U-Net for Brain Tumor CT Semantic Segmentation

Weiyang Chen
Department of Computer Science and Engineering
University at Buffalo
Buffalo, USA
wchen38@buffalo.edu

Kun Lin
Department of Computer Science and Engineering
University at Buffalo
Buffalo, USA
kunlin@buffalo.edu

Wejin Zhu
Department of Computer Science and Engineering
University at Buffalo
Buffalo, USA
weijinzh@buffalo.edu

Zhi Wen Huang
Department of Computer Science and Engineering
University at Buffalo
Buffalo, USA
zhiwenhu@buffalo.edu

***Abstract*—With the emerging field of data science, many of the medical data are now stored in a convenient way for computing applications such as image classification,semantic segmentation, and many machine learning applications. This paper mainly focus on efficient methods of performing semantic segmentation along with UNet model and real world medical data. The goal is surpass the result in the paper [1], and analyze the advantages and disadvantages between two models by tuning hyper-parameters and adding more machine learning techniques such as data argumentation methods.**

1. Introduction

Semantic segmentation [2] [3] is important task in fields of computer vision and image process. We believe we could apply such technique on segmenting and diagnosing brain tumors. Early diagnosis could potentially save patient's life with earlier treatment planning.
The goal of this paper focus on application of U-Net neural network model on set of FLAIR images [4] from The Cancer imaging Archive where these type of image removes signal from the cerebrosphinal fluid in the resulting images. U-net is a type of convolutional neural network that performs well on biomedical image segmentation.

2. Motivation and Related Work

Our project's idea came from a paper called "Semantic Segmentation with Histological Image Data: Cancer Cell vs. Stroma" by Adam Abdulhamid from Stanford University.The main focus of this paper is to segment images into cancerous sections and non-cancerous sections. The author also categorized on the classification behavior on different types of cancers. The model of the project is a Convolution Model called VGG16. VGG16 trains the data and applied different kernel sizes and analyzed the results. This technique also can be used for life expectancy prediction for further research in the medical field. Our interest is to improve this semantic segmentation for brain cancer using a different model.

3. Data

A. Source

The dataset we are using are downloading from Kaggle with link https://www.kaggle.com/mateuszbuda/lgg-mri-segmentation/version.

B. Data-Set Summary

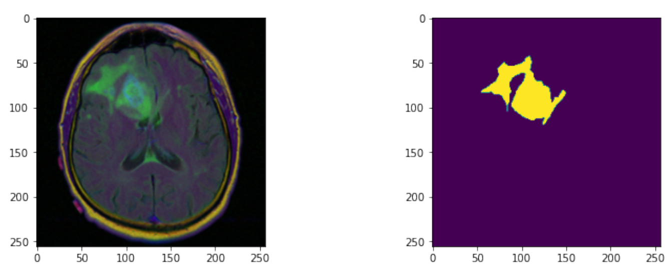

Fig. 1. Left is original Brain MR image and right is its mask

The dataset contains 3929 Brain MR images together with manual FLAIR abnormality segmentation masks. And also some of the Brain MR images contains brain tumor and some others not. Each Brain MR image in the data-set is a RGB image with size (256*256) and each mask is a boolean matrix with size (256*256) which means each index in the matrix is 1 or 0. As the picture shows in the Fig.1, Left one is a original Brain MR image and right one is its mask that the yellow part is the Brain Tumor and it will be 1 in the matrix.

C. Data Pre-Processing

Trans RGB image into Grey image which has only one color channel between 0 to 255. And divide each pixel value by

255 to get a new image which is float type. And respectively reshape the list of image and mask into (# of images, 256, 256, 1). The first dimension is to help image to find its own mask in the new shape of mask list.

D. Data Augmentation

Neural network require a large amount of training images and masks in order to improve its accuracy and avoiding over-fitting. Our data is correspond to 110 patients with 3929 image and labels. One method to accommodate such issue is apply data augmentation image processing technique, where transformation are perform on existing training set to produce more training images and labels. A detailed data augmentation methods is listed in Table 1.

TABLE I
APPLIED DATA AUGMENTATION METHODS AND PARAMETERS

Methods	**Parameters**
Flip horizontally	True (random flip)
Flip vertically	True (random flip)
Rotation	$\pm 45°$
Shift	10% on horizontal and vertical direction
Shear	10% shear
Zoom	$\pm 10\%$

4. APPROACH

A. Overview

We are mainly focus on the Unet network. The detailed architecture of the network is described in the Architecture section. Follow by its architecture, we performed some experiments with the network such as combining Unet with some other deep learning model, tuning the optimizer, loss function and adjusting the number of layers in the network.

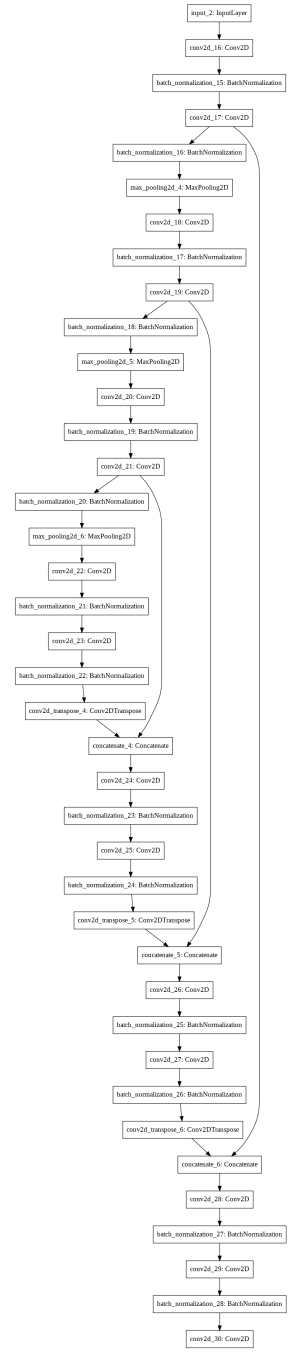

Fig. 2. Unet Implementation

B. Architecture

The overview of the Unet is shown in Figure 3, and our implementation of the model is shown in the Figure 2. [5] If we split the architecture into two chunks, with left side to be the contracting path and right side to be the expansive path. Every time the contracting path would double the number of features and the process is called downsampling. Follow by the downsampling, the upsampling is the expansive path and would halve the number of features.

In the process of downsampling, two 3x3 convolutions are repeated applied that followed by a rectified linear unit (ReLu) activation function and a max pooling with kernel size of 2x2 and stride of 2. In the process of upsampling, we also perform the same procedure as what we did in the contracting path which there are two 3x3 convolutions are repeated applied following by a ReLu activation function. In addition to it, we also concatenate the layers in the expansive path with the corresponding layers in the contracting path. At the final layer a 1x1 convolution is used to map into desired class.

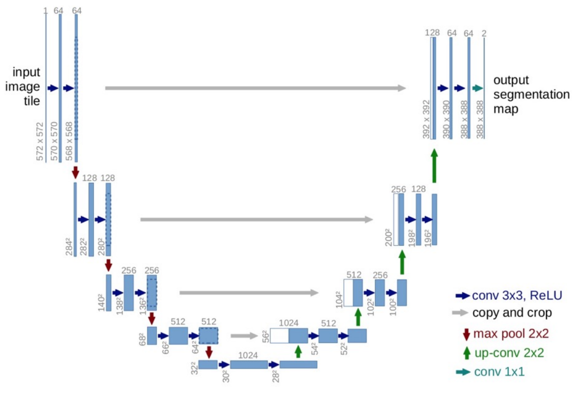

Fig. 3. Basic Unet Architecture

1) Experiment 1: The neural network we build is same as Fig.2, and we apply **Dice loss** as its loss function when we training the samples. **Dice coefficient** is a set similarity measure function, usually used to calculate the similarity of two samples, the value range is [0,1]. Dice Loss is $1 - Dicecoefficient$ so if two samples more similarity, the DL will be smaller.

$$Dice_Loss(A, B) = 1 - \frac{2|A \cap B| + Smooth}{|A| + |B| + Smooth} \quad (1)$$

2) Experiment 2: Keeping neural network the same, and applying **binary cross-entropy** [6] as loss function when we training the samples. Because the the result we want to get is semantic segmentation that shows tumor shape, and binary cross-entropy and apply to two classes classification, so it also is suitable with this problem that to determine whether the tumor is in the certain pixels.

$$BCE_Loss = \sum_x -(T_x \log(P_x) + (1 - T_x) \log(1 - P_x)) \quad (2)$$

The upper formula, T_x is mask and P_x is predicted tumor segmentation.

3) Experiment 3: Keeping neural network the same, and applying **IOU loss** as loss function when we training the samples. IOU loss is similar to the Dice loss that belong to the measurement method of metric learning [7] which metric learning is the task of learning a distance function over objects.

$$IOU_Loss(A, B) = 1 - \frac{A \cap B + Smooth}{A \cup B + Smooth} \quad (3)$$

Smooth is value we defined, usually is 1. A is the model predicted value and B is the true value from mask. IOU is 1 - Intersection between A and B plus smooth to divide by Union between A and B plus smooth.

4) Experiment 4: Keeping neural network the same, and applying **Focal Tversky loss** [8] as loss function when we training the samples. Focal Tversky loss more like the combination of Dice loss and IOU loss.

$$FTL(A, B) = \frac{A \cap B}{A \cap B + \alpha|A - B| + \beta|B - A|} \quad (4)$$

For Focal Tversky loss, we can change the parameter of alpha and beta to get different result. A is the model predicted value and B is the true value from mask. $|A - B|$ represent **False Positive**, $|B - A|$ represent **False Negative**.

5) Experiment 5: Similar to the experiment 1 that using dice loss, but now we will change model that changing the MaxPooling into AveragePooling in first two convolution layer that we reduce the size of the input image.

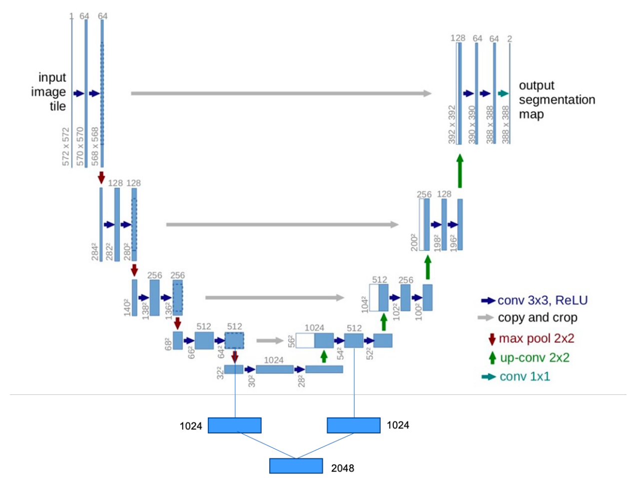

Fig. 4. Modified Unet Architecture

6) Experiment 6: Instead of tuning the loss function to observe if the result can be improved. In this experiment we want to make some changes to the internal structure of the network. As previous Experiment, binary-cross-entropy has best performance so we will use BCE as loss function in this experiment. In the Figure 4, we added one more downsampling block and upsampling block to the original Unet architecture. We hope the modified Unet model can capture more import features than the original model since it has more layers to feature caption and as the result our model can have better segmentation ability.

C. Mean Intersection over Union

We will use the Miou [9] to determine the model performance in the semantic segmentation.

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \quad (5)$$

It calculates the IOU for each category (the intersection of the true label and the prediction result) separately, and then averages the IOUs of all categories.

5. EXPERIMENTAL RESULTS

We will randomly shuffle the training and testing data-set 6 times, and to calculate the Miou score for each model each times to help us to compare models' performance.

A. *Experiment 1 Result*

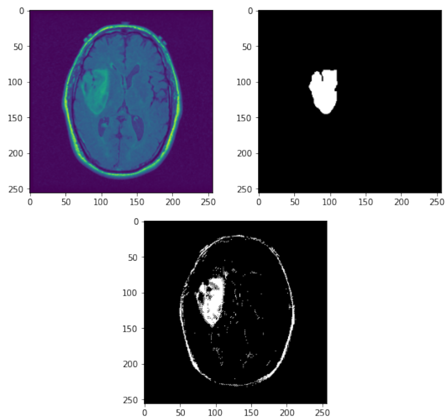

Fig. 5. First Model Original vs Predict

As see the top left is the original image in the test data-set that model never seen before, top right is its mask, the bottom image is the result model predict.The tumor part is roughly segmented, but the pixel value of skull outline in the original picture is big, so the skull outline will also be segmented in the prediction results.

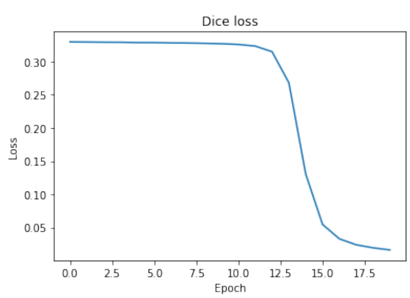

Fig. 6. Dice Loss Graph

At beginning of fewer epochs, the loss decrease a little, and after epoch 10, the loss violent drop in 6 epochs and then goes stable.

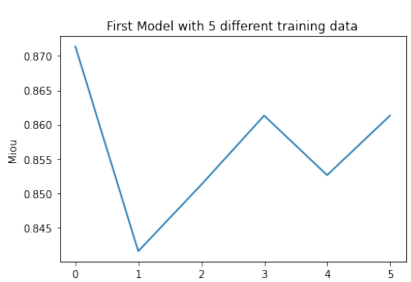

Fig. 7. Model #1 Miou Score

The Miou score of this model is between 87% and 84%, and we can know that the accuracy of model by using Dice Loss is around 85%.

B. *Experiment 2 Result*

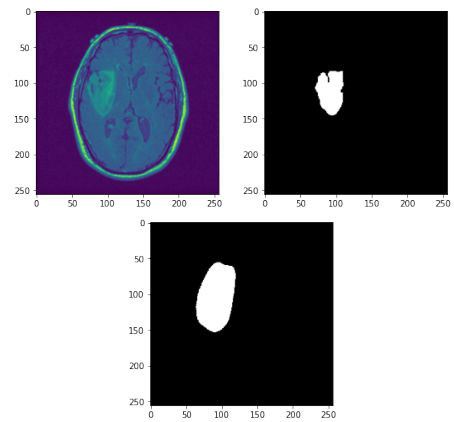

Fig. 8. Second Model Original vs Predict

As see the top left is the original image in the test data-set that model never seen before, top right is its mask, the bottom image is the result model predict. Model predicted that the location of the tumor will be larger than the original, but it succeeded in recognizing the skull outline and eliminated the skull outline during segmentation. I think the reason why the predicted brain tumor is bigger than the actual one is when the image is convoluted and reduced size, it is inevitable that the pixels around the brain tumor get the pixel values of the brain tumor part through the MaxPooling Layer. So when we enlarged the image to the original size, the tumor part also expanded.

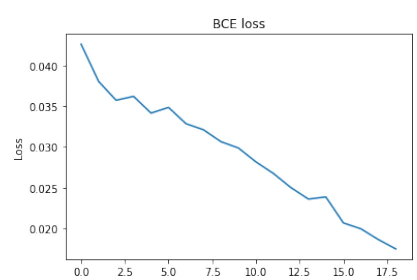

Fig. 9. BCE Loss Graph

The loss value has been showing a downward trend, indicating that the model has the potential to rise.

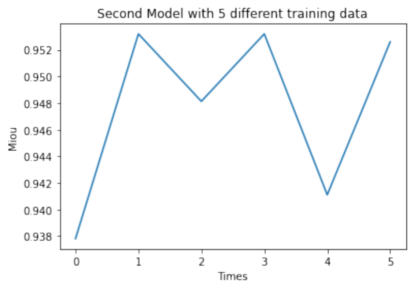

Fig. 10. Model #2 Miou Score

The Miou score of this model is between 95% and 93%, and we can know that the accuracy of model by using Dice Loss is around 94%.

C. Experiment 3 Result

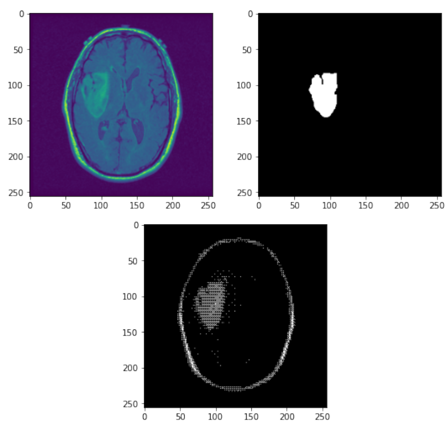

Fig. 11. Third Model Original vs Predict

The result prediction graph by third model is similar to the model that using Dice Loss, that Skull outline still exists and also segment the tumor part is scattered.

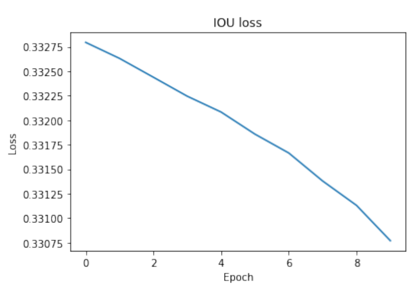

Fig. 12. IOU Loss Graph

IOU loss shows a nearly linear drop in 20 epochs, and starting loss value is pretty small, i think it is the features of our data-set that most areas in the mask are black and even some mask that without tumor is all black and it will make the loss value very small.

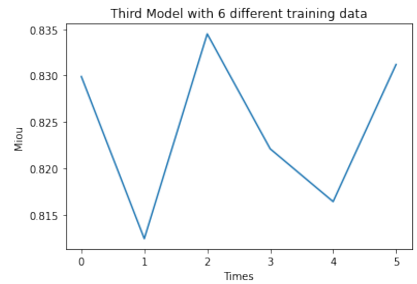

Fig. 13. Model #3 Miou Score

The Miou score of the model with Iou loss is between 81% and 83%, and we can know that the accuracy of model by using Dice Loss is around 82%.

D. Experiment 4 Result

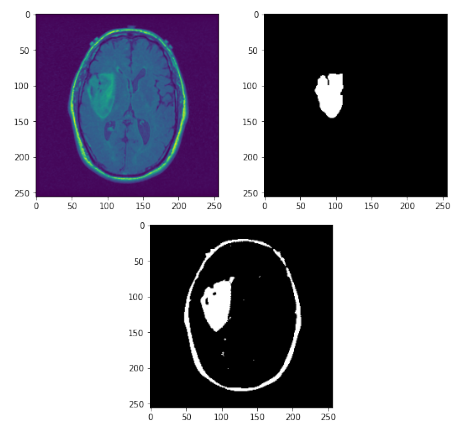

Fig. 14. Fourth Model Original vs Predict

As see the top left is the original image in the test data-set that model never seen before, top right is its mask, the bottom image is the result model predict. The predict result is similar to the first model, but the shape of the tumor is clearer and less noise point in the graph. The disadvantage is that the outline of the skull is more obvious than first model, I think the reason is Focal Tversky Loss make the model pay more attention to highlights and ignore non-prominent points, While highlighting the shape of the brain tumor, the brain outline is inevitably also highlighted.

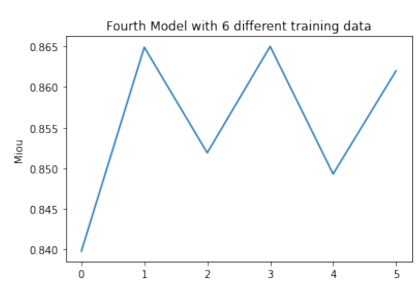

Fig. 15. Model #4 Miou Score

The Miou score of the model with Iou loss is between 84% and 0.87%, and we can know that the accuracy of model by using Dice Loss is around 85%.

E. Experiment 5 Result

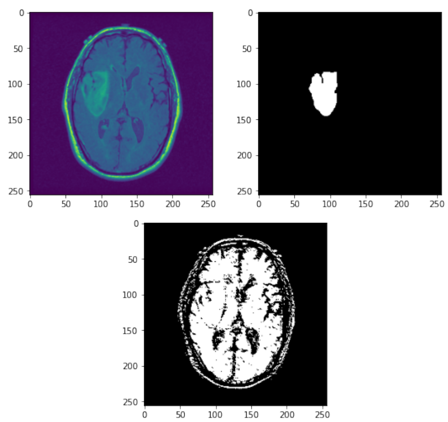

Fig. 16. Fifth Model Original vs Predict

As see the top left is the original image in the test data-set that model never seen before, top right is its mask, the bottom image is the result model predict. Model 5 is change the first model's pooling layer that from MaxPooling to AveragePooling. The result is that the model cannot segment brain tumors from the images. I think the reason is AveragePooling Smoothed out the protruding pixels of the brain tumor during the convolutional, When changing the picture back to its original size, the protruding pixels are no longer prominent. So AveragePooling is unsuitable with image segementation problem. Because this Model is unsuitable for the problem, we will not go further on this model.

F. Experiment 6 Result

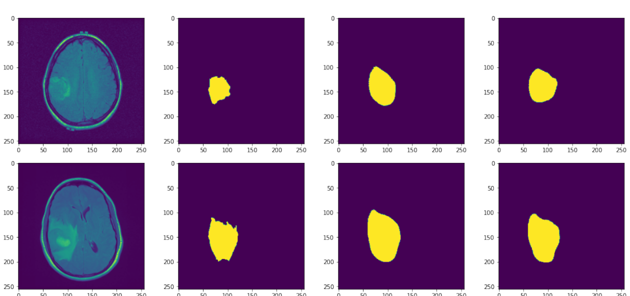

Fig. 17. First column as the image, second column as the mask, third column as the prediction by original Unet, fourth column as the prediction by modified Unet model

After modified the original Unet model, we compare our new model to the original model to compare the segmentation result. In the Figure 17 shown above, we randomly pick two images and their corresponding masks from the data-set. We used our original Unet model to predict the segmentation result, and then we used our modified Unet model to make prediction on the same image as well. The result shows the modified Unet model has slightly better performance than the original Unet model. An additional layer of convolution make the acquired features more detailed and eliminate part of the wrong segmentation prediction.

6. Comparison

In paper [1], it focus on developing a novel 2D fully convoluted segmentation network that is based on the U-Net architecture to boost the segmentation accuracy. They applied soft Dice as loss function and achieving the accuracy around 87%. Which soft dice formula is below:

$$Soft\_Dice = \frac{2TP}{FP + 2TP + FN} \tag{6}$$

It evolved from dice loss. And our paper is main focus on the impact of loss function on this problem. We used four different loss function, the four loss function we used are famous the existing loss function for the different segmentation problem. And in the result compare we can see that in the brain tumor MRI segmentation, the performance of Focal Tversky Loss, Dice Loss and Iou Loss are similar that around 82% to 87%, and binary cross-entropy loss may let model reach to 94% under the same model structure, training data and testing data. As we know, those loss formula is popular on the image segmentation problem. But our data-set contains a large number of MRI without brain tumor, and the mask with that MRI is all black that those Loss function does not work anymore. It will reduce the learning efficiency of the model that model can not find True Positive part, but binary cross-entropy loss function will not be bothered under this limitation. So when the data is not so perfect, binary cross-entropy is a good choice to implement the model. And also under the same setting, if we add downsamplingblock and upsampling block will improve the a little bit of accuracy, but also computing cost will increase.

7. Conclusion

In this paper, we presented some changes that can be made to the current existing U-Net to detect and segment out the brain tumor region. Based on the dataset that were found in Kaggle, which contains 3929 brain MR images together with manual FLAIR abnormality segmentation masks, we have shown that the changes we made can enhance the efficiency and accuracy to the segmentation. Not only we improve the accuracy but also we adopted the data augmentation technique to make the model more robust. We also modified the Unet model so there are one extra downsampling block and upsampling block added to the original model. We have performed six experiments, and each experiment has its own advantages and disadvantages. Among all the experiments we had done, we observed that binary cross-entropy is better suited compare to other loss function when come into two classes segmentation problem. We also observed that our modified model by adding extra convolutional blocks can capture more features of the raw input image and therefore achieve better segmentation ability than the original model. Beyond the better result, the modified model would have longer training time compare to the original model. We hope it gives the secondary verification on tumor contouring, and not only for brain tumor but also other medical analysis as well.

References

[1] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *Medical Image Understanding and Analysis*, M. Valdés Hernández and V. González-Castro, Eds. Cham: Springer International Publishing, 2017, pp. 506–517.

[2] 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*, 2015, pp. 3431–3440.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[4] T. C. I. Archive, "Lgg segmentation datase," 2020.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[6] D. Godoy, *Understanding binary cross-entropy / log loss: a visual explanation*, 2019 (accessed May 10, 2020), https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a.

[7] Wikipedia contributors, "Similarity learning — Wikipedia, the free encyclopedia," 2020, [Online; accessed 10-May-2020]. [Online]. Available: https://en.wikipedia.org/wiki/Similarity_learning#Metric_learning

[8] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," 2018. [Online]. Available: https://www.groundai.com/project/a-novel-focal-tversky-loss-function-with-improved-attention-u-net-for-lesion-segmentation/1

[9] Y. Ye, "Summarization of measurement methods of image semantic segmentation accuracy," 2019. [Online]. Available: https://www.cnblogs.com/yumoye/p/11222388.html