Brain Tumor Segmentation using MRI Image

Project Report of Medical Image Computing
Submitted in partial fulfillment of
the requirements for the degree of
Master of Technology
by

Vidushi Garg(203050027)
Swapnil Malviya(203050067)
Supervisor:
Prof. Suyash P. Awate

Computer Science and Engineering Department
Indian Institute of Technology Bombay
Mumbai 400076 (India)

10 May 2021

Table of Contents

Chapter 1

Introduction

Brain tumors are one of the leading causes of death for cancer patients, especially children, and young people. Gliomas are primary brain tumors which originate from brain cells. Gliomas can be divided into two parts: low-grade gliomas (LGG) and high-grade gliomas (HGG). HGG are malignant brain tumors that require surgery and radiotherapy, and has a poor survival prognosis.

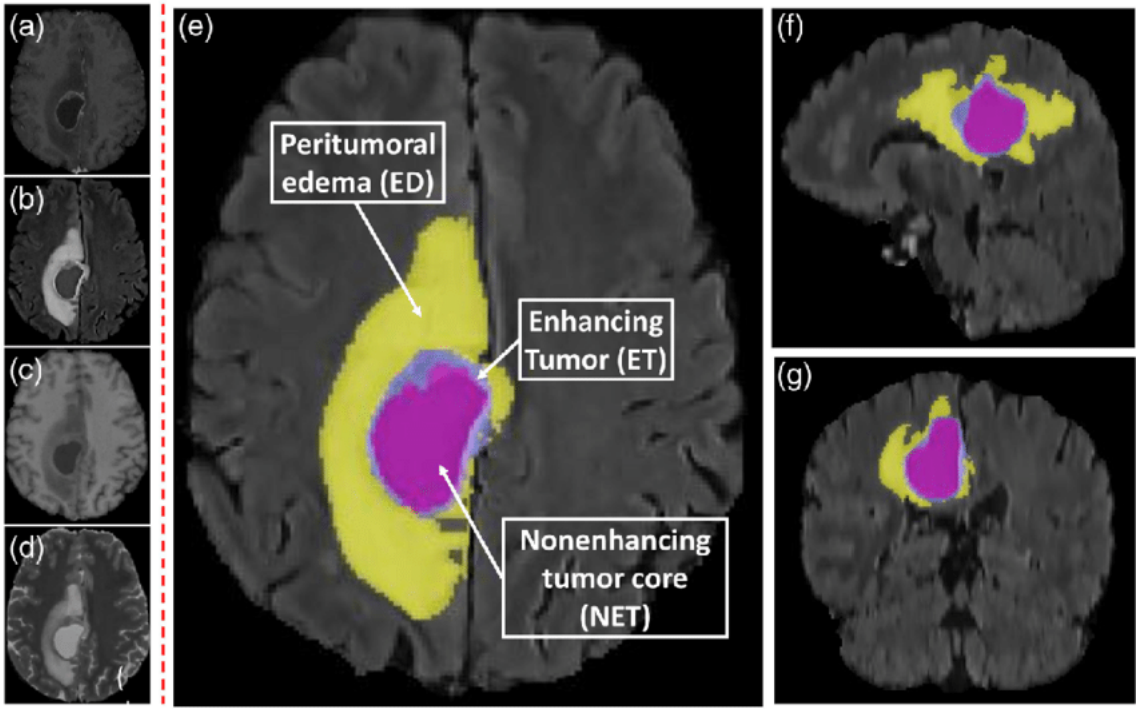


Figure 1.1: (a)T1 (b) T1-weighted (c) Flair (d) T2 (e) axial view of Segmented Brain Tumor MRI Image (f) sagittal view (g) coronal view

Quantitative analysis of brain tumors is critical for clinical decision making. While manual segmentation is tedious, time-consuming, and subjective, this task is at the same

time very challenging to automatic segmentation methods. Deep learning techniques, especially convolutional neural networks are extremely powerful in segmenting the images.

In our project, we have segmented the HGG MRI image into three types of tumor i.e. non-enhancing tumor, enhancing tumor and edema tumor using deep learning models. The BraTS'19 dataset is used for brain tumor segmentation which contains 259 MRI images along with their labelled mask. We have used UNet 3D and ResUnet 3D to label the MRI image.

Chapter 2

BRATS'19 Dataset

Multimodal Brain Tumor Segmentation Challenge 2019 contains 259 MRI scans of patients with high-grade gliomas (HGG). Each image is of size 240x240x155. All BraTS multimodal scans are available as NIfTI files (.nii.gz) and describe a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes, and were acquired with different clinical protocols and various scanners from multiple (n=19) institutions.Menze BH (2019)
Each image contains three types of tumor: Non-enhancing Necrotic tumor (NET), enhancing tumor (ET), and edema (ED) tumor. Further, the segmentation task defines three sub-regions for evaluation, they are 1) the tumor core (TC) including NET and ET, 2) the ET area, 3) the whole tumor (WT) which is the combination of NET, ET and ED.

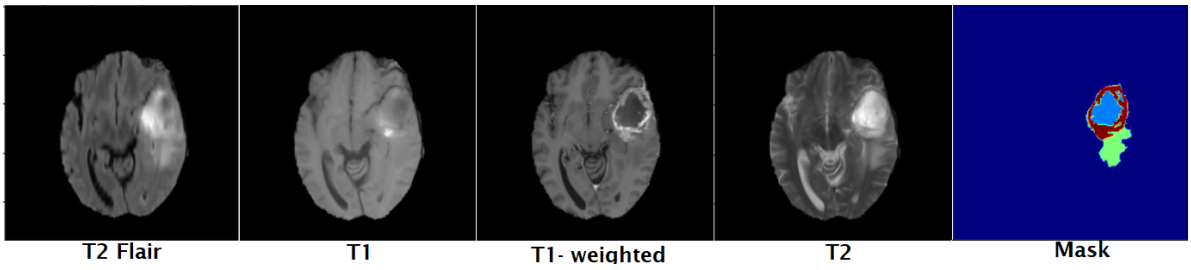


Figure 2.1: Slice of the MRI Image from Brats'19 dataset

Chapter 3

Preprocessing of data

In biomedical field, the dataset is not freely available because the images can only be annotated by experts. Hence, the size of dataset is small. BraTS'19 dataset contains only 259 MRI images along with their labelled mask. To overcome this problem, data augmentation technique is used. For data augmentation, the 3D MRI image is scaled, rotated and translated.

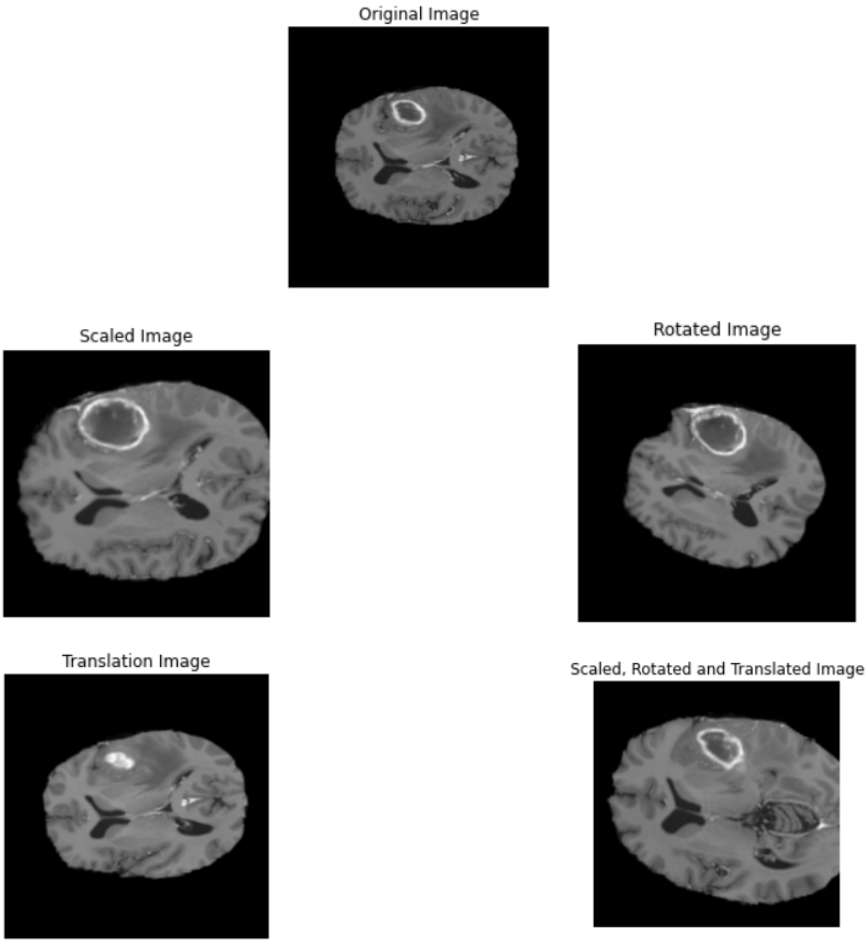


Figure 3.1: Slice of affine transformed MRI Image

Chapter 4

Segmentation Models

4.1 UNet-3D

4.1.1 Introduction

U-Net(Ronneberger *et al.*, 2015) was developed by Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger in 2016. It is a convolutional network for Biomedical Image Segmentation. In biomedical image processing, the desired output should include localization, i.e., a class label is supposed to be assigned to each pixel which is known as semantic segmentation. In this research paper, UNet architecture works with very few training images and yields more precise segmentation with the help of data augmentation.

4.1.2 Architecture

- The network has two parts: the downsampling part and the upsampling part.
- It consists of the repeated application of two 3x3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2x2 max pooling operation with stride 2 for downsampling.
- At each downsampling step, the number of feature channels are doubled.
- In upsampling, 2x2x2 up-convolutional is followed by concatenation of feature map from contracting path, two 3x3x3 convolutional each followed by Relu.
- Concatenation helps in providing localized information from downsampling path to the upsampling path.

- At the final layer a 1x1x1 convolution is used to map each 64- component feature vector to the desired number of classes which is 4 in our case.
- 'Batch Normalization' is added before ReLu layer so that every batch is normalized during training with its mean and standard deviation and global statistics are updated using these values.

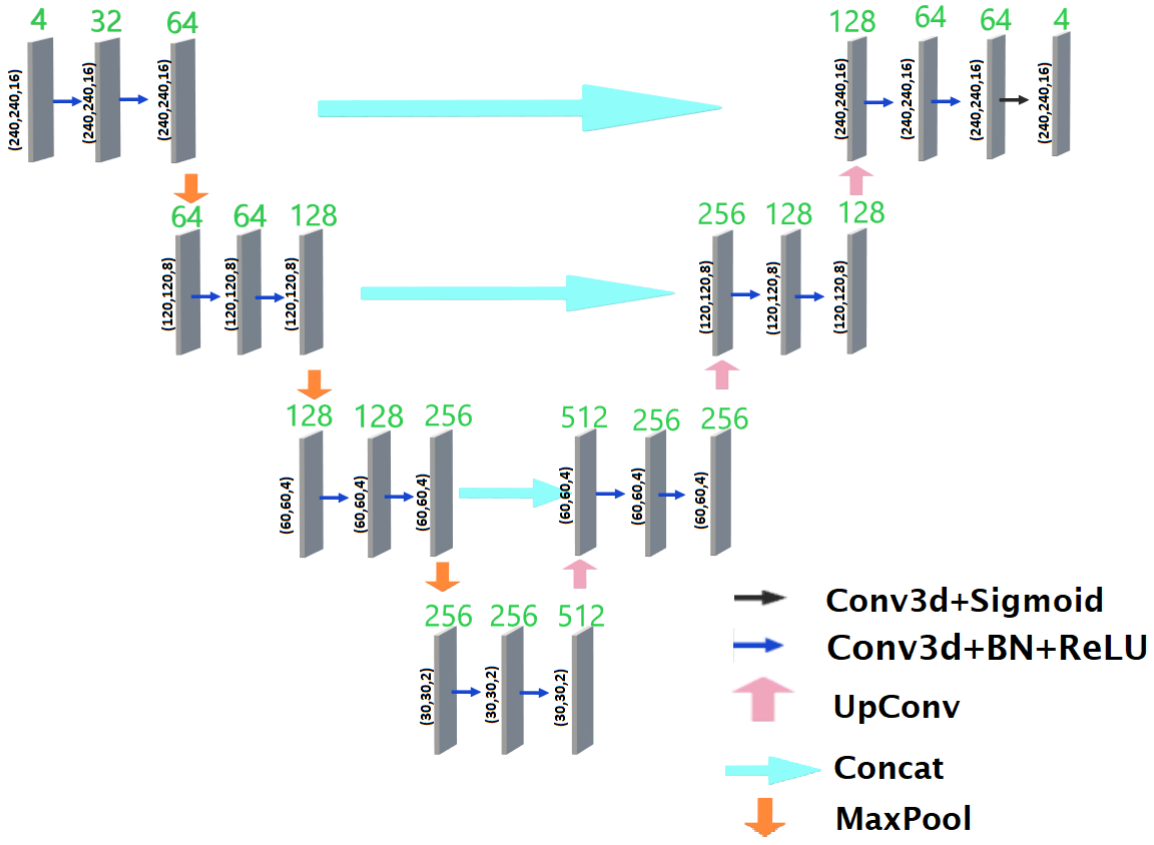


Figure 4.1: UNet architecture

4.2 ResUNet-3D

4.2.1 Introduction

ResUNet is a combination of ResNet and UNet. In ResUnet, Residual neural network acts as an encoder and UNet acts as a decoder. This idea was published by Zeineldin, Ramy A and Karar, Mohamed E and Coburger, Jan and Wirtz, Christian R and Burgert, Oliver in 2020(Zeineldin *et al.*, 2020) for segmenting a 2d slice of MRI images. Since, we need to segment the 3D MRI images so we have modified the 2D ResUnet architecture to 3D architecture.

4.2.2 Residual Block

Due to vanishing gradient problem, the training and test error starts increasing as the number of layers increase. In residual network, "skip-connection" technique is used to overcome the problem of vanishing gradient. The skip connection skips training from a few layers and connects directly to the output. The advantage of skip connection is that if any layer hurt the performance of the architecture then it will be skipped by regularization.

$$y = F(x, \{W_i\}) + x$$

$$F = W_2\sigma(W_1 x)$$

where σ is the RELU function

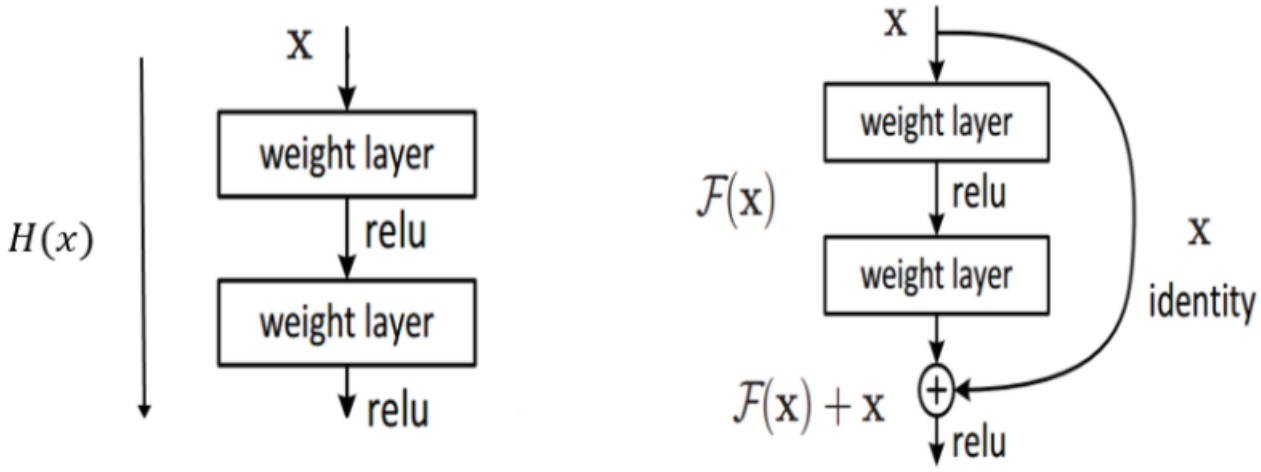


Figure 4.2: Left: Normal CNN. Right: CNN with Residual Connection (Baki, 2016)

4.2.3 Architecture

- The network has two parts: the downsampling part and the upsampling part.
- The downsampling part is similar to a ResNet-50 architecture and the upsampling path is similar to UNet 3D architecture.
- Skip connections have been used in the resnet and identity block to avoid the vanishing gradient problem.
- The architecture consists of the repeated application of two 3x3x3 convolutions (unpadded convolutions), each followed by batch normalization(BN) and rectified linear unit (ReLU)
- A 2x2x2 max pooling operation with stride 2 is used for downsampling.
- At each downsampling step, the number of feature channels are doubled.
- In upsampling, 2x2x2 up-convolutional is followed by concatenation of feature map from contracting path, two 3x3x3 convolutional layer each followed by Relu.
- Concatenation helps in providing localized information from downsampling path to the upsampling path.
- At the final layer a 1x1x1 convolution is used to map each 64-component feature vector to the desired number of classes which is 4 in our case.

4.2.4 ResUnet-2D

We have also the implemented the same architecture using 2D Convolutions. 3x3 kernels were used in the 2D ResUnet Model. 2D ResUnet model does not account for the dependency between different slices while 3D model also captures dependency across the slices. 2D model takes less time to train as compared to 3D ResUnet model. In ResUnet-2D, the model was trained using only the T2-weighted MRI images.

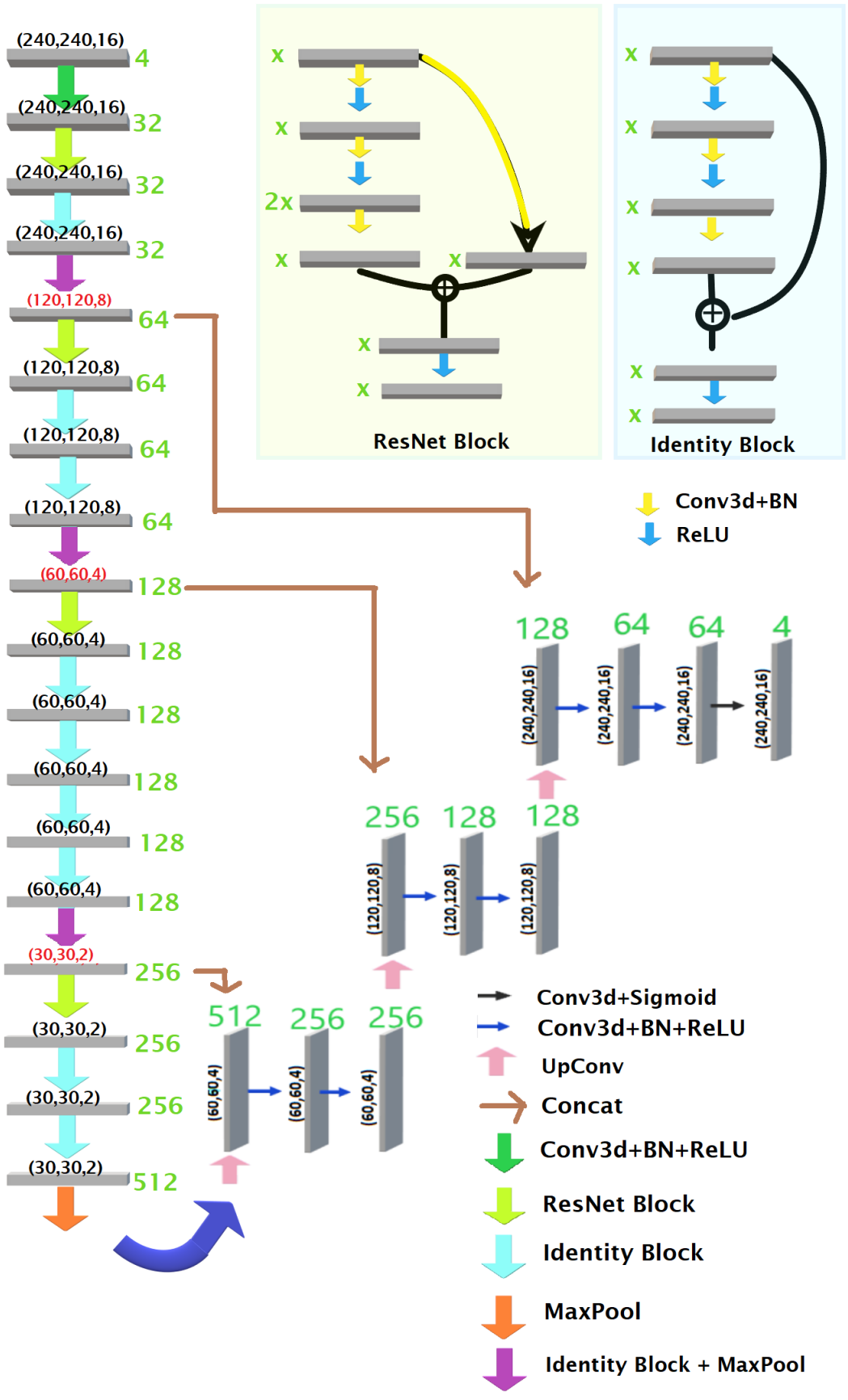


Figure 4.3: ResUNet architecture

4.3 UNet-2D

4.3.1 Introduction

U-Net(Ronneberger *et al.*, 2015) was developed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015. It is a convolutional network for Biomedical Image Segmentation. In biomedical image processing, the desired output should include localization, i.e., a class label is supposed to be assigned to each pixel which is known as semantic segmentation. In this research paper, UNet architecture works with very few training images and yields more precise segmentation with the help of data augmentation.

4.3.2 Architecture

- The network has two paths: the downsampling path and the upsampling path.
- The network contains 32 convolutional layers.
- It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling.
- At each downsampling step, the number of feature channels are doubled.
- In upsampling, 2x2 up-convolutional is followed by concatenation of cropped feature map from contracting path, two 3x3 convolutional each followed by Relu.
- Concatenation helps in providing localized information from downsampling path to the upsampling path.
- At the final layer a 1x1 convolution is used to map each 64- component feature vector to the desired number of classes.
- During training, background loss was also considered while calculating the dice coefficient

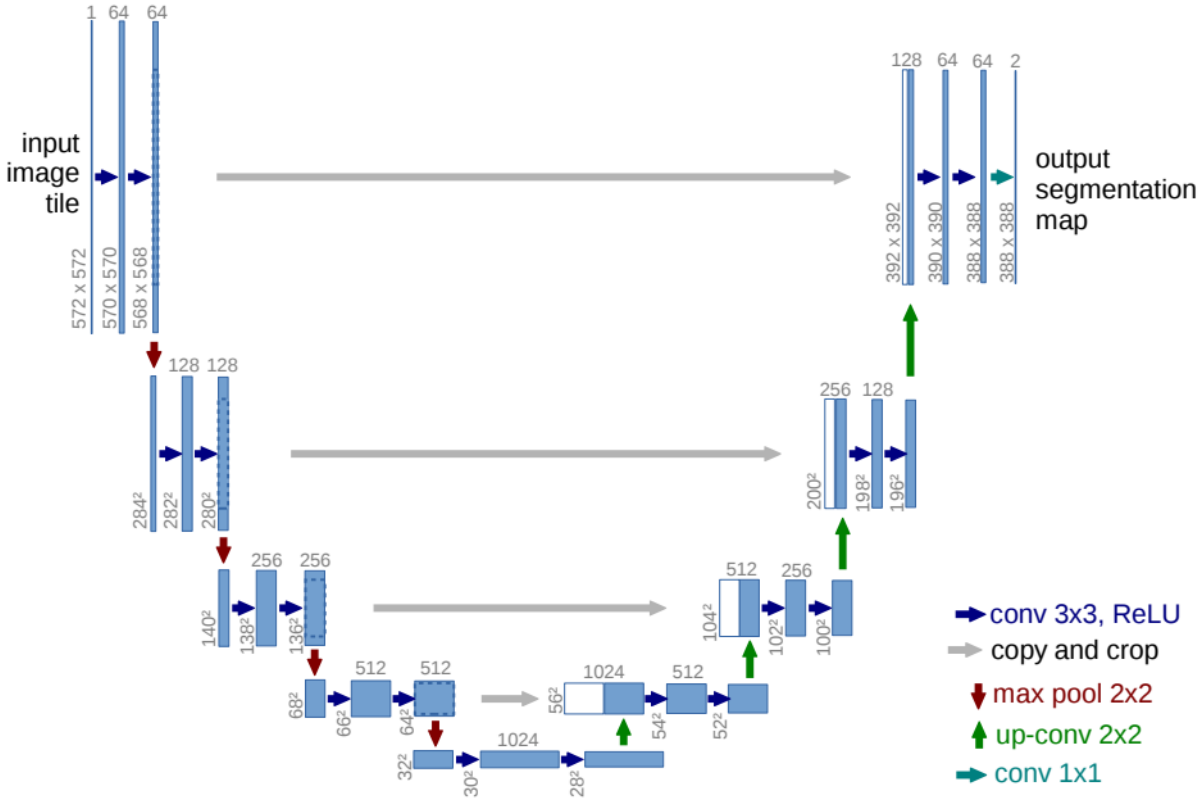
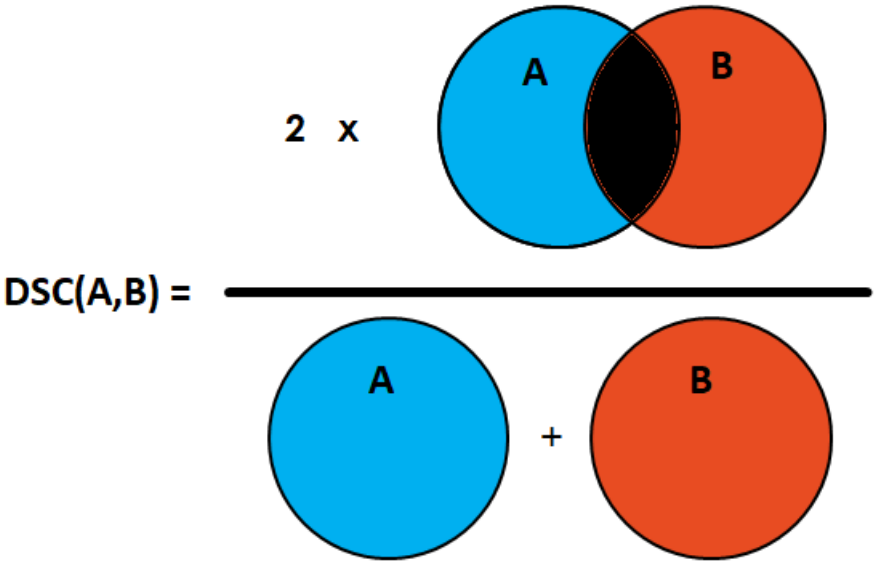


Figure 4.4: UNet architecture (Ronneberger *et al.*, 2015)

Chapter 5

Loss function

Dice similarity coefficient is taken as the metric to calculate loss function. It is a measure of how well the two contours overlap. The Dice index ranges from 0(complete mismatch) to 1(perfect match). Assuming A as the predicted tumor segmented image and B is the ground truth. The dice coefficient is as follows:



$$DSC(f, x, y) = \frac{2 \times \sum_{i,j} f(x)_{ij} \times y_{ij} + \epsilon}{\sum_{i,j} f(x)_{ij} + \sum_{ij} y_{ij} + \epsilon}$$

where,

x = input image

f(x) = predicted image

y = ground truth

ϵ = small number added to avoid division by zero error

Chapter 6

Result

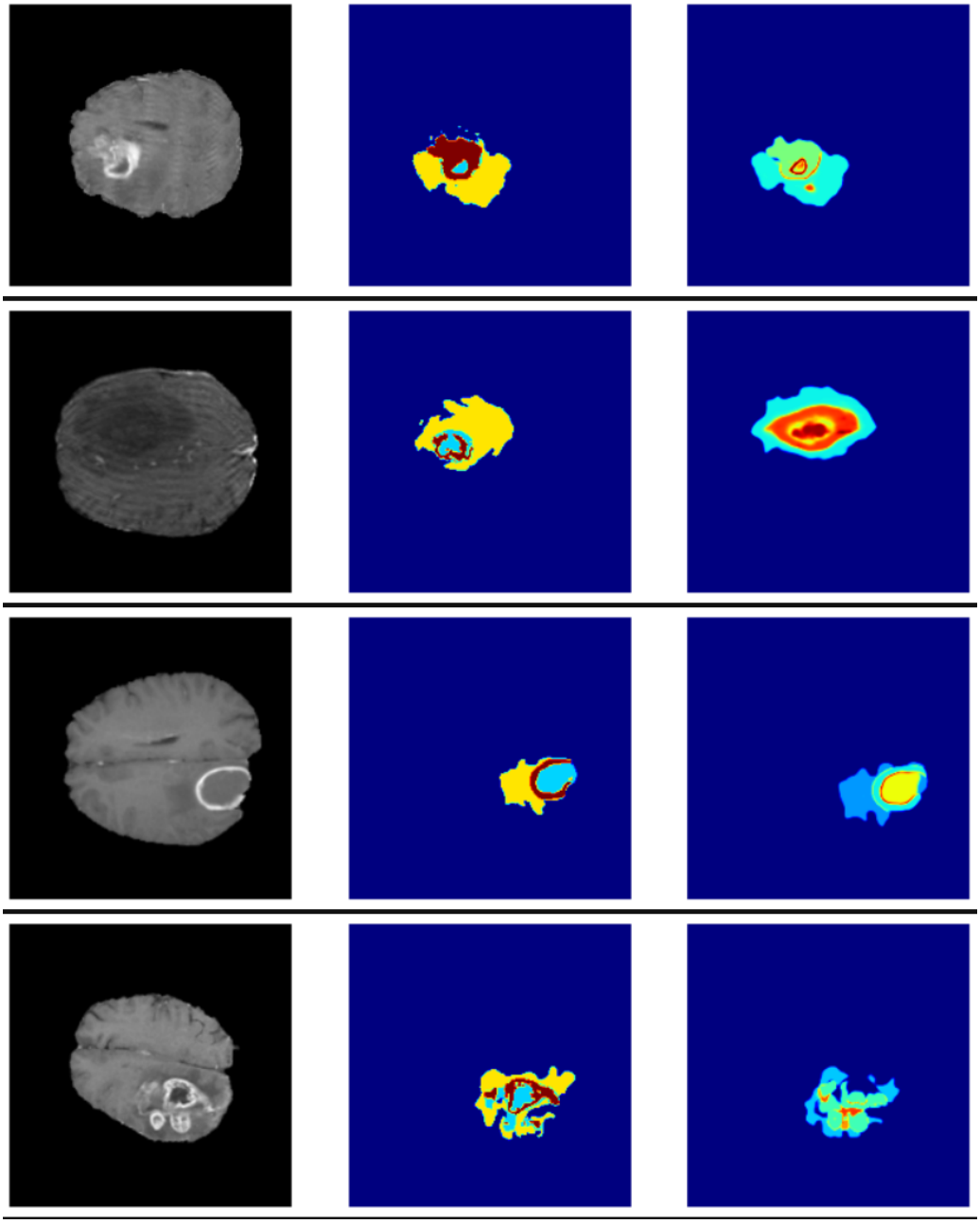


Figure 6.1: For Unet-3D: Left: Input , Middle: Ground Truth, Right: Predicted Mask

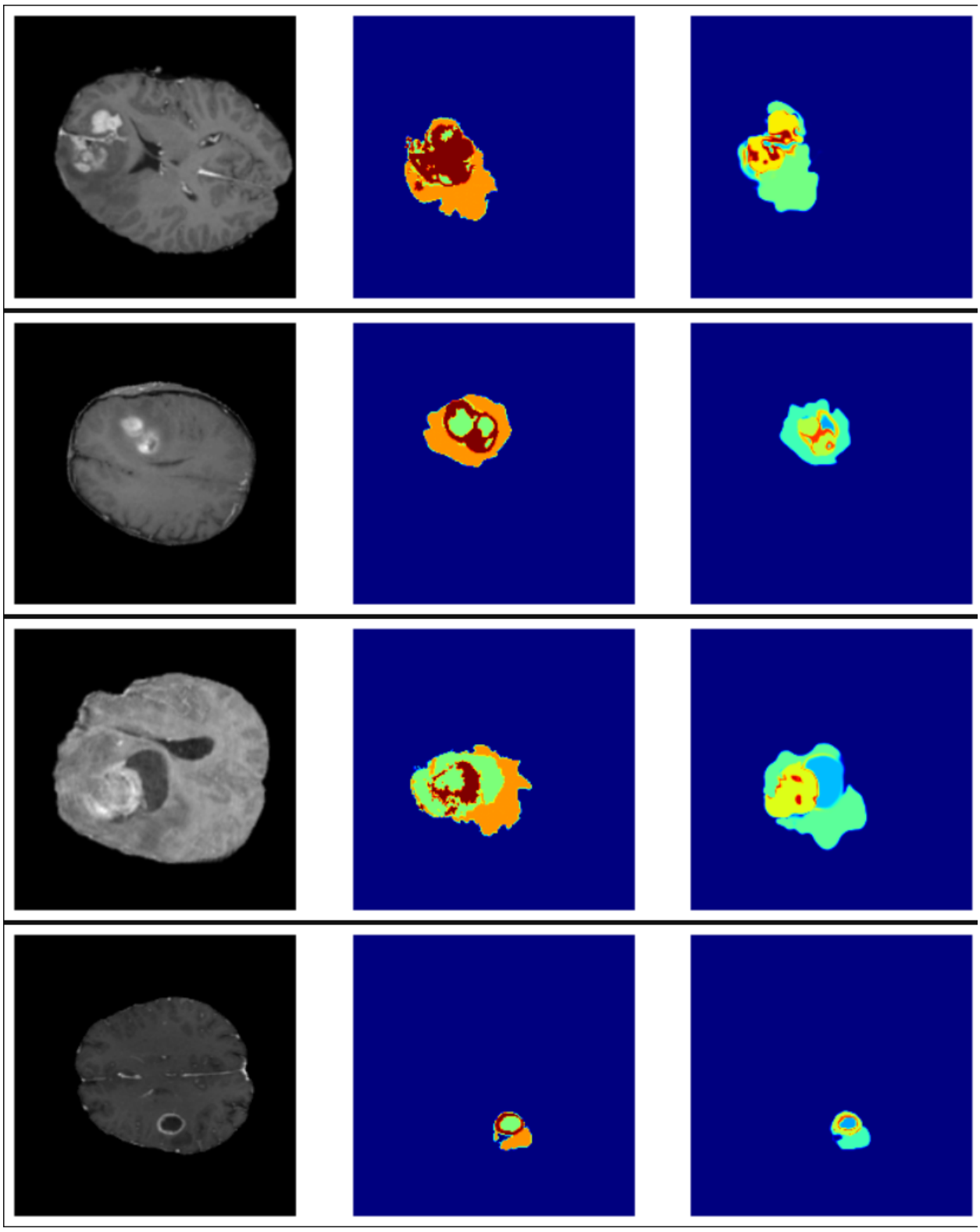


Figure 6.2: For ResUnet-3D: Left: Input Image, Middle: Ground Truth, Right: Predicted Mask

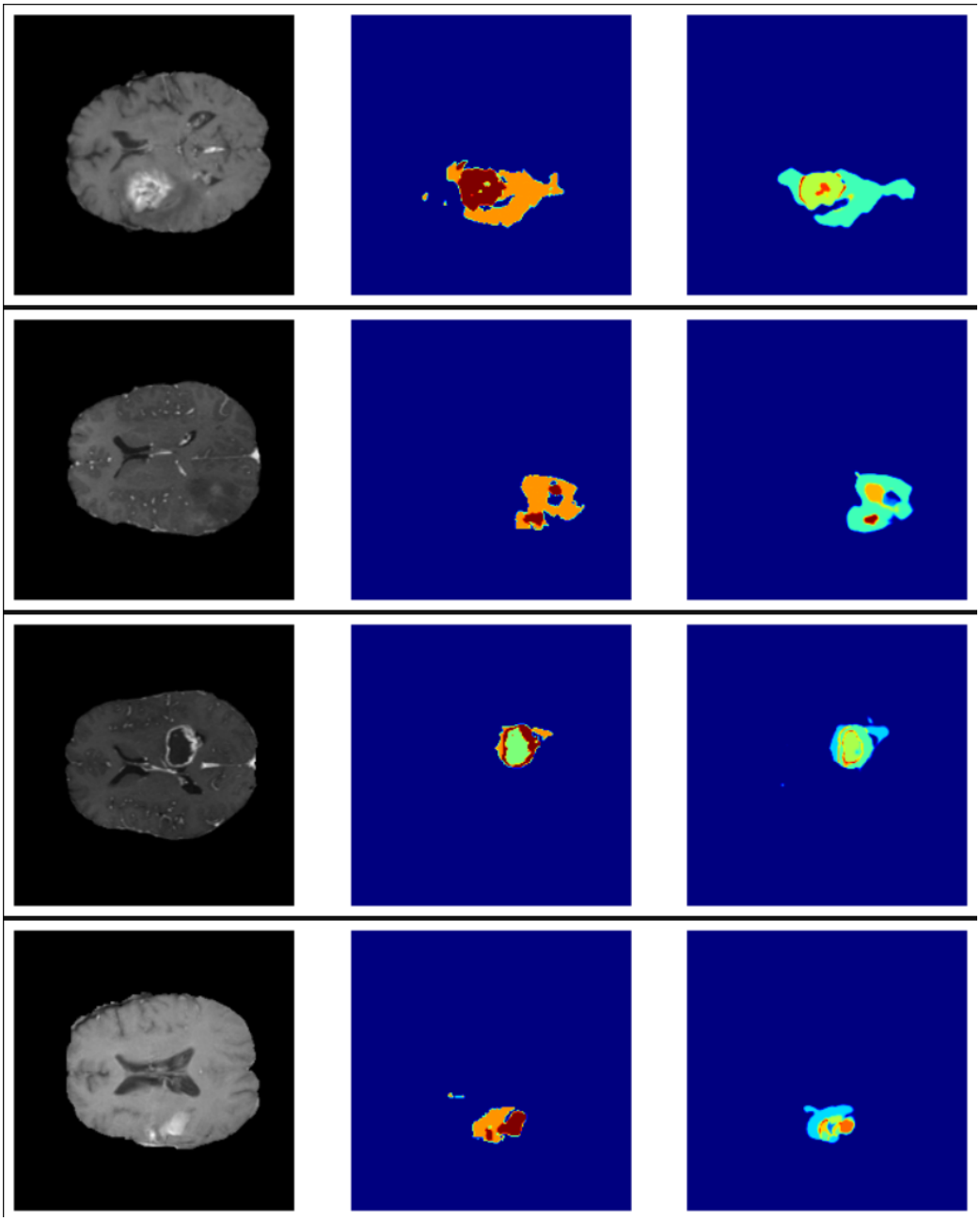


Figure 6.3: For Unet-2D: Left: Input Image, Middle: Ground Truth, Right: Predicted Mask

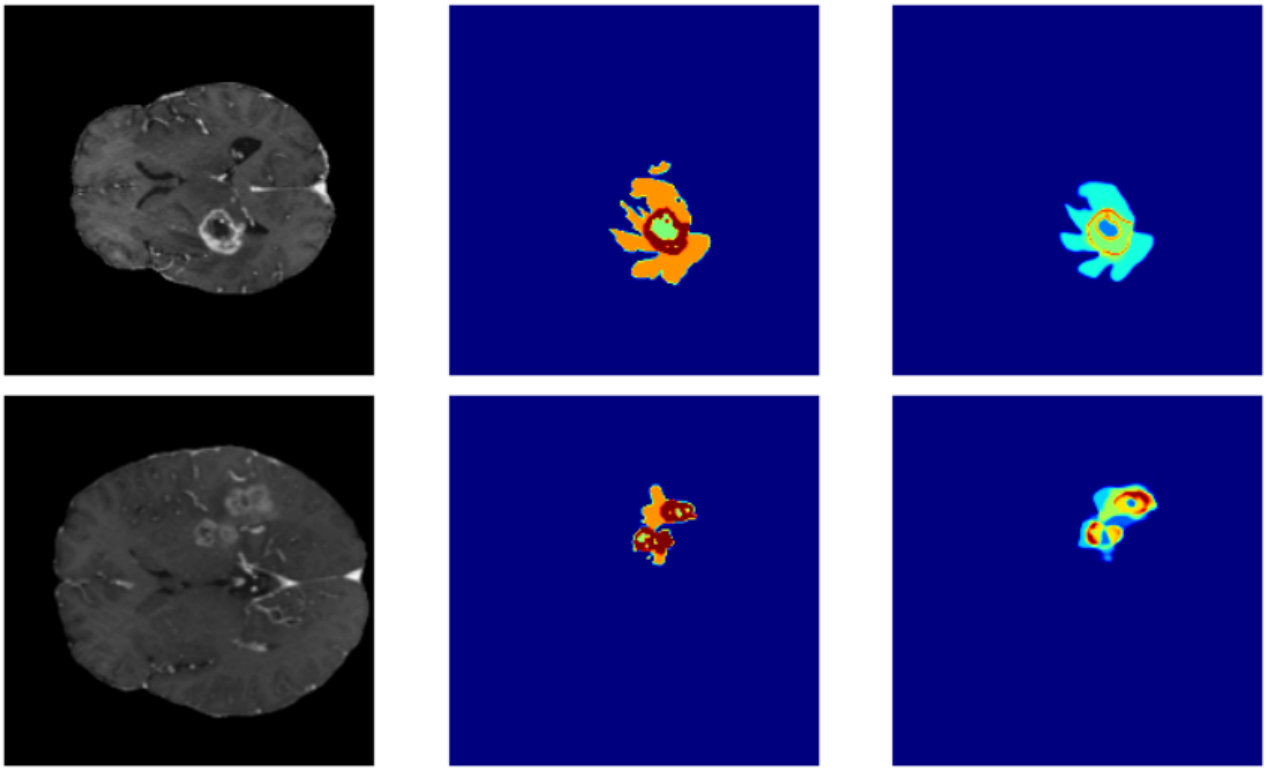


Figure 6.4: For ResUnet-2D: Left: Input Image, Middle: Ground Truth, Right: Predicted Mask

6.1 Accuracy on Training and Test Set

Dice TC = Dice coefficient of Tumor Core(TC) which is the average of dice coefficients of enhancing tumor(ET) and non-enhancing tumor(NET).
Dice ET = Dice coefficient of enhancing tumor(ET).
Dice WT = Dice coefficient of Whole Tumor(WT) which is the average of dice coefficients of enhancing tumor(ET), non-enhancing tumor(NET) and edema tumor(ED).

Model	Epochs	Dice TC	Dice ET	Dice WT
UNet-3D	15	0.35395	0.4332	0.3940
ResUnet-3D	10	0.37995	0.4729	0.4300
Unet-2D	40	0.7510	0.7201	0.7466
ResUnet-2D	30	0.6663	0.6781	0.6815

Table 6.1: Dice Coefficient on Training Set

Model	Epochs	Dice TC	Dice ET	Dice WT
UNet-3D	15	0.37485	0.457	0.4089
ResUnet-3D	10	0.30785	0.3881	0.3614
Unet-2D	40	0.7212	0.7090	0.7166
ResUnet-2D	30	0.6443	0.6981	0.6622

Table 6.2: Dice Coefficient on Test Set

6.2 Conclusion

Based on Quantitative results, ResUNet performed better than UNet. The ResUnet-3D model was trained for only for 10 epochs while Unet-3D was trained for 15 epochs, still the performance of ResUnet-3D was similar to Unet-3D. If the 3D model is trained for more than 40 epochs then the accuracy will improve. The hyperparameters can be tuned further to improve the accuracy of the 2D model.

References

Baki, 2016 Aug, "Microsoft presents : Deep residual networks,"

Menze BH, J. A., 2019, "Cbica image processing portal by center for biomedical image computing and analytics, university of pennsylvania,"

Ronneberger, O., Fischer, P., and Brox, T., 2015, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Springer). pp. 234–241.

Zeineldin, R. A., Karar, M. E., Coburger, J., Wirtz, C. R., and Burgert, O., 2020, "Deepseg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance flair images," *International journal of computer assisted radiology and surgery* **15**, 909–920.