training and Validation Dice Coefficient initial learning phase for LGG Segmentation.

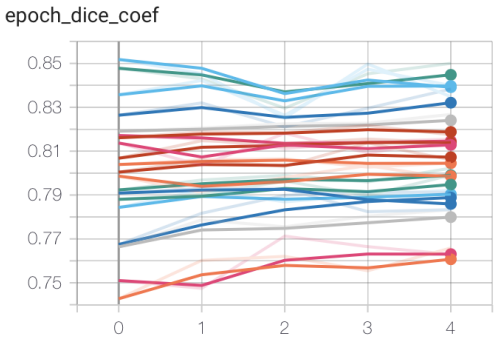

Figure 9. Training and Validation Dice Coefficient in active learning phase for Melanoma Segmentation.

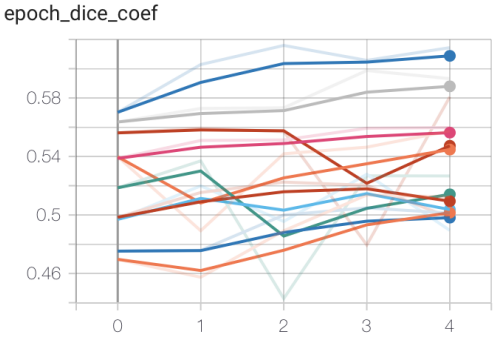

Figure 10. Training and Validation Dice Coefficient active learning phase for LGG Segmentation.

27(12):2591–2600, 2017. 1