MULTIPLANAR CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

by

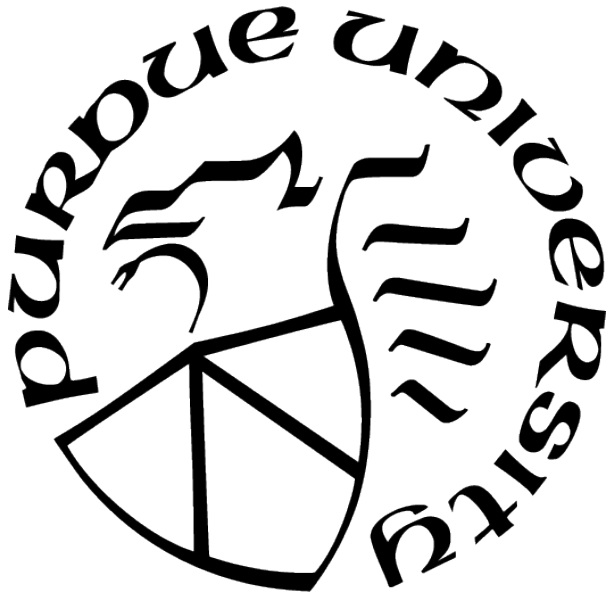
Somosmita Mitra

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Electrical and Computer Engineering



School of Electrical and Computer Engineering

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Thomas Talavage, Chair

School of Electrical and Computer Engineering

Dr. Chris Brinton

School of Electrical and Computer Engineering

Dr. Mary Comer

School of Electrical and Computer Engineering

Approved by:

Dr. Dimitri Peroulis

ACKNOWLEDGMENTS

I would like to express my gratitude to Prof. Thomas Talavage for his unwavering support and encouragement throughout my graduate study. I would also like to thank Prof. Chris Brinton for his words of encouragement, support, and corrections throughout my degree. Thank you to Prof. Mary Comer for her inquisitive questions and feedback through this experience. Finally, I am grateful to my mother and my friends for supporting me through this experience.

TABLE OF CONTENTS

LIST OF TABLES 6

LIST OF FIGURES 7

ABSTRACT 12

1 INTRODUCTION 13

1.1 Brain Tumors 13

1.2 Imaging 14

1.3 MR Imaging for Brain Tumors 14

1.4 Objective 16

2 BACKGROUND 17

2.1 Machine Learning 17

2.1.1 Neural Learning 18

2.1.2 Deep Learning 20

2.2 Using Artificial Intelligence for Brain Tumor Segmentation 25

3 EXPERIMENTATION 30

3.1 Image Segmentation Methodologies 30

3.1.1 Convolutional Neural Networks 30

3.1.2 Generative Adversarial Networks 31
3.1.3 Conditional Generative Adversarial Networks 33
3.2 Dataset 33
3.3 Methodology 35
3.3.1 Pre-Processing 36
3.3.2 Network Design 40
3.3.2.1 Segmenter/Generator Design 43
3.3.2.2 Discriminator Design 47
3.3.3 Training 48
4 SUMMARY 59
REFERENCES 60

LIST OF TABLES

1.1 Comparison between T1 vs T2 vs FLAIR MRI sequences 16
3.1 Dice Scores for the Proposed Network 51
3.2 Specificity for the Proposed Network 51
3.3 Sensitivity for the Proposed Network 51

LIST OF FIGURES

1.1 Image patches show from left to right: the whole tumor visible in FLAIR (A), the tumor core visible in T2 (B), the enhancing tumor structures visible in T1c (blue), surrounding the cystic/necrotic components of the core (green) (C). Segmentations are combined to generate the final labels of the tumor structures (D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core(blue). 14

2.1 The mathematical function representing a neuron. Here, the weights (w_1,w_2,w_3,w_4) are multiplied by the inputs (x_1,x_2,x_3,x_4) and a bias term (b) is added to this neural unit. The figure shows a 4 input configuration. The output of the neuron is passed through an activation function to generate the final outcome (y) 18

2.2 A Single layer Neural network, with three neurons in the input layer, 4 neurons in the hidden layer and 2 neurons in the output layer. 19

2.3 A Multi Layer Perceptron with two hidden layers. There are 3 neurons in the input layer, 4 neurons in the first hidden layer, 3 neurons in the second hidden layer and 2 neurons in the output layer. 21

2.4 A Convolution defined over a 5 x 5 initial image with a filter of size 3 x 3. The filter/kernel slides over the image and the resultant of the transformation is produced in the output image. The filter sliding over the first row is demonstrated in this figure. 22

2.5 A typical Convolution Neural Network Structure. A single channel brain MR image is input through a three stage CNN with a hierarchical convolution, activation and sub-sampling structure. In each stage there is a Convolution feature map generated, referred to as C1, C2 and C3 respectively, these feature maps are then passed through an activation function introducing non-linearity in the feature space and is followed by a sub sampling feature map, referred to as S1, S2 and S3 respectively. The sub-sampled feature map from the third stage is flatted into a one dimensional feature array and connected to a fully connected neural network. This results in a binary classifier which detects the presence of tumor in the input image. Here, Stage 1 detects patterns in the pixel range, Stage 2 detects edges and Stage 3 performs object recognition. This mimics the human receptive field perception system. 23

2.6 A Convolution Neural Network catering to a 3 D input and a 3 D feature map. The convolutional, non linear and subsampling layers are followed by a mapping to a fully connected neural network. As visible in this figure, there is a multiplicative increase in the number of network parameters due to the addition of a third dimension for each feature map and input. 27

3.1 A Generative Adversarial Network with a generator G and a discriminator D. The real images and the generated images are sampled and passed into the discriminator D. As training progresses the task of the network is to get better at differentiating between the real image and the generated image, while simultaneously training the generator G to produce more real like generated images. 31

3.2 A Conditional Generative Adversarial Network with a generator G and a discriminator D. The real images and the conditionally generated images, along with extra modal information/labels are sampled and passed into the discriminator D. As training progresses the task of the network is to get better at differentiating between the real image and the conditionally generated image, while simultaneously training the generator G to produce more real like conditionally generated images. 33

3.3 Multimodal 2-dimensional MR image slices of the human brain post processing and labelling. The predicted labels are evaluated by merging three regions viz. whole tumor (WT:NCR/NE + ET + ED), tumor core (TC:NCR/NET + ET), and enhancing tumor (ET) as shown. 34

3.4 Three gray-scale images A, B, C, are permuted over the Red, Green and Blue channels to produce combinations of FLAIR, T1C and T2 channels. Each combination is transformed to the CIE- $L*a*b*$ format and displayed between option a to f. The six combinations are ABC, ACB, BAC, BCA, CAB, CBA. Of these we can use wither option a or option b. They have distinctly highlighted brain tumor regions and the rest of the brain matter is considerably suppressed. 36

3.5 On the left we have the Red channel which has the FLAIR MR image. There are non tumor regions in this image which have the same intensity as the tumor region. On the right we have the $L*$ channel of the CIE $L*a*b*$ converted FLAIR image. Although the whole image is more uniformly spread (due to normalization), the tumor region still strikingly stands out from its Red channel gray image equivalent. 37

3.6 The gray scale single channel equivalent of the Green channel is presented here. On the left we have the Green channel which contains the T1C MR image. There are non tumor regions in this image which have the same intensity as the tumor region. On the right we have the $a*$ channel of the CIE $L*a*b*$ converted T1C image. Although the whole image is more uniformly spread (due to normalization), the tumor region still strikingly stands out from its Green channel gray image equivalent. 38

3.7 The gray scale single channel equivalent of the Blue channel is presented here. On the left we have the Blue channel which contains the T2 MR image. There are non tumor regions in this image which have the same intensity as the tumor region. On the right we have the $b*$ channel of the CIE $L*a*b*$ converted T1C image. Although the whole image is more uniformly spread (due to normalization), the tumor region still strikingly stands out from its Blue channel gray image equivalent. 39

3.8 The axial view of the brain is a xy planar view and allows us to maintain the spatial connectivity in brain tumor and brain matter in the xy plane. On the other hand, the coronal view of the brain is a yz planar view. It maintains the spatial connectivity between brain matter and tumor in yz plane. The sagittal plane deals with the xz plane and helps us in our task by maintaining bran matter and tumor connectivity in the xz plane. 40

3.9 In each three dimensional brain MR image, we generate slices in the xy/axial plane, yz/coronal plane, and the xz/sagittal plane. Our proposed algorithm runs this data through three segmenters. The sagittal segmenter is fed image slices in the xz plane, the axial segmenter is fed images in the xy plane and the coronal segmenter has yz planar images supplied to it. The label map across each segmenter is weighted with a 1/3 coefficient. We then compute the final label map/segmented image and compute a dice loss with the ground truth and a binary cross entropy loss in the discriminator. Details of these losses are provided in Section 3.3.3. The combined loss of these metrics is used to optimize the parameters of each segmenter. 41

3.10 On the left we have the loss function being optimised without skip connections. As visible, the surface is undulating and rough. This leads to the gradient descent algorithm being stuck in the local minimas. On the right we have the loss/optimization space being must smoother and significantly reducing the possibility of local minimas. This leads to a more accurate and faster parameter optimization![8] 42

3.11 Represented here is an autoencoder- decoder network with three convolutional layers in each down-sampling stage of the encoder/analysis network. The input dimensions reduce from 256×256 to 32×32 in the bottleneck component and is then up-sampled in consecutive three convolutional layers to the output layer of 256×256. It can be noted that the dimensions of the input and output layer need not be exactly the same. They are represented to be of similar dimensions to maintain coherence through all figures. 43

3.12 Represented here is an autoencoder- decoder network with three convolutional layers in each down-sampling stage of the encoder/analysis network. The input dimensions reduce from 256×256 to 32×32 in the bottleneck component and is then up-sampled in consecutive three convolutional layers to the output layer of 256×256. There are skip connections from the first layer of the network to the $n-1$ layer, second layer to the $n-2$ layer and third layer to the $n-3$ layer. Here, n is considered to be 7 with six layer comprising of 3 sequential convolutions. followed by max-pooling and an activation function. The final layer is a fully connected network leading to the output. 44

3.13 Represented here is an autoencoder- decoder network with three convolutional layers in each down-sampling stage of the encoder/analysis network. The input dimensions reduce from 256×256 to 32×32 in the bottleneck component and is then up-sampled in consecutive three convolutional layers to the output layer of 256×256. There are skip connections from the first layer of the network to the $n-1$ layer, second layer to the $n-2$ layer and third layer to the $n-3$ layer. Here, n is considered to be 7 with six layers comprising of 3 sequential convolutions. followed by max-pooling and an activation function. The final layer is a fully connected network leading to the output. This network also has additional skip connections from the third layer to the $n-2$ and $n-1$ layer. Along with skip connections from the second layer to the $n-1$ layer. 46

3.14 The discriminator has four convolution layers, with each followed by max pooling, rectified linear unit activation and batch normalization. They are then, passed through two fully connected layers and output to a single neuron with a binary classification function. 47

3.15 Our proposed Conditional Generative Adversarial Network with a generator G and a discriminator D. The ground truth and the segmenter produced output, along with label information is sampled and passed into the discriminator D. As training progresses the task of the network is to get better at differentiating between the real image and the conditionally generated image, while simultaneously training the generator G to produce more real like conditionally generated images. 48

3.16 The training loss of the proposed network is illustrated. 51

3.17 On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset. 52

3.18 On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset. 53

3.19 On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset. 54

3.20 On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset. 55

3.21 On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset. 56

3.22 On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset. 57

ABSTRACT

Brain tumor sub region segmentation is a challenging problem in Magnetic Resonance imaging. The tumor regions tend to suffer from lack of homogeneity, textural differences, variable location, and their ability to proliferate into surrounding tissue. The segmentation task thus requires an algorithm which can be indifferent to such influences and robust to external interference. In this work we propose a conditional generative adversarial network which learns off multiple planes of reference. Using this learning, we evaluate the quality of the segmentation and back propagate the loss for improving the learning. The results produced by the network show competitive quality in both the training and the testing data-set.

1. INTRODUCTION

Brain Tumors are the presence of abnormal cells in the tissue of the nervous system which could be benign or malignant depending on the location, mass growth and extent of proliferation of the tumor cells. The human brain is one of the more complex structures of the living world and is primarily responsible for the control, thought and memory operations by humankind. Any aberration in this complex structure can lead to devastating consequences which include death. A very common type of brain tumor is glioma or cancer in the glial cells. This can occur both in the spinal cord and the brain region.

1.1 Brain Tumors

The glial cells are the supporting tissues which surround the neuron structures in the central nervous brain. While gliomas themselves make up to 30% of the tumors in the brain, when it comes to malignancies, in the brain, gliomas comprise of 80% of the more deadly form of cancer. Their rate and extent of proliferation makes them difficult to segment and oftentimes the minutest delay in treatment could prove to be fatal for the patient.

Segmentation refers to the delineation of the different tumor regions. They can be broadly classified as three main portions: whole tumor, enhancing tumor and the core tumor region. These areas help determine the stage of the tumor and the extent of its spread in the healthy tissues of the brain. The wider the spread, the tougher it is to remove without causing unintended harm to the healthy brain cells surrounding these regions. These healthy brain cells are responsible for different vital actions of the human body, which include breathing, circulation of blood, co-ordination of movement, thoughts and influence critical decision making abilities. Medical practitioners often rely on their visual perception of the tumor while making these delineations during surgery. This procedure is prone to human error and oftentimes the classification of the region varies from one surgeon to another. Manual segmentation is also laborious and involves considerable amount of time spent in differentiating between the cancerous and non cancerous regions of the brain.

Operative accuracy of the tumor regions is vital for the removal of the cancer cells and the accuracy involved in brain tumor segmentation can largely influence the survival of the patient. This is where an image based procedure for brain tumor segmentation can prove to be valuable albeit it still has its own challenges.

1.2 Imaging

Magnetic Resonance Imaging or MRIs have been extensively used to study brain tumors due to their ease of use and ability to produce accurate description of the size and heterogeneity of the region. With MRIs the imaging can be obtained and viewed from all three planes of orientation in a three dimensional space- axial, sagittal and coronal. This has facilitated studies on the development of automatic and semi automatic segmentation algorithms for the delineation of the various tumor regions present in gliomas.

The difference between tissue and tumor in MRI scans is characterized by two different relaxation times known as T1 and T2. T1 or longitudinal relaxation time weighted MR sequences denote the decay constant in the z direction for spinning protons to realign with the external magnetic field for different tissues. T2 or transverse relaxation time denotes the decay constant required by spinning protons to lose their phase coherence in the xy plane, or the plane perpendicular to the static magnetic field.

1.3 MR Imaging for Brain Tumors

T1 and T2 weighted MR images produce contrast in images, which become more apparent when dealing with tissues of various densities. In the case of brain tumors, these help provide significant information on the tumor core, enhancing and whole tumor regions. T1C or post contrast with Gadolinium T1 MR sequences reduce the T1 time and result in a higher contrast between tumor regions and healthy brain tissues. Fluid Attenuated Inversion Recovery or FLAIR is another commonly used MRI channel in the study of brain tumor segmentation. MR image sequences in this channel nullifies the effect of cerebral fluid and helps obtain a clearer sketch of the glioma.

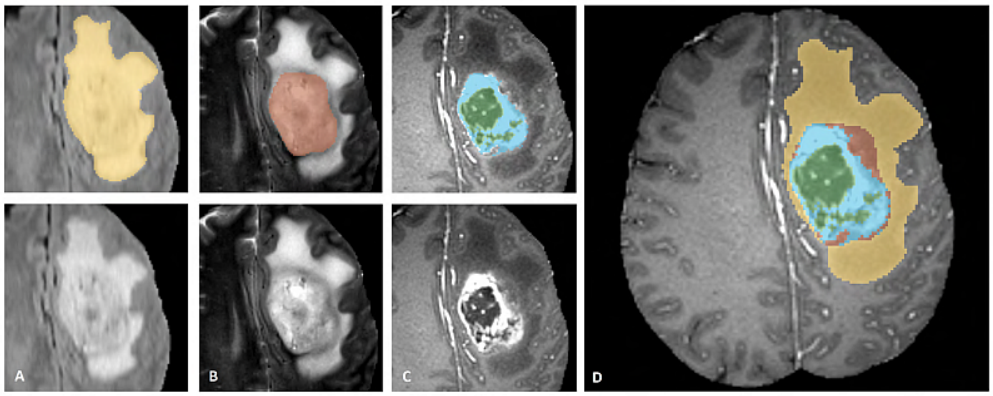


Figure 1.1. Image patches show from left to right: the whole tumor visible in FLAIR (A), the tumor core visible in T2 (B), the enhancing tumor structures visible in T1c (blue), surrounding the cystic/necrotic components of the core (green) (C). Segmentations are combined to generate the final labels of the tumor structures (D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core(blue).

We describe each channel in much detail, due to their scope in providing complimentary information describing the tumor regions, their extent of spread and their localization in the brain, as observed in Fig. 1.1

The tendency of Gliomas to proliferate into surrounding tissue areas without displacing healthy tissue boundaries makes the borders blurry and difficult to separate. Brain tissue images often include artifacts like stroke, gliosis (the change in glial cells due to damage to the central nervous system), and the presence of blood clots. These tend to have similar appearances as tumor regions in the MR images. This, coupled with the varied locations, size and appearances of brain tumors or gliomas in any part of the human brain make it tougher to localize, delineate and classify as tumor.[16]

Table 1.1. Comparison between T1 vs T2 vs FLAIR MRI sequences

Tissue	**T1 Weighted**	**T2 Weighted**	**FLAIR**
Cerebrospinal Fluid	Dark	Bright	Dark
White Matter	Light	Dark Gray	Dark Gray
Cortex	Gray	Light Gray	Light Gray
Fat	Bright	Light	Light
Inflamation	Dark	Bright	Bright

Table 1.1 provides us with a comparison between the different MRI channels used for brain tumor segmentation. These channels together provide critical information which imbibes the tumor features involved in segmentation.

The properties of the MR imaging techniques, i.e. relaxation times, helps make the whole tumor visible in FLAIR, the tumor core visible in T2, the enhancing tumor structures along with the core visible in T1c. However, tumor MR images often tend to suffer from lack of homogeneity across MRI scanners, the presence of scanner specific noise and the noise induced from patient's movement in the scanner. The segmentation task from MR images thus requires an algorithm which can be indifferent to such influences and robust to external interference.

1.4 Objective

We have established that the task of manually segmenting the regions of a brain tumor by any medical practitioner is difficult and prone to human error. In addition to this, the time taken to produce a singular glioma segmentation map is intensive and can cause considerable delay in patient diagnosis and prognosis. The presence of internal scanner noise for MR images and the noise induced by subject movement is another addendum which produces variance between image to image for a single subject in the same scanner as well as for different subjects in the same scanner.

To overcome these adversities in the realm of Brain Tumor Segmentation through the use of MR Imaging techniques, we propose a generative model for brain tumor segmentation, which is described in details in Chapter 3. The data used and the scope of our contribution is highlighted in Chapter 2. We provide our state of the art results obtained in Chapter 4 and describe our finding in Chapter 5.

2. BACKGROUND

Accurate tumor region detection and delineation of segments is crucial for diagnosis, treatment planning, operative accuracy, evaluation of outcome and prognosis. The extensive amount of detail associated with each MR Image of the brain and the large amount of data made available from a variety of source MR scanners, generates the need for a computer based algorithm to demarcate the regions. The complexity of such a task has inspired progress in the development of computer based algorithms. We discuss some of the existing algorithms and techniques used in brain based MR Image segmentation followed by the description of our proposed model, with further elaboration in Chapter 3.

2.1 Machine Learning

With the advent of technology and increase in processor capabilities, the world has seen an increase in the ability of machines to mimic human cognition systems to a more primitive definition of neural learning. Traditionally, machine learning[39] has been the training of computer based algorithms on sets of data- which could range from a singular dimension to the multi dimensional spectrum. This has been followed by their evaluation of performance using a testing set which is independent from the training set. There are three broad categories of machine learning algorithms: Supervised, Unsupervised and Reinforced learning.

Supervised Learning, defined in the scope of machine learning, involves the training of algorithms with a set of input and output pairs. Here, the input is fed into the computer based algorithm, which while training maps it to its output. The algorithm is then penalised or rewarded based on its accuracy obtained in the mapping task.

Unsupervised Learning, within the scope of machine learning or artificial intelligence, is defined as the drawing of inferences from the hidden patterns in data. In other words, the computer based algorithm observes the input and tries to group them based on their similarity of features. This form of learning does not involve any external feeding of output labels and thus, is without human supervision.

The third form of learning in the realm of machine intelligence, that is called reinforced learning. This type of learning depends on feedback which is reinforced into the learning

system as either a positive or a negative signal. This can be seen as being similar to training behavior pattern in children. The main idea being good behavior is rewarded and bad behavior is penalised. Through this process the system learns to move in the direction of what can be construed as good behavior and successfully optimizes the learning framework without the involvement of predefined output labels.

Since the time of inception of artificial intelligence, a term which is often interchangeably used with machine learning, the availability of resources, both in the form of data and processor capabilities has played a major role in determining the progress of such algorithms. Gathering of data is a mammoth task, especially in the bio-medical image field of application. When dealing with Magnetic Resonance Imaging, we come across scanners with various magnet configurations, imaging procedures and modalities. What makes it especially hard for a researcher in such a field is generating a uniform basis for data gathering and classification.

With time, as the speed of the processors increased, the human supervision involved in computer based algorithms have seen a certain degree of reduction. This has led to the refinement of algorithms to cater to the gradual absence of human manipulation on systems- based on visual analysis or statistical data. In other words, with time computer based learning algorithms have been provided with an increased degree of freedom while exploring the feature space of the input. It is with this development that machine learning has metamorphosed into its more advanced and independent form of algorithmic versatility known as deep learning. The ability of such models to work on larger quantities of data and learn from a more fundamental feature format has led to its success in classification, segmentation, analysis and generation tasks.

2.1.1 Neural Learning

The introduction of the neuron[33] as a mathematical function that mimics the functioning of the neurons in the human brain brought in the possibility of neural learning. Here a neural unit, as shown in Fig. 1.2 represents a mathematical function with inputs, which are multiplied by the weights of these connections.

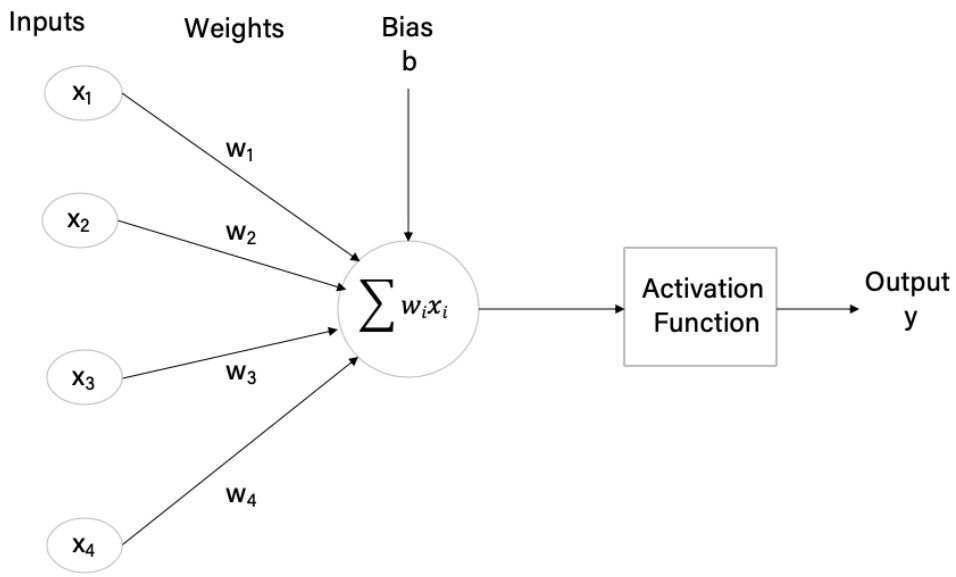


Figure 2.1. The mathematical function representing a neuron. Here, the weights (w_1,w_2,w_3,w_4) are multiplied by the inputs (x_1,x_2,x_3,x_4) and a bias term (b) is added to this neural unit. The figure shows a 4 input configuration. The output of the neuron is passed through an activation function to generate the final outcome (y)

This was followed by the addition of a neural bias and the resultant output was produced through an activation function. Using multiple neural units in a a multi layer framework, the multiple layer perceptron [46] came into existence. The traditional definition of a multi layer perceptron or MLP required the arrangement of one or more neural units in a layer and the repetition of at least three such layers. The input to the neurons in these layers are either the data from the training set or the output of the previous layer. The weights in each connection are adjusted depending on the gradient of the loss of the resultant output produced by the network. The layer where the input data is fed into the network is called the input layer and the last layer of the network is called the output layer. Any layers in between the input and the output layer are known as the hidden layers.

MLPs were restricted to the single hidden layer configuration- due to the lack of large enough processors and the multiplied need for weight and bias parameters with each additional layer. This single hidden layer configuration could not solve the logical XOR problem

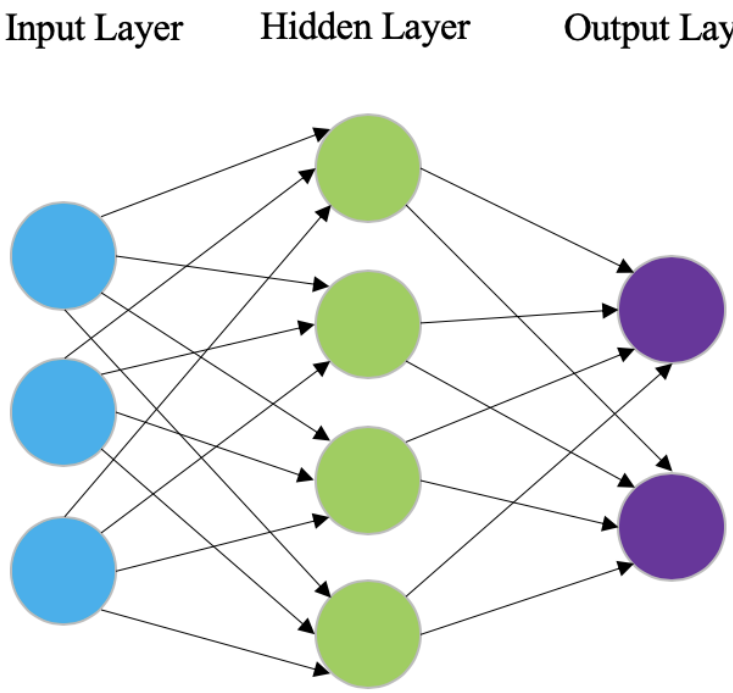


Figure 2.2. A Single layer Neural network, with three neurons in the input layer, 4 neurons in the hidden layer and 2 neurons in the output layer.

and led to a stagnation in the field of Neural Learning, which exposed a single layer perceptron's failure to work on non-linearly separable problems. Due to this Artificial Neural Networks were not a part of popular Machine Learning research for some time. This, however, did not prevent models like the Neocognitron [14] to be developed. Much progress could not be made on such models due to the limitations in processor speeds, capability and memory.

2.1.2 Deep Learning

When comparing between Deep Learning and traditional Machine Learning algorithms, solely based on its capacity to handle data, it would not be incorrect to claim that deep learning caters to a much larger data group and has a higher model complexity than any other machine learning framework. This provides users with a quantitative advantage, as long as the resource requirements are met.

In theory however, the contribution of the deeper model framework resonates to more fundamental basis and caters to a more human like or logical process of thinking. Computer based algorithms for the purpose of learning consists of a certain number of trainable parameters which are known as hyper-parameters. The larger number of parameters we can supply an algorithm with, the higher degrees of learning can occur. However, this comes with its own set of constraints. In an ideal framework, with infinite amount of time and parameters, computer based algorithms can learn and perform well on a multitude of tasks [17]. With limited resources, i.e. the lack of infinite parameters and time, stems the use of a small parameter space and a resource constrained training time. Deep Learning provides the advantage of a larger parameter space and a method for computer based algorithms to simulate, in a more primitive form the human process of cognition.

Multi Layer Perceptrons or MLPs were revived with the Neocognitron [14] and the implementation of the back propagation algorithms for neural networks with more than one hidden layers[49]. Backpropagation is defined as the propagation of the first derivative of the loss with respect to the neurons in the previous layers which are carried backward from the output layer to the input layer to tune the weight and bias parameters[24].

MLPs are fully connected networks. In a MLP, each neuron in a layer is connected to all neurons in the next layer. This leads to a multitude of parameters which are hard to train and result in over fitting. Overfitting is a phenomenon characterised by the computer based algorithm learning redundant features, which are present in the training set but does not help in obtaining a similar or improved accuracy in the testing test[55]. This is often observed when the training accuracy observes an improvement in its score but the validation accuracy observes a deterioration or stagnation in performance. Validation dataset is a subset of the training dataset which is used to evaluate the performance of the network while training, the testing dataset is different from the validation dataset and not used to guide the network while training.

The Neocognitron introduced a form of visual perception based learning where the neuron had a hierarchical arrangement and took advantage of the grouping of data from a simpler structural level to a more complex level of patterns. Instead of having each neural unit connecting to its predecessors and successors, the receptive field of the neural unit was used

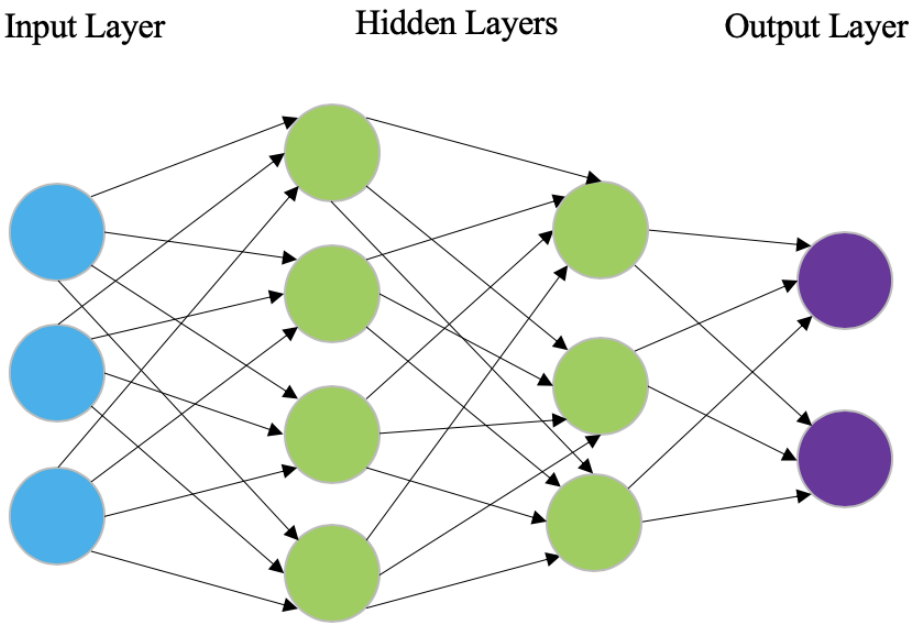


Figure 2.3. A Multi Layer Perceptron with two hidden layers. There are 3 neurons in the input layer, 4 neurons in the first hidden layer, 3 neurons in the second hidden layer and 2 neurons in the output layer.

to determine the connectivity of a single neuron. This form of neural network was called a Convolutional Neural Network or CNNs[27].

In a Convolutional Neural Network the model is distributed between multiple stages of incremental learning[25]. Each stage in a CNN comprises of a combination of convolutional layers, non-linear processing units and subsampling layers. The convolutional kernel defined performs multiple transformations which help uncover crucial features from internally correlated data points within the receptive field of the kernel[23].

The output is then passed through an activation function in the non-linear processing unit. The activation functions helps generate different non-linear patterns depending on the responses from the activation functions. This facilitates the learning of a variety of semantic differences and similarities in the image. The use of an activation function helps introduce non-linearities in the feature space, which ultimately help in learning abstract patterns. Ultimately the output of the non-linear processing unit is fed into the sub-sampling layer. This is also known as the pooling layer.

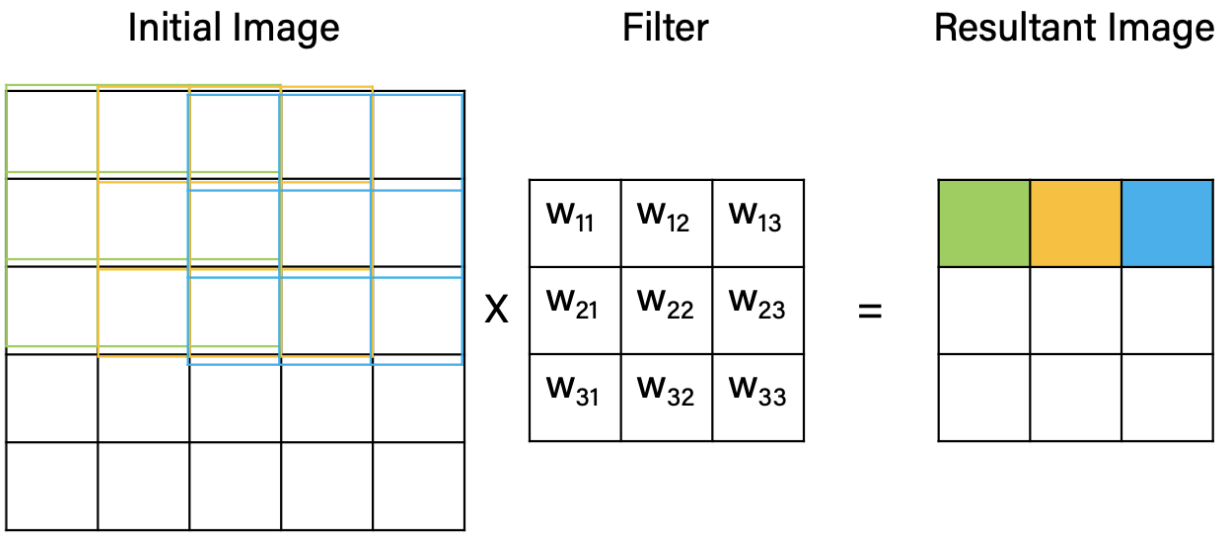


Figure 2.4. A Convolution defined over a 5 x 5 initial image with a filter of size 3 x 3. The filter/kernel slides over the image and the resultant of the transformation is produced in the output image. The filter sliding over the first row is demonstrated in this figure.

The introduction of pooling after the inclusion of non-linearities is to consolidate the results and make the resultant output space, location and orientation invariant. It also helps prevent geometrical distortions from causing discrepancies in the output[51]. These layers are then repeated in a variety of combinations as best found to meet the requirement of the task. This is followed by a single dimensional representation of the output neurons and their mapping onto a fully connected neural network to produce the final outcome. This structure of a CNN facilitates learning from the input fed while exploiting the spatial and temporal correlation present in data.

Due to the hierarchical arrangement of the layers in a Convolutional Neural Network, the human visual perception system is simulated to provide a receptive field based learning structure. This imbibes in it an internal feature extractor ability and helps the network easily learn the internal representations from the raw pixels of the image. The introduction of weight sharing between neurons in a convolutional layer reduces the high resource requirements that were present in a multi layer perceptron.

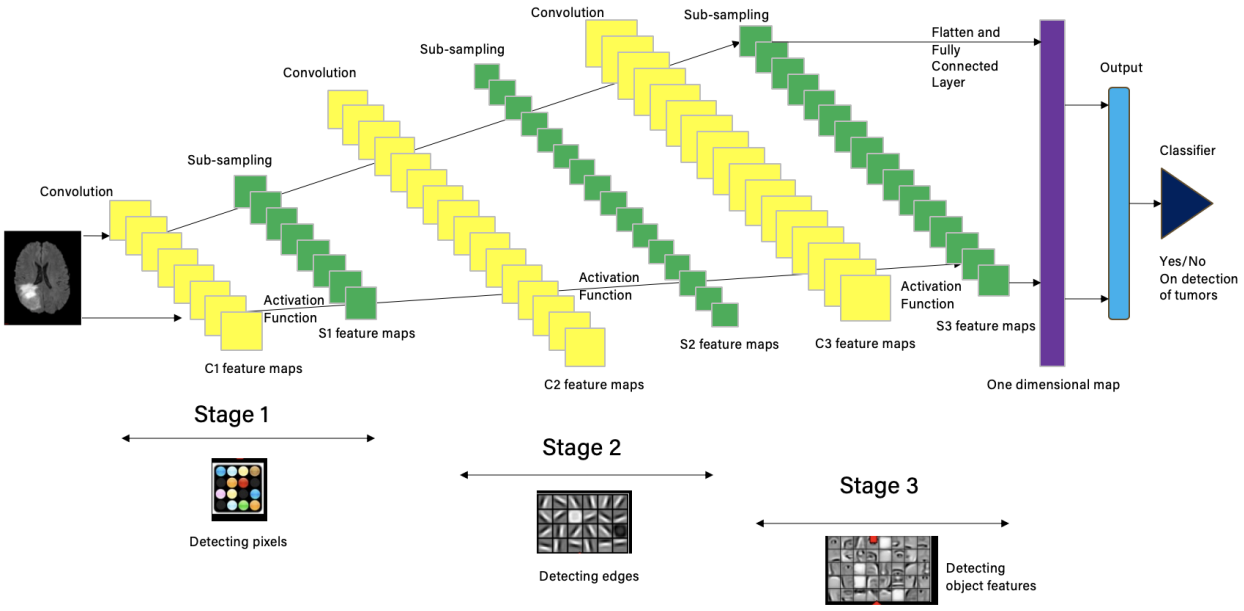


Figure 2.5. A typical Convolution Neural Network Structure. A single channel brain MR image is input through a three stage CNN with a hierarchical convolution, activation and sub-sampling structure. In each stage there is a Convolution feature map generated, referred to as C1, C2 and C3 respectively, these feature maps are then passed through an activation function introducing non-linearity in the feature space and is followed by a sub sampling feature map, referred to as S1, S2 and S3 respectively. The sub-sampled feature map from the third stage is flatted into a one dimensional feature array and connected to a fully connected neural network. This results in a binary classifier which detects the presence of tumor in the input image. Here, Stage 1 detects patterns in the pixel range, Stage 2 detects edges and Stage 3 performs object recognition. This mimics the human receptive field perception system.

This deep learning model provided a means to visualize brain tumor segmentation from simple structures to complex tumor patterns. The ability of CNNs to identify patterns which are location and orientation invariant, help promote research in Medical Image Processing with considerable progress. We talk about some of the convolution based attempts at addressing this problem in the next section.

2.2 Using Artificial Intelligence for Brain Tumor Segmentation

Artificial Intelligence and Machine Learning are two terms that are often interchangeably used in the world today. While Artificial Intelligence or AI signifies the creation of intelligent machines which can ideally simulate human intelligence, machine learning is used to designate the learning performed by computer based software or machines for the purpose of being able to perform the task independently after a certain training period. This is congruent to the human learning system, where a child or novice receives training and uses that in future endeavours. This also holds true in the case of medical practitioners who are trained to distinguish between the various layers in a brain tumor and delineate them pre surgery or evaluate their outcome post surgery. The task here has been to develop an automatic brain tumor segmentation system which can automate the delineation of the various regions of interest and help in pre and post surgery decision making. The conclusion to perform surgical removal of the tumor region may be influenced by the output of such an algorithm.

The prevalent automatic brain tumor segmentation models can be broadly classifies as either generative or discriminative model based approaches[36][37]. There are benefits in both methods. For a generative model based method, the network requires a prior probability distribution of the image atlas[19]. Here the output of such a model is usually an outline detection task which generates regions of similarity, in this situation- tumor density and texture[2]. This ultimately highlights the internal pixel correlations in an MR image which correspond to a pre-specified distribution. In a discriminative model based approach, the task of brain tumor segmentation is viewed in a more classification of the tissue layer fashion. Here, each pixel is categorized as being with tumor or without tumor or with a variety of densities which signify the tumor heterogeneity. These algorithm depend on the tumor feature and the accuracy of the classification algorithm and have traditionally utilized machine learning algorithms like support vector machines[48][29][34], random forests [53][42]. These algorithms have largely depended on the introduction of non linearities in the feature space and manual transformations to the MR brain images. A major approach used here has been the identification of tumor regions or a voxel/pixel wise classification of tumor and

their extraction using the internal correlations between the post transformation structures. This has been the blueprint of research along these lines.

The existence of a collection of brain MR image data with standardization techniques which could form an uniform basis of comparison between MRI from different scanners and units of measurement helped provide a large enough data set which ultimately facilitated deep learning techniques to be applied in this realm of medical image processing. Convolutional Neural Networks, being a major part of modern deep learning research, were deployed in the Multimodal Brain Tumor Segmentation (BraTS) challenges[61][20][56][41][1] since the year 2014. There is an increasing body of research on computational algorithms which focus on the task of brain tumor segmentation. However, the lack of public access datasets to design and test the validity of such research is the roadblock in the process. For private datasets it becomes harder to compare segmentation techniques between pre-existing and newly developed approaches. The imaging techniques, machine specification, image modalities and conditions of imaging make it difficult to bring about a uniform scale of comparison. The other factors which deter this process are the type of the tumor- which could be high grade glioma (glioblastoma multiforme) or low grade tumors, primary or secondary tumor or the texture and heterogeneity of the tumor regions influence the classification accuracy. In addition to this the instance in time when the scan was taken determines the outcome. In other words the presence of post surgical cavities, radiotherapy effects and the age of both the patient and the tumor determine the effectiveness of these machine learning algorithms. BraTS has developed one of the largest datasets of brain MR data accompanied by the expert delineation of the relevant tumor sub-regions.

The introduction of convolutional neural networks in the field of medical image processing led to a significant improvement in result statistics. These models outperformed the prior set state of the art performance metrics of other traditional machine learning algorithms. The use of CNNs in the field of image segmentation based on semantics[26], image classification[15] and detection of regions of interest or objects[32], as relevant to the research problem, provided developers the inspiration to apply them in brain tumor sub region detection and delineation.

Since then, the commonly used dataset for brain tumor segmentation has been the open access BraTS database. We have developed our proposed algorithm, which is described in Chapter 3, using this dataset.

The three dimensional nature of brain MR Images allows for the use of a variety of CNN structures to implement the task of brain tumor segmentation. As with any other volumetric image, the relationship between each voxel is defined over its immediate neighborhood. Exploiting this data to advantage progress in research has led to the development of three dimensional and two dimensional approaches to address this task. The increase in resource requirements for any standard 2D kernels/filters, as demonstrated in Fig. 2.5, is determined by the quantity of filters in each convolutional layer and the number of such layers in the model. The convolution function is defined for a 2D slice of the data as demonstrated in Fig. 2.4. For applications in 3D, utilizing the full spread of the volumetric data, 3D kernels could be deployed, with 3 dimensional feature maps and their corresponding sub sampling and mapping. Here the resource consumption is considerably more as the number of parameters are linearly increased for each additional layer of convolution and multiplied for each such layer with the units of the input in the third dimension.

Networks that use 3D kernels[35][11][12][9] over 2D[59][40][21] require a larger resource pool, is more expensive to train and require the creation of a larger network model. Even though this kind of a structure, can potentially take complete advantage of the 3D MR image, the resource constraints often force researchers to opt for the 2D approach. The use of a plethora of network structures, including the convolutional and fully connected neural network approach, has led to a wide variety of structural modifications and layers combinations to be produced.

Most approaches with a 2D Convolutional Neural Network, break down a brain MR Image into 2 dimensional patches or slices of the original data. This makes the network treat each individual slice as an independent entity. The task of the neural network is then made into a classification problem where each patch is now categorized as either a tumor core, enhancing tumor or a whole tumor region. Depending on the weighted approximate of all overlapping slices in a central two dimensional patch region, the final classification is made and a segmentation label is associated with the region. The drawback of such a two

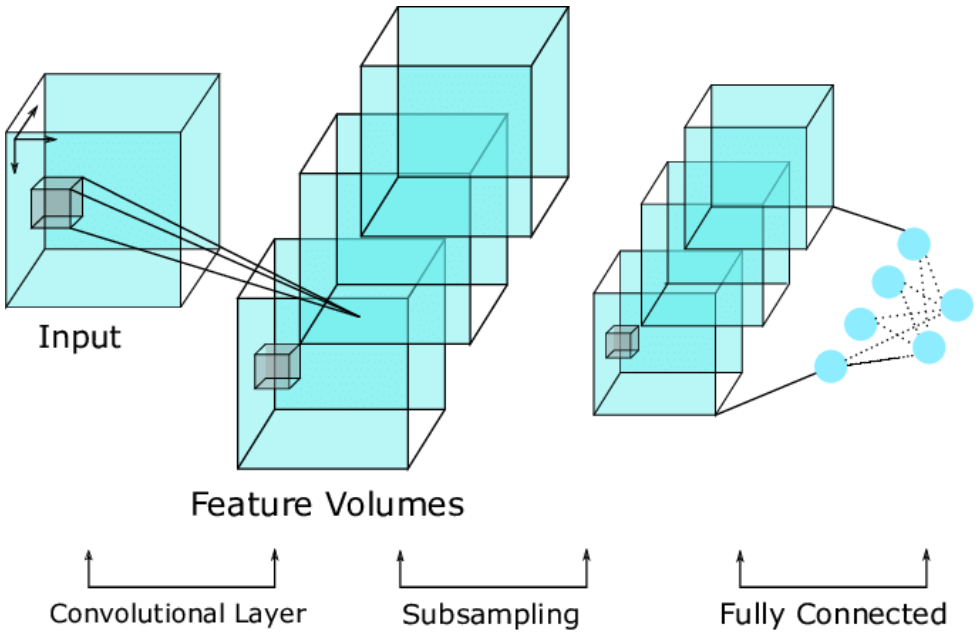


Figure 2.6. A Convolution Neural Network catering to a 3 D input and a 3 D feature map. The convolutional, non linear and subsampling layers are followed by a mapping to a fully connected neural network. As visible in this figure, there is a multiplicative increase in the number of network parameters due to the addition of a third dimension for each feature map and input.

dimensional approach lies in the treatment of each slice as an independent entity. In reality the tumor spreads in a three dimensional spectrum and the exclusion of the two other axes for tumor sub region classification leads to the loss of crucial information.

Network structural modifications have been introduced in many research problems as a means to supplement this loss. Some of them are particularly observed in the realm of medical image analysis techniques like knee cartilage segmentation[43]. Here, the authors try to break up the learning in three specific planes and weight them with the use of a softmax classification layer[31].

Literature[13][52][58] provides instances such as the use of orthogonal patches with information on the texture and pattern of the medical data. This is finally brought together by the fully convoluted pathway of the CNN [13]. Research in the field of Lung Computed Tomography (CT) images, especially in the detection of pulmonary nodules for the purpose of segmentation [52] [58] display the use of angular patches and the training of multiple

CNNs corresponding to these angles and their amalgamation for the final segmentation task has opened up the realm of possibilities when it comes to the use of modified 2D CNNs for segmentation tasks in the medical image analysis spectrum.

While various machine learning techniques have come into play in the world of data processing, the scope of their algorithmic applicability are limited by the data supplied. In other words, data pre-processing plays an important role in determining how successful/accurate our classifier can be. This is true for textual data, signals, images, videos and any other form of data used in such tasks. In the case of medical images, especially when the data is sourced from different scanners or machines and subjected to noise in image acquisition- this task becomes harder.

One way to mitigate this discrepancy is to normalize the data using standard normalization techniques which include subtracting the mean of each volume and dividing by the standard deviation. However, in the case of brain MR images, the intensities in the image volume vary depending on the brain tissue being referenced for standardization purposes. To simplify that, the mean of the intensities of different MRI volumes correspond to different areas of the brain depending on external exposure, intensities of surrounding regions and noise. Thus, a simple normalization procedure as defined above, is not sufficient for high accuracy image segmentation tasks.

Our proposed algorithm cuts across both the spheres of generative and discriminative properties and offers an image atlas based approach while keeping the distinct features of texture and heterogeneity in consideration when performing the task of image segmentation. We also implement the transformation of raw MR image data into the Lab or CIELAB color space, which uses intensity and illumination and facilitates normalization of image data for the task of brain tumor sub region segmentation. More detail has been provided in Chapter 3.

3. EXPERIMENTATION

In this chapter, we use the Multi-modal MRI brain images from the Brain Tumor Segmentation Challenge (BraTS)[36][3] [6][4][5] for the task of brain tumor segmentation using Conditional Generative Adversarial Networks (GANs)[18] [38][22]. This involves the segmentation of multiple Magnetic Resonance Image(MRI) channels into pre-defined brain tumor sub regions. The tumor sub regions are manually segmented by certified radiologists following a uniform labelling protocol. This forms the ground truth in our segmentation task.

3.1 Image Segmentation Methodologies

The progress in research on the use of Convolutional Neural Networks(CNNs), have led to them being widely implemented in a variety of image prediction and segmentation problems. Of particular interest is the problem of image labelling in a multi channel MR image. The use of multi channel magnetic resonance imaging enhances the richness of representation in each voxel. The labelling task now, is to use the information available to classify the voxel representation as whole tumor, tumor core or enhancing tumor regions. Our methodology relies on the fundamental properties of each individual channel conveying critical information on tumor density, texture, shape and extent of spread. All of these properties, combined helps in the creation of an independent loss metric but with a conditioned segmentation map.

3.1.1 Convolutional Neural Networks

The main role of a CNN is to minimize the loss function, which ultimately determines the quality of the result. In a problem of this nature, the loss is denoted by similarity metrics between the ground truth and the generated segmentation. Here, even though the training is automatic, we still have a substantial amount of control over the designation of the loss function. This results in the model learning to minimize that which the human developer heuristically determines to be the best representation of the task at hand. This has been seen to work well with most image prediction tasks. However, when it comes to segmentation, our

focus no longer lies in the realm of summing up the distance/difference between the number of pixels representing the ground truth and the generated image as shown in Equation 3.1

$$d\left(GroundTruth, GeneratedImage\right) = \sqrt{\sum_{\mathrm{i}=1}^{n}\sum_{\mathrm{j}=1}^{n}\left(GroundTruth_{\mathrm{i,j}} - GeneratedImage_{\mathrm{i,j}}\right)^2} \quad (3.1)$$

It becomes a binary match/mismatch between labels for each voxel in the image. The input images for this problem are three dimensional, hence, each unit is referred to as a voxel.

Generalising a loss such as the Euclidean distance function (Equation 3.1) between voxels, result in blurring or an average smoothing of the area. If we want the output to be sharp, similar to the ground truth, and resilient to scanner noise then our task benefits from the specification of a more abstract goal. We propose that the network learns the higher level goal of optimizing the similarity of the generated segmentation to the given ground truth. Here, we no longer specify the loss function but rather ask the network to both devise and minimize the loss. This uniquely helps us devise a spatially independent loss metric for evaluation in the case of brain MR images.

3.1.2 Generative Adversarial Networks

Generative Adversarial Networks(GANs) have the ability of performing this task by training a generator-discriminator combination of neural networks[10][44] [50]. GANs work on minimizing the generator loss while simultaneously training a discriminator to distinguish between the actual and generated ground truth. Their advantage lies in the ability of GANs to be able to customize the loss function depending on the training set provided.

In this kind of network[18][38], there is a generator G and a discriminator D. The generative model tries to capture the data distribution while the discriminator model tries to differentiate between the output from the generator and the actual ground truth/real image provided. This is demonstrated in Figure 3.1

The distribution of data provided to the generator p_g is learnt over the input x. Here, the generator G utilizes a prior random noise vector $p_z(z)$ to form a mapping to the generated

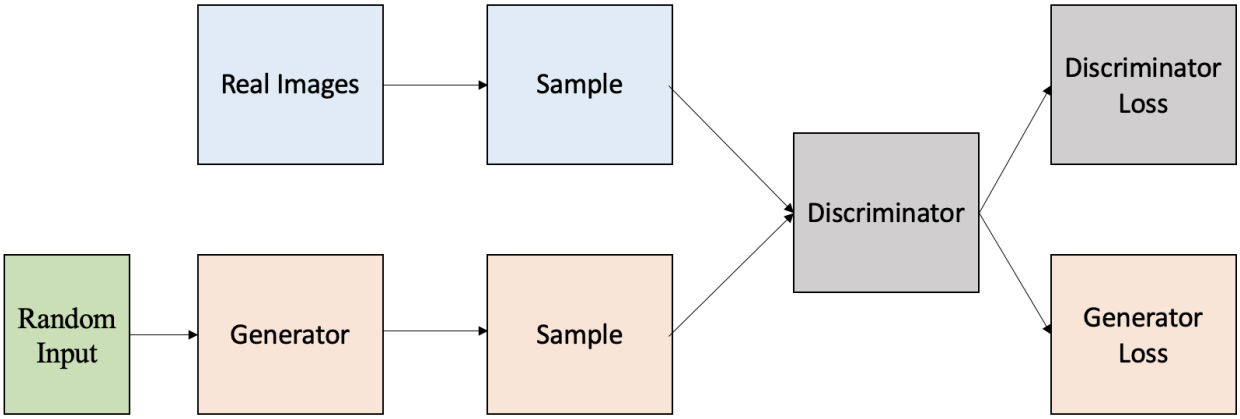


Figure 3.1. A Generative Adversarial Network with a generator G and a discriminator D. The real images and the generated images are sampled and passed into the discriminator D. As training progresses the task of the network is to get better at differentiating between the real image and the generated image, while simultaneously training the generator G to produce more real like generated images.

output $G(z;\theta_g)$, where z is the random input vector/noise and θ_g represents the parameters being optimized by the generator network. While in the case of the discriminator D, the network learns whether the sample belongs to the data x or is a product of the generator distribution p_g. This is learnt over $D(x;\theta_d)$, where θ_d represents the parameters of the model being optimized by the discriminator network.

G and D are trained together, resulting in both discriminator and generator loss. These losses are used to optimize the parameters of the adversarial model. The generator network G optimizes θ_g using the loss described by $\log(1-D(G(z)))$. At the same time the discriminator network D optimizes θ_d utilizing the loss denoted by $\log(D(x))$. Finally, the entire network can be denoted by $A(G,D)$, where the role of the entire model is represented by Equation 3.2

$$min_G max_D A(D,G) = E_{x\sim p_{data(x)}}[\log(D(x))] + E_{z\sim p_{z_{(z)}}}[\log(1-D(G(z)))] \quad (3.2)$$

3.1.3 Conditional Generative Adversarial Networks

Where a GAN learns a generative model of data, conditional generative adversarial networks[18] [38][22] learn from a conditioned prior image/input. The output of the generator is based off this input image and an additional label or extra modal information k. Here, the basis of the generator network G is biased to produce output that is based off the input distribution $p_z(z)$ and the extra information k.

This joint distribution of $p_z(z)$ and k work together to compose a hidden representation which optimizes the generator G model parameters θ_g. The discriminator D learns to optimize its parameters θ_d from the conditionally generated images and the real image. This entire network can be represented as $A(G, D)$ by the Equation 3.3

$$min_G max_D A(D, G) = E_{x \sim p_{data(x)}}[\log(D(x|k))] + E_{z \sim p_{z_{(z)}}}[\log(1 - D(G(z|k)))] \quad (3.3)$$

Figure 3.2 is an accurate representation of the conditional generative adversarial network. Here the prior image/label provided acts as the bias in the production of conditioned intermediate generated image. The loss function for the generator $G$is modified to include this additional information k as $\log(1 - D(G(z|k)))$. The discriminator D, also has its loss function modified to $\log(D(x|k))$. This is ultimately used to update and optimize its parameters θ_d.

3.2 Dataset

The dataset contains 210 high grade glioma (HGG) and 75 low grade glioma (LGG). The gliomas are tumors in the brain as per Fig. 1.1 All of these images are used for training and 66 unlabelled test cases are used for validation. The dataset contains multiple channel views of the MRI multi modal imafes. In particular, we have the MR image channels for T1C, T1, T2, and FLAIR. Each of these volumes are of dimensions $155 * 240 * 240$ resolution. These images are first meticulously oriented and based on the framework of the anatomical template, it is then skull stripped, and scaled to $1mm^3$ voxel resolution.

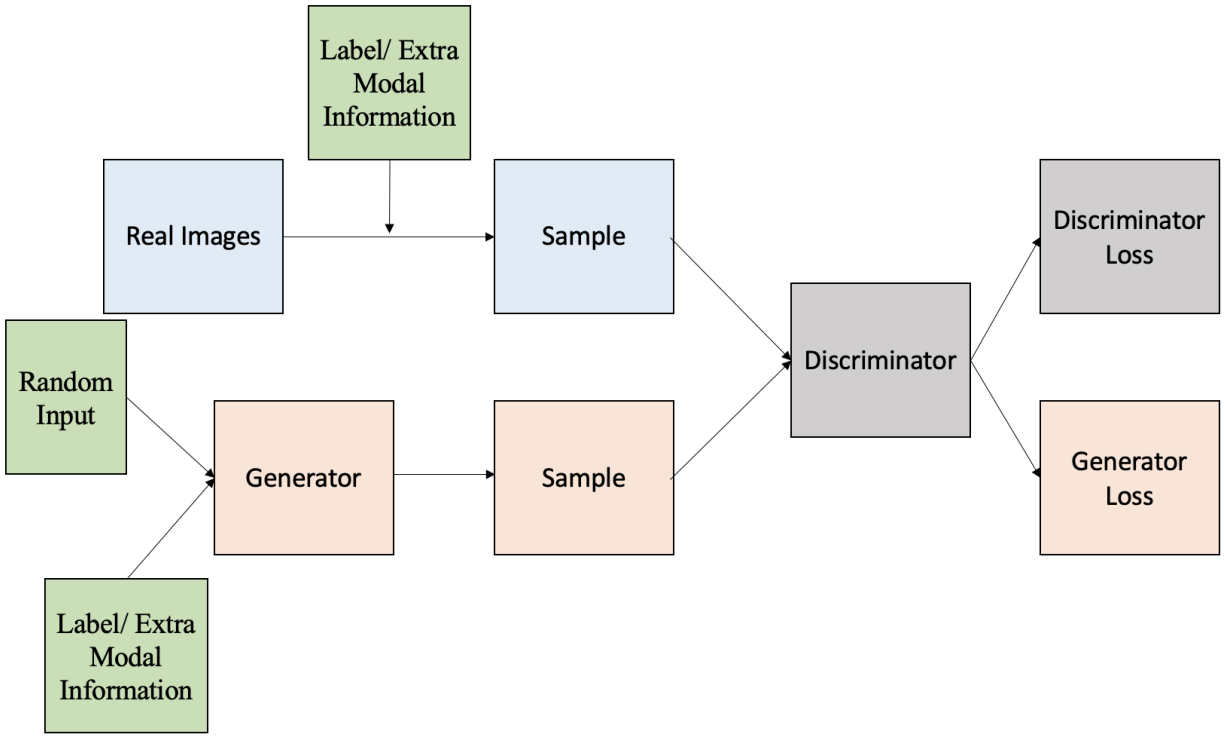


Figure 3.2. A Conditional Generative Adversarial Network with a generator G and a discriminator D. The real images and the conditionally generated images, along with extra modal information/labels are sampled and passed into the discriminator D. As training progresses the task of the network is to get better at differentiating between the real image and the conditionally generated image, while simultaneously training the generator G to produce more real like conditionally generated images.

Manual segmentation of the tumor sub-regions is done by experts, following a uniform annotation protocol across all patients. The protocols used here were revised and approved by medically certified neuro-radiologists. In the end, the predicted labels are evaluated by merging three regions viz. whole tumor (WT:NCR/NE + ET + ED), tumor core (TC:NCR/NET + ET), and enhancing tumor (ET) as shown in Figure 1.1. The NCR region represents the necrotic-cystic core, the NET signifies the non enhancing tumor region and the ED represents the edema region. The conglomeration and labelling of these regions together into the whole tumor, tumor core and enhancing tumor regions are done to benefit the medical practitioner operating on a human being with a brain tumor of such nature.

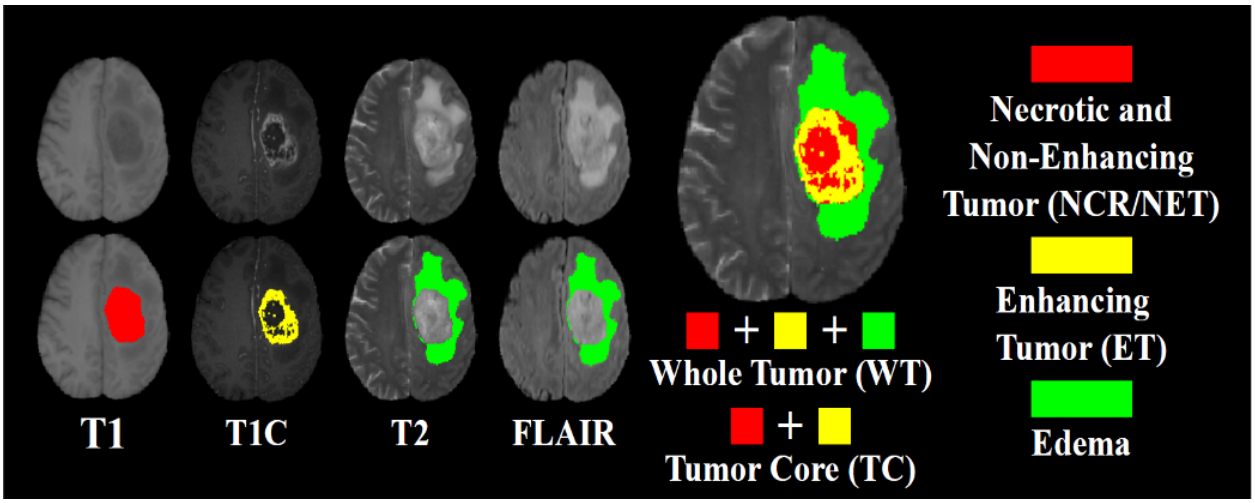


Figure 3.3. Multimodal 2-dimensional MR image slices of the human brain post processing and labelling. The predicted labels are evaluated by merging three regions viz. whole tumor (WT:NCR/NE + ET + ED), tumor core (TC:NCR/NET + ET), and enhancing tumor (ET) as shown.

3.3 Methodology

The automatic segmentation of brain tumor sub regions from medical images is important for treatment planning, prognosis and diagnosis for a person suffering with brain gliomas. This serves as an additional aid to the medical practitioner as they make a decision on the next step in patient outcome evaluation. With the increase in the use of convolutional neural networks in solving such segmentation tasks, there has been an increased reliance on 2 dimensional approaches, the reason behind this trend is the resource constraints and limitations in computation power. The 3 dimensional nature of the data warrants a 3 dimensional approach, such that can exploit the spatial relationship between voxels but also be memory efficient.

We propose a multiplanar conditional generative adversarial network (MPcGAN) to segment the brain tumors into spatial sub regions. Our generator follows the structure of a U-Net[45] with multiple skip connections[57] between different stages of feature extraction, more details are provided in the subsection on network design. The novel design of our skip

generator supplies rudimentary information from the initial layers of the network to the later refined feature extraction layers.

We also propose a loss metric using the dice loss[54] and binary cross entropy loss[60]. The discriminator works on a patch based evaluation metric, where we follow a PatchGAN[28] mechanism which penalises structures at a local level of style statistic.

Finally, our proposed approach treats the 3 dimensional images as contiguous slices of 2 dimensional data in the xy plane, yz plane and xz plane. We refer to these as the axial plane, coronal plane and the sagittal plane respectively. Our algorithm demonstrates top performance while treating each sequence of brain images as a sequence of 2 dimensional contiguous brain image slices. The combined generated segmentation map of these multi-planar sequences of brain images are bagged[8]. The generator map after bagging is then used to formulate the final segmentation map. This is further used to calculate and back propagate the dice and binary cross entropy loss, which finally updated the optimization parameters of the segmenter- discriminator model.

3.3.1 Pre-Processing

As discussed in the Introduction, various brain tumor regions differ by density, texture, composition and structure. This results in different T1, T2 and FLAIR intensity values output through a magnetic resonance(MR) imaging scanner. These intensities are used to locate tumor core, edema, enhancing tumor and other necrotic regions.

Color images are often constructed from a sequence of images which are layered in accordance to the color channels of red, green and blue (RGB). These can be disintegrated into three gray scale images, in 3! ways and re-integrated back as a color/RGB image. As the color information does not directly correspond to the color of the imaged tissue, the combination of channels used for composing the RGB combination is not detrimental to the color information of the imaged tissue, tumor or edema regions.

In this particular problem statement we use the T1C, T2 and FLAIR MR image channels in the construction of our RGB channels. The omittance of the T1 channel from the BraTS images is attributed to the lack of extra information between the T1C and T1 channels. In

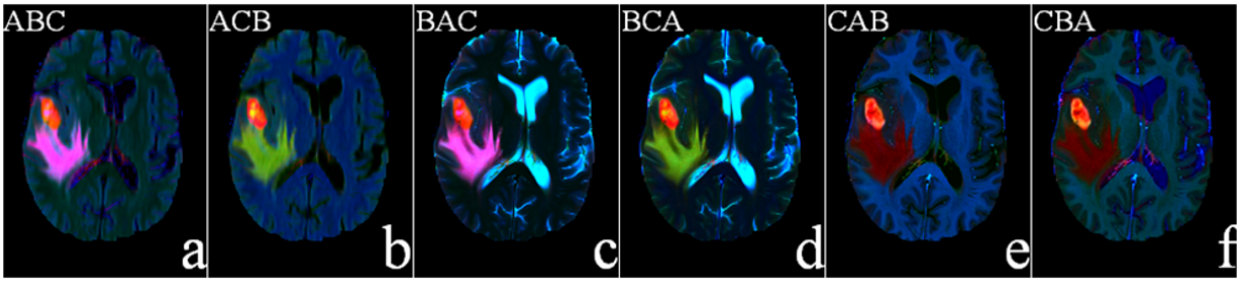
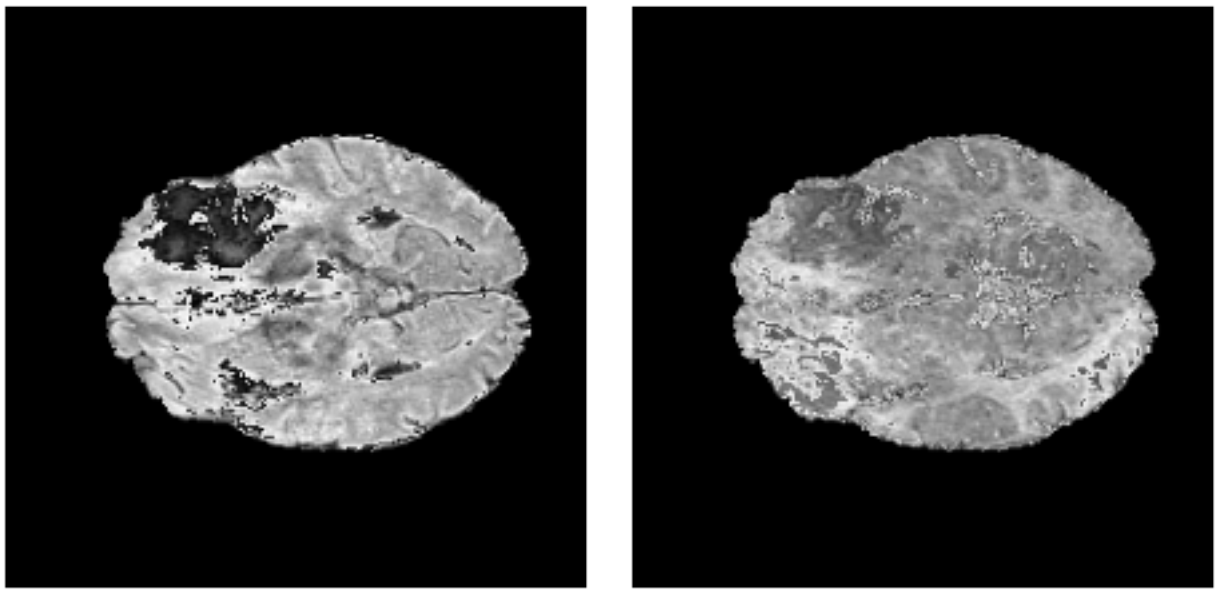
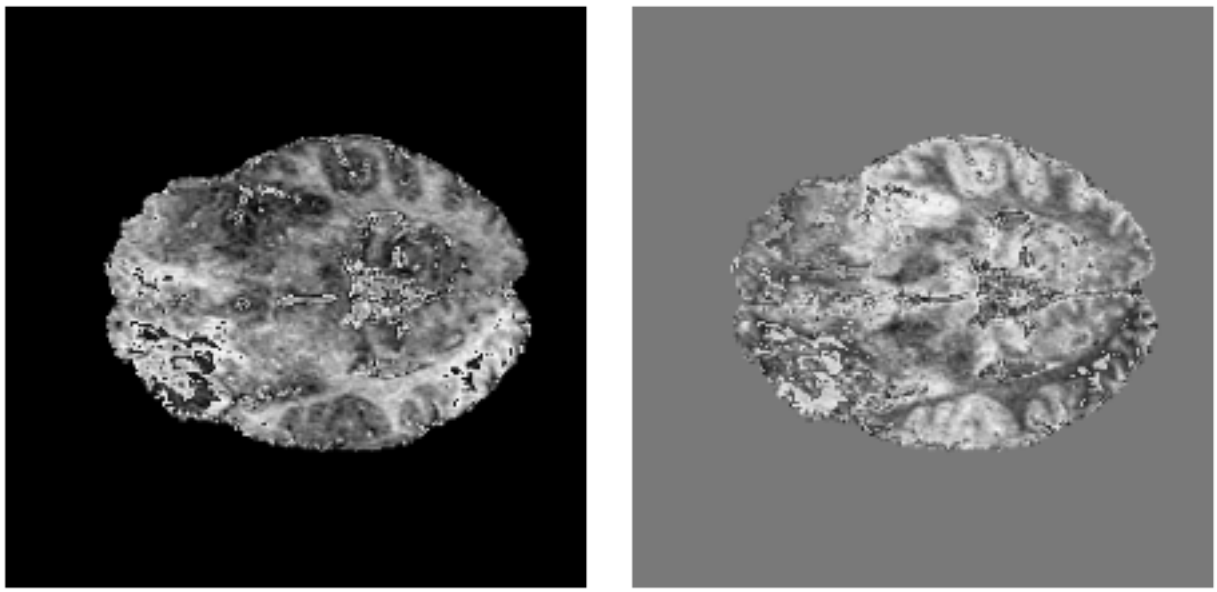


Figure 3.4. Three gray-scale images A, B, C, are permuted over the Red, Green and Blue channels to produce combinations of FLAIR, T1C and T2 channels. Each combination is transformed to the CIE- $L*a*b*$ format and displayed between option a to f. The six combinations are ABC, ACB, BAC, BCA, CAB, CBA. Of these we can use wither option a or option b. They have distinctly highlighted brain tumor regions and the rest of the brain matter is considerably suppressed.



(a) (b)

Figure 3.5. On the left we have the Red channel which has the FLAIR MR image. There are non tumor regions in this image which have the same intensity as the tumor region. On the right we have the $L*$ channel of the CIE $L*a*b*$ converted FLAIR image. Although the whole image is more uniformly spread (due to normalization), the tumor region still strikingly stands out from its Red channel gray image equivalent.



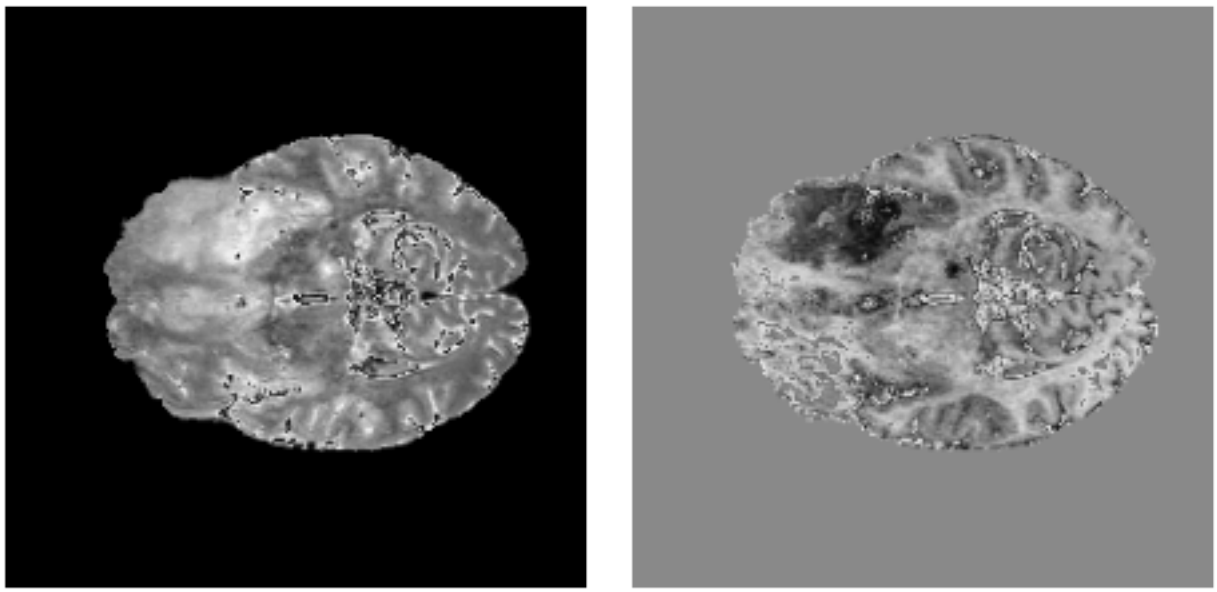
(a) (b)

Figure 3.6. The gray scale single channel equivalent of the Green channel is presented here. On the left we have the Green channel which contains the T1C MR image. There are non tumor regions in this image which have the same intensity as the tumor region. On the right we have the $a*$ channel of the CIE $L*a*b*$ converted T1C image. Although the whole image is more uniformly spread (due to normalization), the tumor region still strikingly stands out from its Green channel gray image equivalent.

our experiments with our network, the resultant segmentation map does not significantly vary with the addition of a fourth channel. However, our network complexity and resource consumption does increase considerably with a fourth image channel. For this reason, we move ahead with T1C instead of both T1 and T1C.

Of the six different ways to combine the T1C, T2 and FLAIR channel into a RGB color image spectrum, there are only two which has the tumor region distinctly differentiated from the rest of the brain region. This is a desirable outcome, as we do not want our model to annotate brain matter as tumor regions, especially in an image as complex and detailed as that of the human brain. A side by side comparison is provided in Figures **??**, **??** and **??**.

Then, we transform the RGB image to the CIE—$L*a*b*$ color space to bring aboout a higher level of contrast between the brain matter and the glioma. Converting an image into $L*a*b*$ from RGB, results in the separation between the two layers, luminosity and



(a) (b)

Figure 3.7. The gray scale single channel equivalent of the Blue channel is presented here. On the left we have the Blue channel which contains the T2 MR image. There are non tumor regions in this image which have the same intensity as the tumor region. On the right we have the $b*$ channel of the CIE $L*a*b*$ converted T1C image. Although the whole image is more uniformly spread (due to normalization), the tumor region still strikingly stands out from its Blue channel gray image equivalent.

chromaticity. Figure **??** displays the six combinations of Red, Blue and Green channels with placeholders A, B and C. The A, B, C combination is permuted over the FLAIR, T1C and T2 channels. Of the six combinations we notice the option a and option b work best to our advantage. The tumor region is distinctly separable from the brain image and benefits our segmentation task. Additionally, the chromaticity and luminosity transfers the image into a independent color space than that of RGB. This leads to differential voxel intensity values and supplements our task of brain tumor sub region segmentation.

The processed input image is then normalized and broken down into two dimensional slices along the axial, coronal and the sagittal planes for feeding into the generator/segmenter network for segmentation purposes. The three planes being considered are displayed in Figure 3.8.

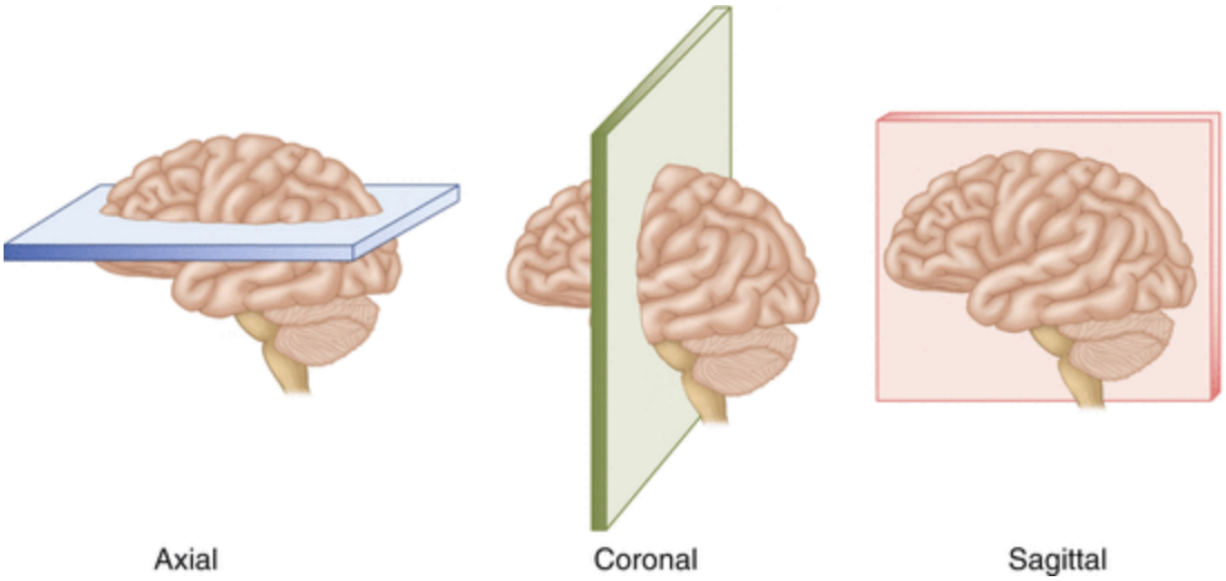


Figure 3.8. The axial view of the brain is a xy planar view and allows us to maintain the spatial connectivity in brain tumor and brain matter in the xy plane. On the other hand, the coronal view of the brain is a yz planar view. It maintains the spatial connectivity between brain matter and tumor in yz plane. The sagittal plane deals with the xz plane and helps us in our task by maintaining bran matter and tumor connectivity in the xz plane.

3.3.2 Network Design

The MICCAI (Medical Image Computing and Computer-Assisted Intervention) BraTS data comprises of both High-Grade gliomas and Low-Grade gliomas, colloquially known as brain tumors. The segmentation of such tumors is often difficult and prone to human error while delineation. With the advent of deep learning-based technologies, this task of segmenting the tumor core (TC), enhancing tumor (ET) and specifying the whole tumor (WT) region from human brain images have been considerably benefited. Every year the MICCAI BraTS challenge releases Magnetic Resonance Imaging (MRI) volumes from four channels T1, T1C, T2 and FLAIR. The segmentation task is to delineate the above TC, ET and WT regions from these MRI scan images.

When feeding three dimensional images or image patches into a deep learning network, the use of a 3 D kernel is required. This induces considerable overhead which slows down

the learning of networks. Again, should we not use the 3D features associated with voxel to voxel mapping, then we tend to lose the properties that influence the tumor region by virtue of its proximity to a particular segment or segment layer. In our proposed model we treat the images as coronal, axis and sagittal planes. While calculating the probability of a pixel belonging to a certain segment, we weight the segmentation performed by each model and deduce our final label.

Our main algorithmic contribution lies in the use of a Generative Adversarial Network (GAN) which is conditioned on the three channels of the MRI scan image. Each slice of a sequence of scans, that are shuffled in the dataset, is treated as a two-dimensional entity. We feed this through a segmenter (more traditionally referred to as a generator in a GAN framework). This is a three-layer down sampling, followed by a three-layer up sampling network. Up sampling refers to the increase in the dimensions of the fed feature map whereas down sampling refers to the reduction in dimensions. Our proposed algorithm uses an increment and reduction factor of multiples of two in each dimension.

Each layer on our segmenter comprises of a convolutional, max-pooling and batch normalization layer. In addition to this, the network has the output of the first layer passed on as the input to the sixth layer, the second layer to the fifth layer and the third to the fourth layer. We also account for additional skip connections as illustrated in Figure 3.13. While the skeleton represents a more traditional U-Net[45] structure, our skips from the third layer is passed on to the fifth and sixth layers. Additionally there is a skip connection from the second to the sixth layer.

A skip connection is a direct pathway from an initial layer to a later layer. This helps mitigate the vanishing gradient problem. Often with multiple layers of a network there is loss of information in the pipeline, this loss can prove detrimental to parameter optimization. We performed multiple experiments with and without these skip connections and found our proposed segmenter model to be the optimal solution when evaluated with keeping other model parameters constant. The use of skip connections between intermediate layers proved to improve the efficiency of our proposed multi planar generative adversarial network. This can be attributed to the complex nature of brain images and the fine features present in its content. Finally, the output of the segmenter is the label map which needs to maximize its

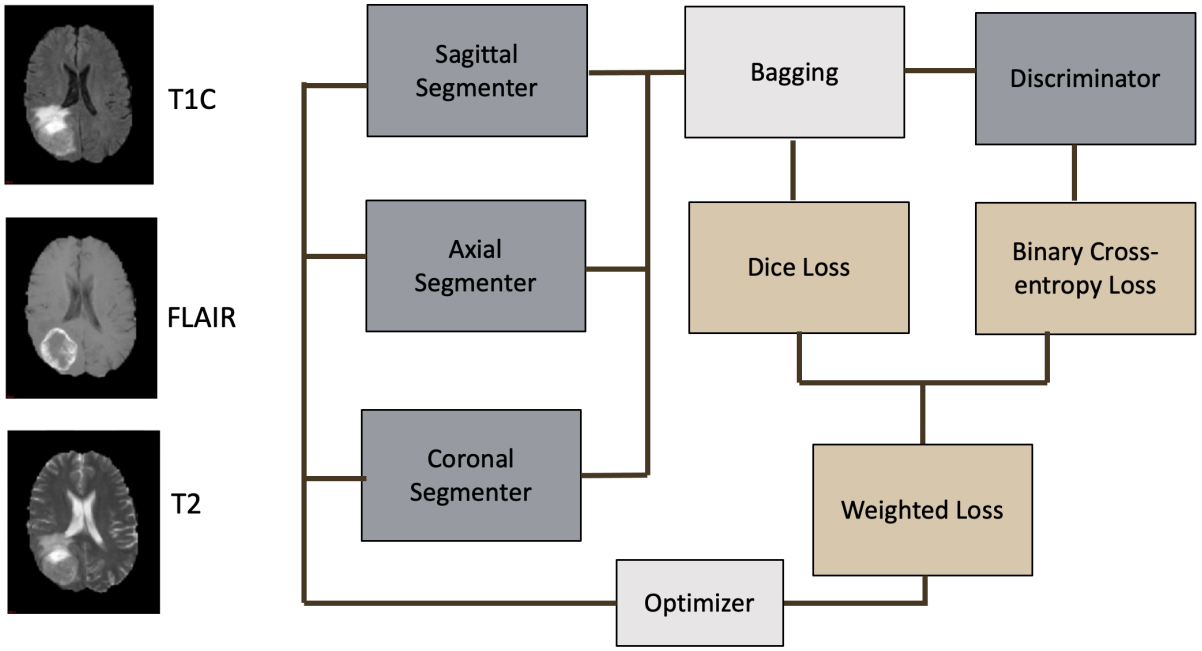


Figure 3.9. In each three dimensional brain MR image, we generate slices in the xy/axial plane, yz/coronal plane, and the xz/sagittal plane. Our proposed algorithm runs this data through three segmenters. The sagittal segmenter is fed image slices in the xz plane, the axial segmenter is fed images in the xy plane and the coronal segmenter has yz planar images supplied to it. The label map across each segmenter is weighted with a 1/3 coefficient. We then compute the final label map/segmented image and compute a dice loss with the ground truth and a binary cross entropy loss in the discriminator. Details of these losses are provided in Section 3.3.3. The combined loss of these metrics is used to optimize the parameters of each segmenter.

similarity to the ground truth (label maps which have been manually segmented into the TC, ET and WT region by a group of doctors specializing in this field).

The design of the network is critical for outcome evaluation. In this section we will focus on the two distinct components of any Generative Adversarial Network model: generator and the discriminator. The generator for this task is the segmenter and is referred to as such for the rest of this section. Pictorially the complete network is displayed in Figure 3.9.

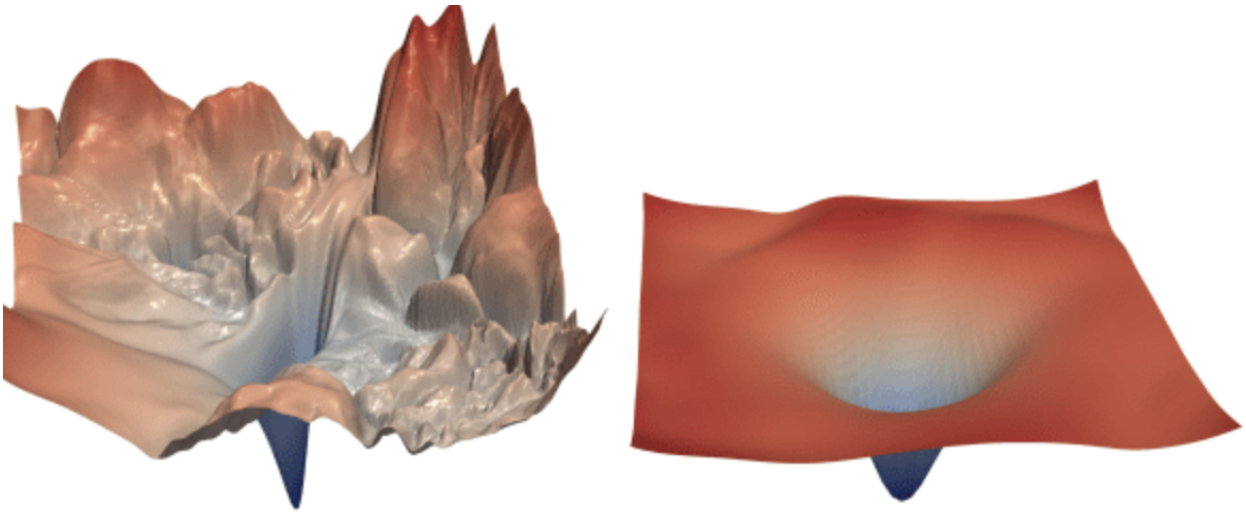


Figure 3.10. On the left we have the loss function being optimised without skip connections. As visible, the surface is undulating and rough. This leads to the gradient descent algorithm being stuck in the local minimas. On the right we have the loss/optimization space being must smoother and significantly reducing the possibility of local minimas. This leads to a more accurate and faster parameter optimization[1]

3.3.2.1 Segmenter/Generator Design

The segmenter is an important part of the proposed multi planar conditional generative adversarial network model. It is responsible for the establishment of connections between pixels in the 2D slices of each image. As the convolution layer learns the features from its rudimentary stages in the initial layer, the network learns to fine tune and form segmentation maps. As the requirement of this segmenter is to input a 256×256 image and produce a 256×256 segmentation map, we decided to move with a autoencoder-decoder [47][7] structure of convolutional neural network.

The autoencoder - decoder framework of a convolutional neural network structure is unique due to a number of factors. Primary of which, is its ability to transform a n dimensional input space to a m dimensional output space. This is especially beneficial for this task of transformation between MR image space and a segmentation label map. Secondarily, the encoder/step down portion of the network works as the analysis wing of the model. Here,

[1]The image has been taken from [30]

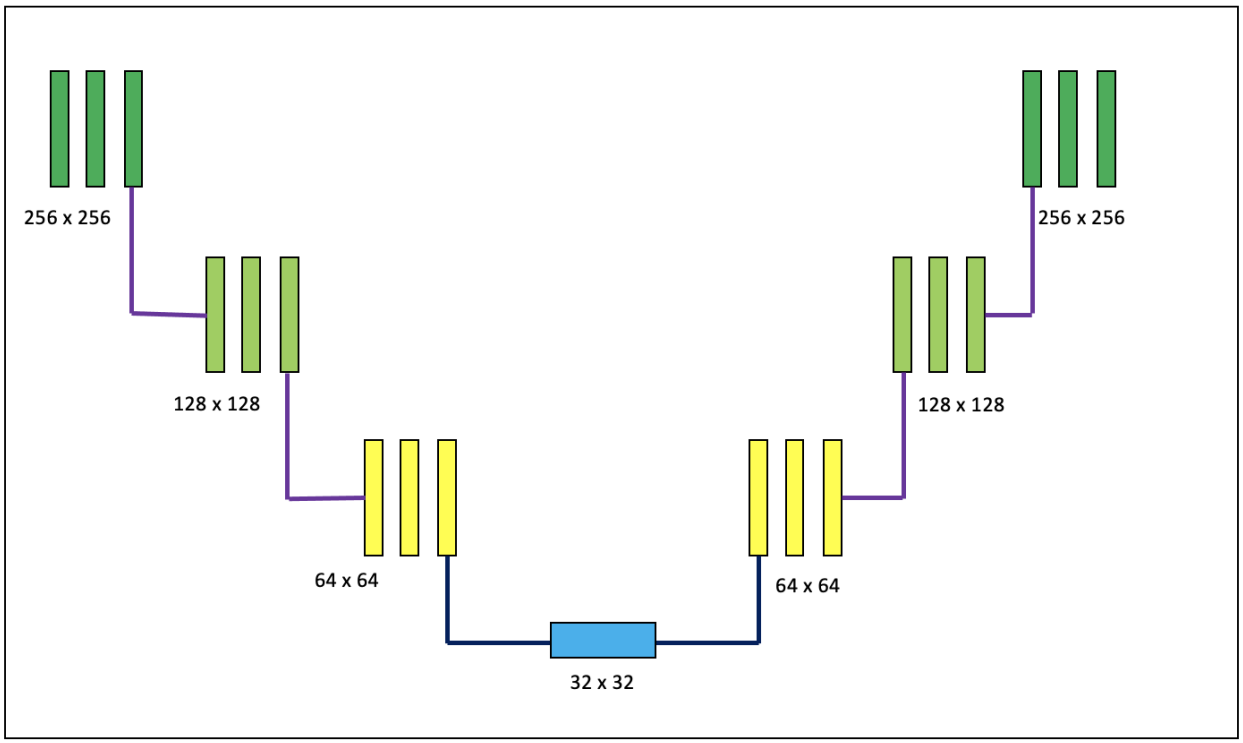


Figure 3.11. Represented here is an autoencoder- decoder network with three convolutional layers in each down-sampling stage of the encoder/analysis network. The input dimensions reduce from 256×256 to 32×32 in the bottleneck component and is then up-sampled in consecutive three convolutional layers to the output layer of 256×256. It can be noted that the dimensions of the input and output layer need not be exactly the same. They are represented to be of similar dimensions to maintain coherence through all figures.

the redundant features, scanner noise, non crucial information for segmentation purposes is filtered through to the lowest level 32×32. This is also referred to as the bottleneck layer and plays an indispensable role in latent space representation.

From the 32×32 image space, we have the decoder or the synthesis portion of the network up-scaling the feature maps, all the way to the 256×256 segmentation label space. This analysis-synthesis combination of the autoencoder-decoder network makes it a suitable tool for image space transformations, filtering out of redundancies, and noise. A skeleton structure is provided in Figure 3.11.

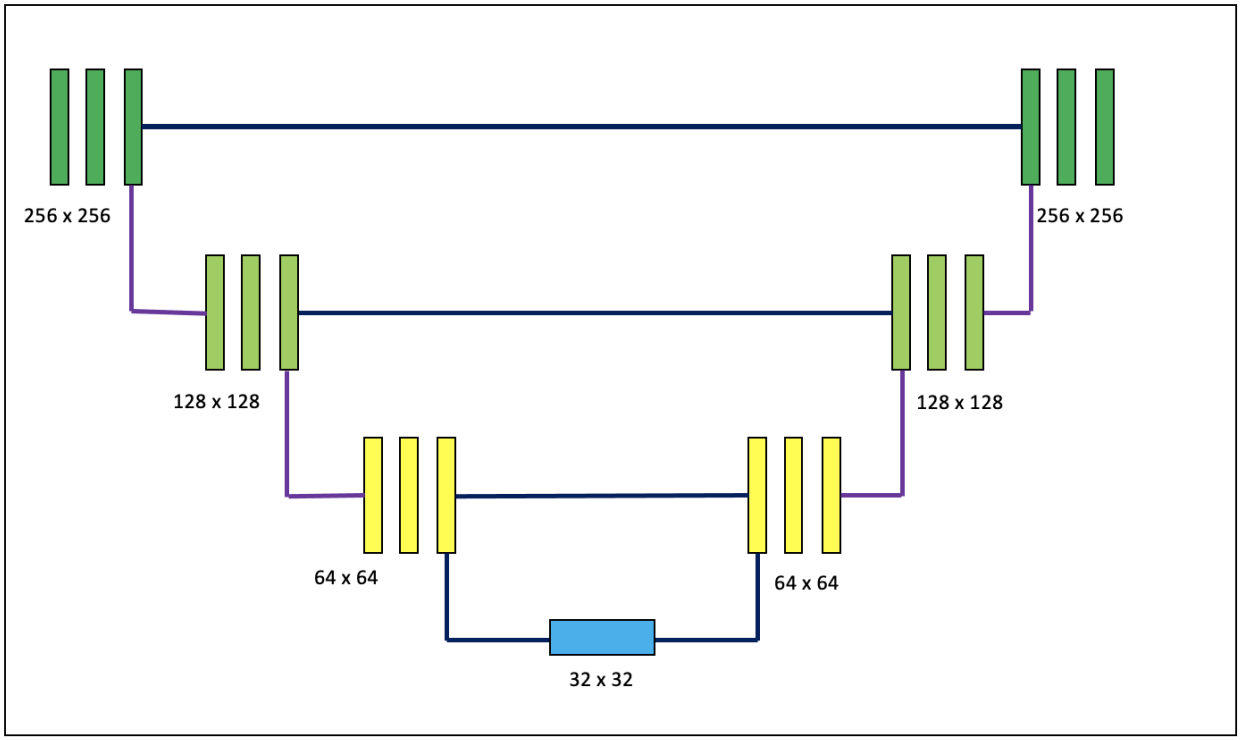


Figure 3.12. Represented here is an autoencoder- decoder network with three convolutional layers in each down-sampling stage of the encoder/analysis network. The input dimensions reduce from 256×256 to 32×32 in the bottleneck component and is then up-sampled in consecutive three convolutional layers to the output layer of 256×256. There are skip connections from the first layer of the network to the $n-1$ layer, second layer to the $n-2$ layer and third layer to the $n-3$ layer. Here, n is considered to be 7 with six layer comprising of 3 sequential convolutions. followed by max-pooling and an activation function. The final layer is a fully connected network leading to the output.

However, this system as observed in Figure 3.11 has certain shortfalls. The main of which being the loss of information and features. This can lead to a vanishing gradient problem. The vanishing gradient arises when the parameters of the network cannot be updated or modified sufficiently in each iteration due to the lack of a gradient large enough to make an impactful difference/learning. This can be attributed to the loss of feature differentials in the later layers of the network.

Mitigating this requires the introduction of low level features in the high level convolutions of the neural network structure. This is structurally illustrated in Figure 3.12. With the

introduction of skip connections from the first layer of the network to the $n-1$ layer, second layer to the $n-2$ layer and third layer to the $n-3$ layer, for a n layered convolutional network; the classification accuracy is observed to me much higher and the vanishing gradient problem is considerably mitigated[45].

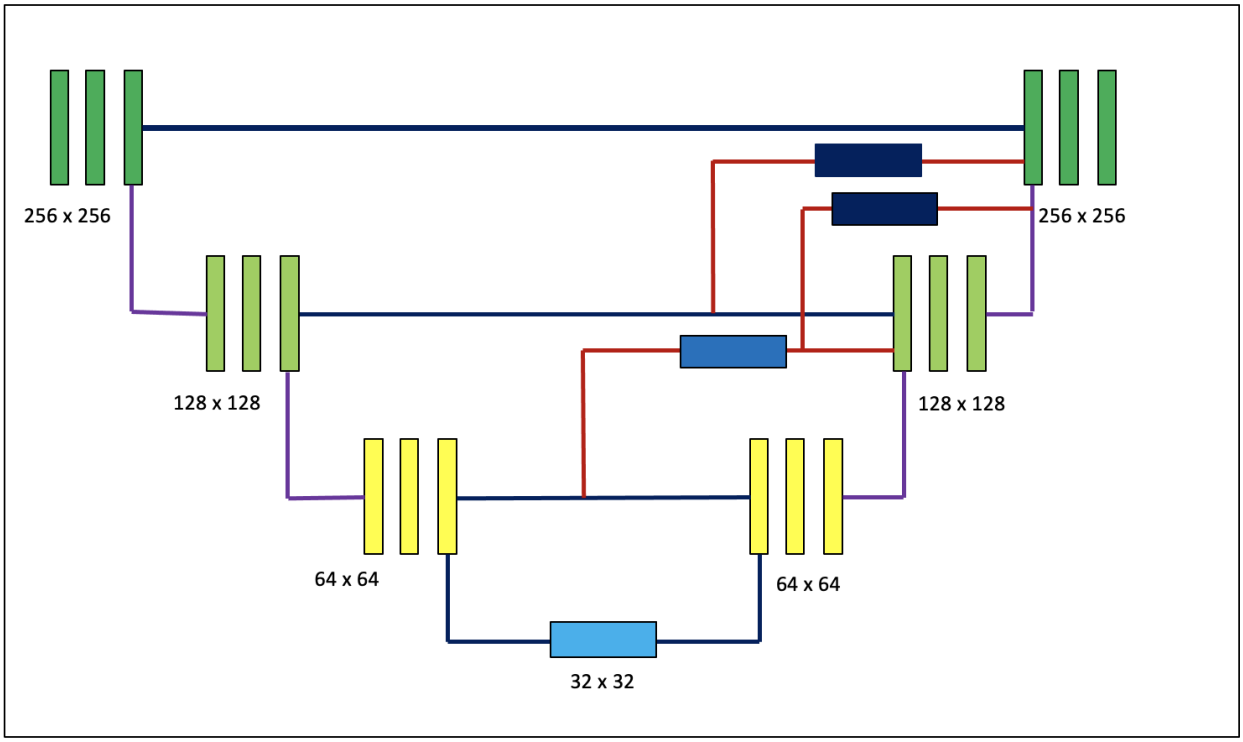


Figure 3.13. Represented here is an autoencoder- decoder network with three convolutional layers in each down-sampling stage of the encoder/analysis network. The input dimensions reduce from 256×256 to 32×32 in the bottleneck component and is then up-sampled in consecutive three convolutional layers to the output layer of 256×256. There are skip connections from the first layer of the network to the $n-1$ layer, second layer to the $n-2$ layer and third layer to the $n-3$ layer. Here, n is considered to be 7 with six layers comprising of 3 sequential convolutions. followed by max-pooling and an activation function. The final layer is a fully connected network leading to the output. This network also has additional skip connections from the third layer to the $n-2$ and $n-1$ layer. Along with skip connections from the second layer to the $n-1$ layer.

Drawing from this observation, we propose the inclusion of a further set of skip connections. More specifically, we introduce a skip connection from the second layer of grouped

convolutions to the $n-1$ layer, this is followed by a skip connection from the third layer of grouped convolutions to the $n-1$ and $n-2$ layer. This is the final structure of our segmenter architecture.

In this framework, the basic autoencoder-decoder structure, as witnessed in Figure 3.11 is maintained for the removal of scanner noise, brain matter that does not significantly contribute to tumor sub region detection and noise present in the convolved feature maps extracting tumor region specific information. The large number of parameters of this network, prevents backpropagation of gradient through the network's large parameter space. Overcoming this with skip connections as observed in Figure 3.12 proves to not provide optimal results for complex input data as that of the human brain.

To supplement the learning, prevent the vanishing gradient problem and provide adequate low level features to the high level complex feature space, we provide the network illustrated in Figure 3.13. This network offers us the opportunity to view the image space through four different views of the region of interest in the sixth convolutional layer. It also offers three different views of the region of interest in the fifth convolutional layer. Together, the network is able to remove noise, highlight tumor sub-regions and provide a layered view of the MR image, thus facilitating segmentation.

3.3.2.2 Discriminator Design

The discriminator is a component of the adversarial learning framework. Where the segmenter tries to create label maps of tumor sub regions from the MR Image, the discriminator trains itself in achieving better accuracy in differentiating between the provided segmentation map and the generated segmentation map.

To develop the discriminator for our network we use four convolution layers, with each followed by max pooling, rectified linear unit activation and batch normalization. Finally, they are passed through two fully connected layers and output to a single neuron with a binary classification function.

The job of the discriminator is to act as a classifier that can differentiate between the segmenter generated images (classified as false images – 0) and the ground truth (classified

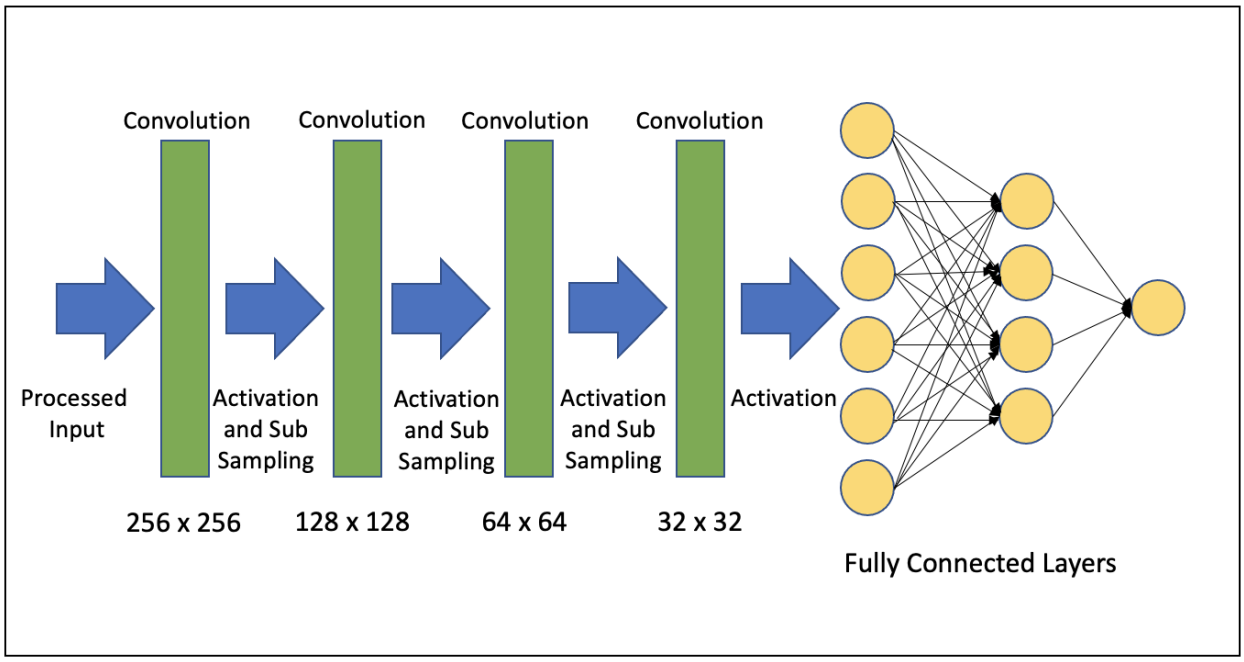


Figure 3.14. The discriminator has four convolution layers, with each followed by max pooling, rectified linear unit activation and batch normalization. They are then, passed through two fully connected layers and output to a single neuron with a binary classification function.

as true images – 1). The loss functions used here is a Binary Cross-entropy. We evaluate the performance of the discriminator using the binary classifier where the ground truth images are provided a target prediction value of 1 and the generated images are provided a target prediction value of 0. Then, to evaluate the performance of the segmenter above the dice loss, we use the binary crossentropy loss where the generated images are passed with a target prediction value of 0. This gives us a measure of how closely the segmenter's output mimics the ground truth.

3.3.3 Training

With the initial training run of the network, known as the first epoch, the loss is considerably high. This would translate to the knowledge that the segmenter is not able to generate a label map which is maximally similar to the ground truth. For each ground truth

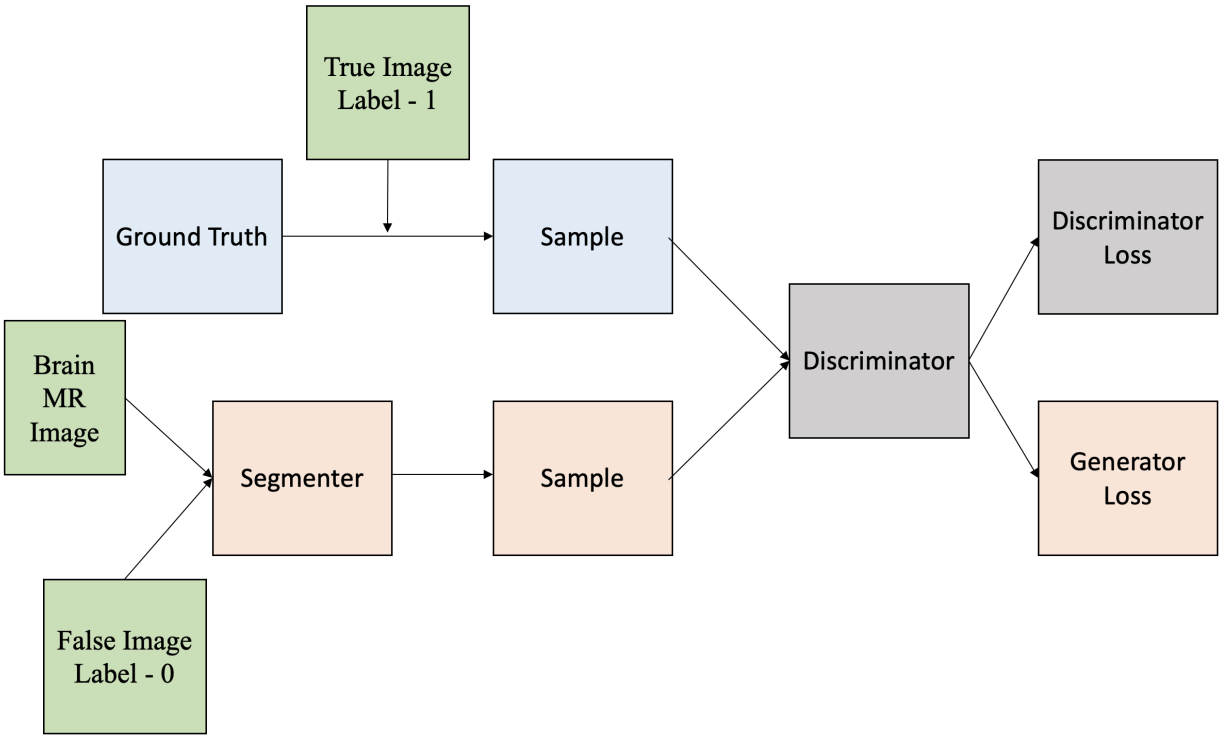


Figure 3.15. Our proposed Conditional Generative Adversarial Network with a generator G and a discriminator D. The ground truth and the segmenter produced output, along with label information is sampled and passed into the discriminator D. As training progresses the task of the network is to get better at differentiating between the real image and the conditionally generated image, while simultaneously training the generator G to produce more real like conditionally generated images.

label,i.e., core tumor, whole tumor and enhancing tumor regions, we calculate the dice loss which generates from the Sorensen Dice Coefficient[54] for crisp label detection. Here, all pixels which belong to the label category and have not been segmented as that label are used to determine the loss for that particular label. In the end, the total loss is formulated as the average of the individual losses of each category in the label. The dice score can be mathematically represented by $2|A \cap B|/(|A| + |B|)$, where A is the ground truth and B is the generated image. When we subtract the dice score from 1, we obtain the dice loss.

In addition to the dice loss, the segmenter is supplied another loss function which arises as a penalty for not achieving the maximal similarity to the ground truth. This loss function is

calculated as the loss incurred when the discriminator tries to classify the Segmenter's output as the ground truth. In other words, we feed the segmenter generated images through the Discriminator and calculate the loss incurred in classifying it as true images belonging to classification category 1. The gradient of both these losses are fed into the Segmenter to update its parameters for the next training cycle or epoch.

The distribution of data provided to the segmenter p_g is learnt over the input x. Here, the segmenter G utilizes a prior random noise vector $p_z(z)$ to form a mapping to the generated output $G(z;\theta_g)$, where z is the random input vector/noise and θ_g represents the parameters being optimized by the generator network. While in the case of the discriminator D, the network learns whether the sample belongs to the data x or is a product of the generator distribution p_g. This is learnt over $D(x;\theta_d)$, where θ_d represents the parameters of the model being optimized by the discriminator network.

This joint distribution of $p_z(z)$ and k work together to compose a hidden representation which optimizes the segmenter G model parameters θ_g. The discriminator D learns to optimize its parameters θ_d from the segmented images and the ground truth. This entire network is represented as $S(G,D)$ by the Equation 3.4

$$ăă min_G max_D S(D,G) = E_{x\sim p_{data(x)}}ă[\log(D(x|k))] + E_{z\sim p_{z_{(z)}}}[\log(1-D(G(z|k)))] \quad (3.4)$$

The label provided acts as the bias in the production of conditioned intermediate generated image. The loss function for the segmenter $G$is modified to include this additional information k as $\log(1-D(G(z|k)))$. The discriminator D, also has its loss function modified to $\log(D(x|k))$. This is ultimately used to update and optimize its parameters θ_d. We additionally induce a L1 loss metric between the segmented image and the ground truth. Thus, modifying the equation of our proposed network to be Equation 3.5

$$min_G max_D S(D,G) = E_{x\sim p_{data(x)}}[\log(D(x|k))]+$$

$$\mathrm{E}_{z\sim p_{z_{(z)}}}[\log(1-D(G(z|k)))] + E_{z\sim p_{z_{(z)}},x\sim p_{data(x)}}[||x-G(z|k)||] \quad (3.5)$$

The objective of this modification is to not only help the segmenter fool the discriminator but also land a segmented image as closely representative of the ground truth image as possible.

We train three segmenter models at a time. With the execution of an entire epoch, each of the coronal, saggital and axial segmenters are evaluated by recomposing the sequence of images back to its initial frame of reference. The proposed algorithm weights the labels for each plane with a coefficient 1/3 as all planes are assumed to be equally important. We then compute the dice score and the dice loss for the three segmenter models, pass it through the discriminator and backpropagate the losses.

Table 3.1. Dice Scores for the Proposed Network

Data	**Enhancing Tumor**	**Whole Tumor**	**Tumor Core**
Training	0.81	0.91	0.86
Validation	0.80	0.89	0.85
Testing	0.80	0.88	0.84

Table 3.2. Specificity for the Proposed Network

Data	**Enhancing Tumor**	**Whole Tumor**	**Tumor Core**
Testing Set	0.81	0.88	0.80

Table 3.3. Sensitivity for the Proposed Network

Data	**Enhancing Tumor**	**Whole Tumor**	**Tumor Core**
Testing Set	0.79	0.88	0.85

We train our model on 80 epochs and use early stopping to determine its termination. The training loss is visualized in Figure 3.16, and the results on the testing dataset are demonstrated in Figures 3.17 to 3.22.

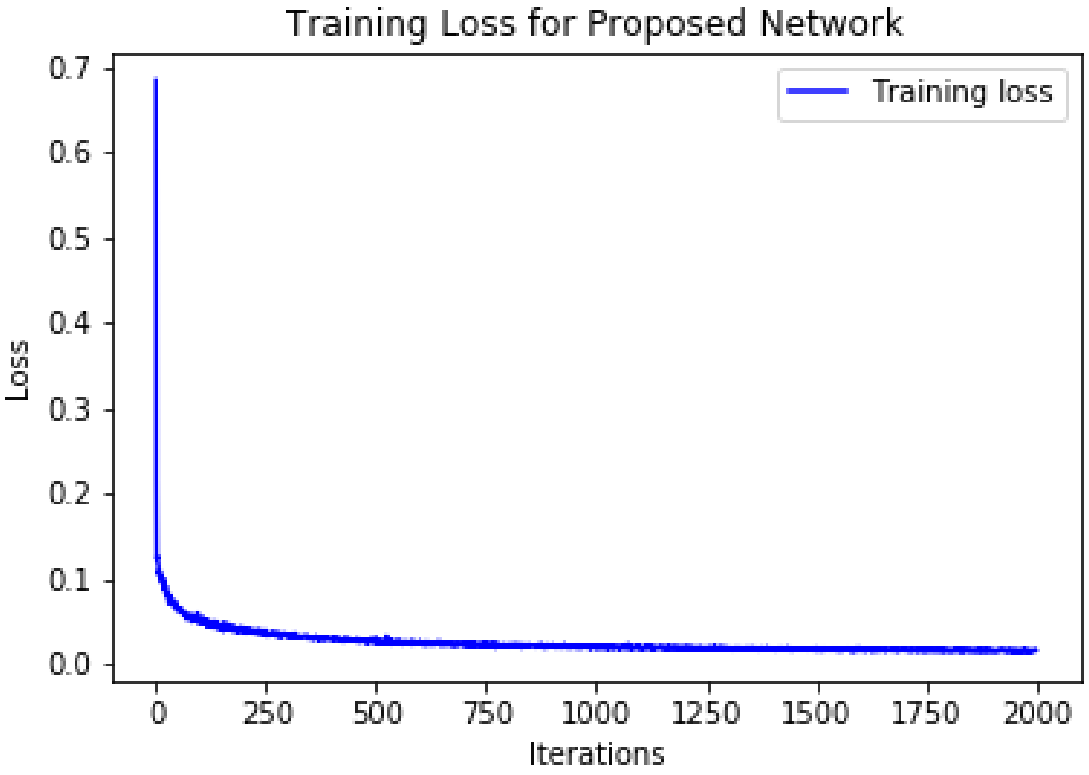
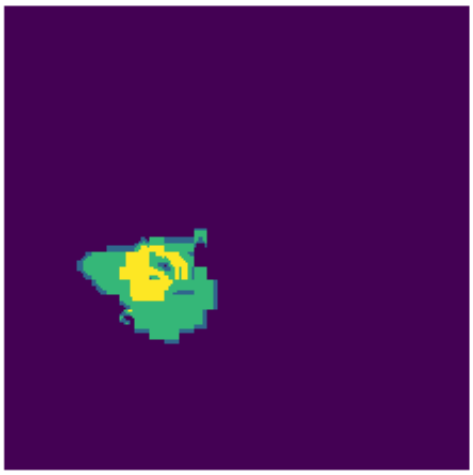


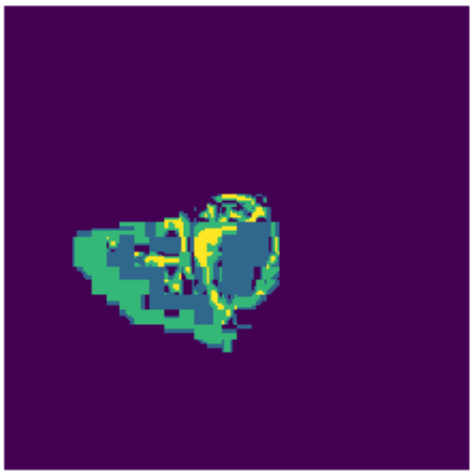
Figure 3.16. The training loss of the proposed network is illustrated.



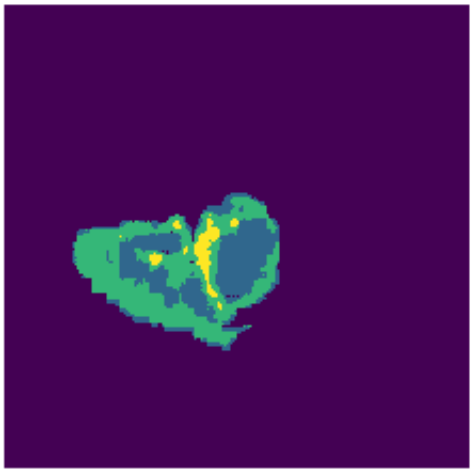
(a)

(b)

Figure 3.17. On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset.

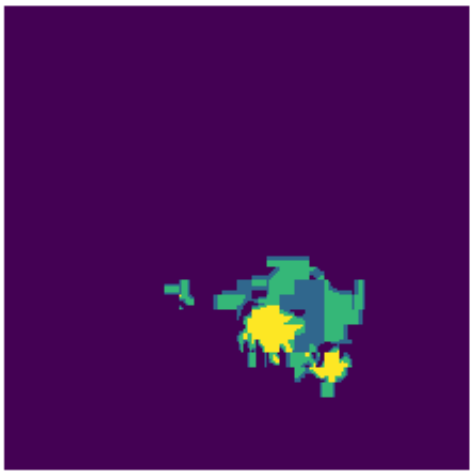


(a)

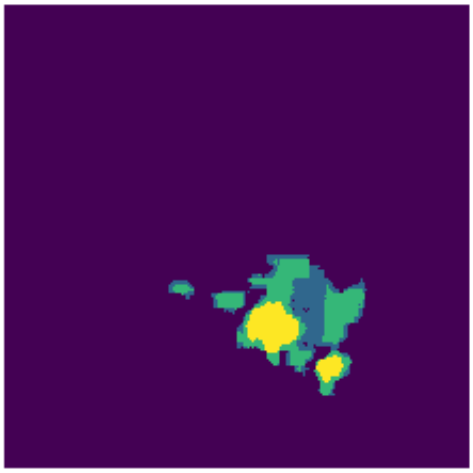


(b)

Figure 3.18. On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset.

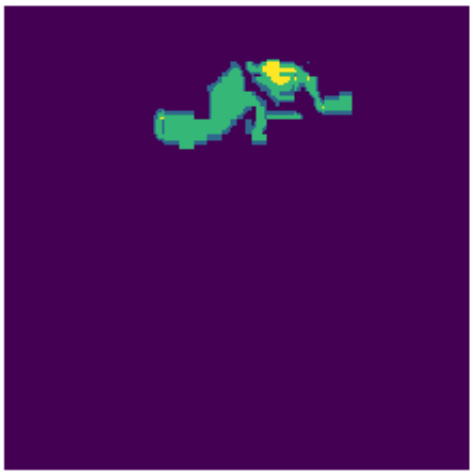


(a)

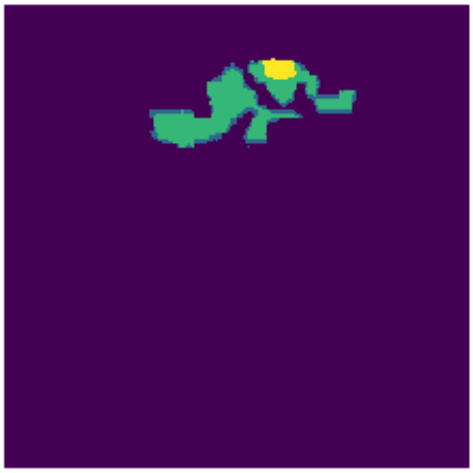


(b)

Figure 3.19. On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset.

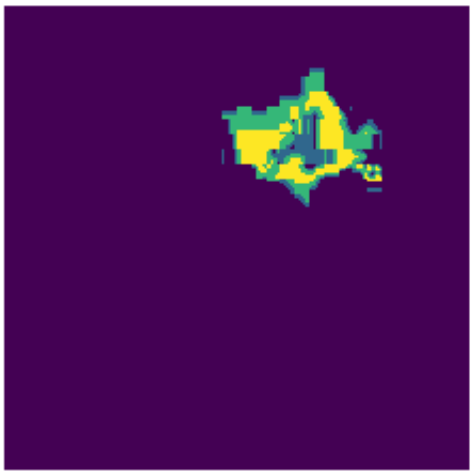


(a)

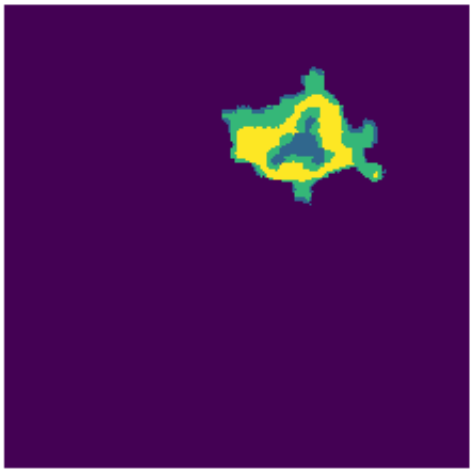


(b)

Figure 3.20. On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset.

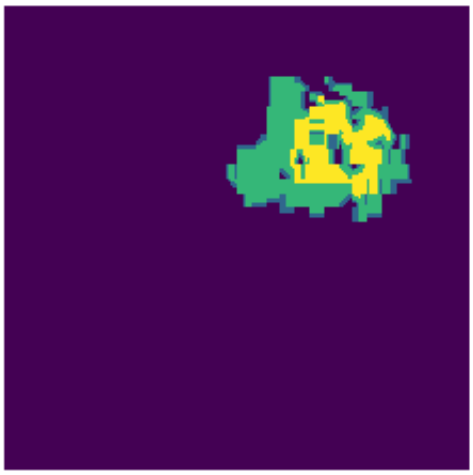


(a)

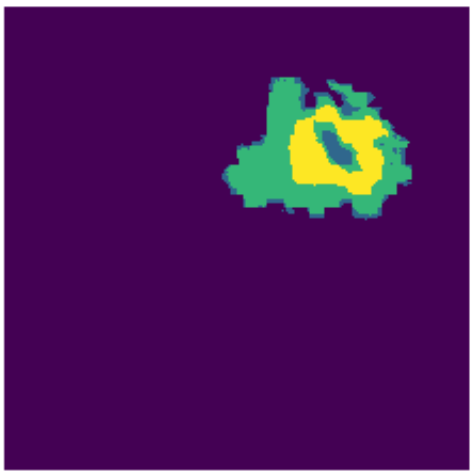


(b)

Figure 3.21. On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset.



(a)



(b)

Figure 3.22. On the top we have the ground truth and on the bottom we have its corresponding predicted segmentation. Results are obtained on the testing dataset.

4. SUMMARY

Deep neural learning algorithms have demonstrated competitive results in Brain Tumor sub region segmentation. Their popularity in overcoming this challenging task stems from their powerful feature learning capability. The features extracted from these extensive learning processes ultimately help in delineating tumor sub regions, through their high differential abilities and dynamic memory systems. Like any optimization technique, the formulation of the problem loss function, learning conditions and parameters play an important role in determining outcome. The results in our work provide a promising approach to solving three dimensional data problems from a two dimensional perspective. Our proposed network learns a custom loss applicable in any three dimensional image space.

The two dimensional approach helps in overcoming overhead incurred with the usage of three dimensional kernels. Any increase in resource consumption due to the training of three segmenter modules is justified by the sequential views obtained in the three planar levels. These sequential views in all three planes, sagital, axial and coronal, are unique and serve to increment the size of the dataset. They provide adequate data points for the training of two dimensional kernels and helps circumvent the problem of overfitting. The traditional three dimensional kernel, would require a larger amount of training data for a complex networks, such as ours. Dividing this problem up into three separate planar views of two dimensional kernels help in not only increasing the size of the dataset, but also keeps each planar view independent from the other. Ultimately, this achieves in proving a multi dimensional view of the problem while avoiding the learning of redundancies on the part of the model.

Additionally, the update in model parameters with a weighted loss gradient computation steers the optimization function towards a steady minima. Ultimately, this work tackles the problem of brain tumor segmentation and is successful in providing competitive results.

REFERENCES

[1] Mikael Agn et al. "Brain Tumor Segmentation Using a Generative Model with an RBM Prior on Tumor Shape". In: vol. 9556. Jan. 2016, pp. 168–180. ISBN: 978-3-319-30857-9. DOI: 10.1007/978-3-319-30858-6_15.

[2] Meritxell Bach Cuadra et al. "Atlas-Based Segmentation of Pathological MR Brain Images Using a Model of Lesion Growth". In: *IEEE transactions on medical imaging* 23 (Nov. 2004), pp. 1301–14. DOI: 10.1109/TMI.2004.834618.

[3] Spyridon Bakas et al. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features". In: *Scientific Data* 4.1 (Dec. 2017), p. 170117. ISSN: 2052-4463. DOI: 10.1038/sdata.2017.117.

[4] Spyridon Bakas et al. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features". In: *Scientific Data* 4.1 (Dec. 2017), p. 170117. ISSN: 2052-4463. DOI: 10.1038/sdata.2017.117.

[5] Spyridon Bakas et al. "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge". In: *arXiv:1811.02629 [cs, stat]* (Apr. 2019). arXiv: 1811.02629.

[6] Spyridon Bakas et al. *Segmentation Labels for the Pre-operative Scans of the TCGA-LGG collection*. 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF. URL: https://wiki.cancerimagingarchive.net/x/LIZyAQ.

[7] Dor Bank, Noam Koenigstein, and Raja Giryes. *Autoencoders*. 2021. arXiv: 2003.05991 [cs.LG].

[8] Leo Breiman. "Bagging Predictors". In: *Machine Learning* 24.2 (Aug. 1996), pp. 123–140. ISSN: 1573-0565. DOI: 10.1023/A:1018054314350. URL: https://doi.org/10.1023/A:1018054314350.

[9] Adrià Casamitjana et al. "3D Convolutional Neural Networks for Brain Tumor Segmentation: A Comparison of Multi-resolution Architectures". In: Apr. 2016, pp. 150–161. ISBN: 978-3-319-55523-2. DOI: 10.1007/978-3-319-55524-9_15.

[10] Emily L Denton et al. “Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks”. In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes et al. Curran Associates, Inc., 2015, pp. 1486–1494. URL: http://papers.nips.cc/paper/5773-deep-generative-image-models-using-a-laplacian-pyramid-of-adversarial-networks.pdf.

[11] Pavel Dvorak and B. Menze. “Structured prediction with convolutional neural networks for multimodal brain tumor segmentation”. In: (Oct. 2015), pp. 13–24.

[12] Pavel Dvořák and Bjoern Menze. “Local Structure Prediction with Convolutional Neural Networks for Multimodal Brain Tumor Segmentation”. In: *Medical Computer Vision: Algorithms for Big Data*. Ed. by Bjoern Menze et al. Cham: Springer International Publishing, 2016, pp. 59–71. ISBN: 978-3-319-42016-5.

[13] Karl Fritscher et al. “Deep Neural Networks for Fast Segmentation of 3D Medical Images”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Ed. by Sebastien Ourselin et al. Cham: Springer International Publishing, 2016, pp. 158–165. ISBN: 978-3-319-46723-8.

[14] Kunihiko Fukushima. “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”. In: *Biological Cybernetics* 36.4 (Apr. 1980), pp. 193–202. ISSN: 1432-0770. DOI: 10.1007/BF00344251. URL: https://doi.org/10.1007/BF00344251.

[15] R. Girshick et al. “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 580–587.

[16] M. Goetz et al. “DALSA: Domain Adaptation for Supervised Learning From Sparsely Annotated MR Images”. In: *IEEE Transactions on Medical Imaging* 35.1 (2016), pp. 184–196.

[17] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[18] Ian Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[19] A. Gooya et al. “GLISTR: Glioma Image Segmentation and Registration”. In: *IEEE Transactions on Medical Imaging* 31.10 (Oct. 2012), pp. 1941–1954. ISSN: 0278-0062, 1558-254X. DOI: 10.1109/TMI.2012.2210558.

[20] Mohammad Havaei et al. “Brain Tumor Segmentation with Deep Neural Networks”. In: *Medical Image Analysis* 35 (May 2015). DOI: 10.1016/j.media.2016.05.004.

[21] K. Hu et al. “Brain Tumor Segmentation Using Multi-Cascaded Convolutional Neural Networks and Conditional Random Field”. In: *IEEE Access* 7 (2019), pp. 92615–92629.

[22] Phillip Isola et al. *Image-to-Image Translation with Conditional Adversarial Networks*. 2018. arXiv: 1611.07004 [cs.CV].

[23] Koray Kavukcuoglu et al. “Learning Convolutional Feature Hierachies for Visual Recognition”. In: *Advances in Neural Information Processing Systems (NIPS 2010)*. Vol. 23. 2010.

[24] Henry J Kelley. “Gradient theory of optimal flight paths”. In: *Ars Journal* 30.10 (1960), pp. 947–954.

[25] Asifullah Khan et al. “A survey of the recent architectures of deep convolutional neural networks”. In: *Artificial Intelligence Review* (Apr. 2020). ISSN: 1573-7462. DOI: 10.1007/s10462-020-09825-6. URL: https://doi.org/10.1007/s10462-020-09825-6.

[26] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[27] Y. LeCun et al. “Backpropagation Applied to Handwritten Zip Code Recognition”. In: *Neural Computation* 1.4 (Winter 1989), pp. 541–551.

[28] Chuan Li and Michael Wand. *Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks*. 2016. arXiv: 1604.04382 [cs.CV].

[29] H. Li and Y. Fan. "Label propagation with robust initialization for brain tumor segmentation". In: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. 2012, pp. 1715–1718.

[30] Hao Li et al. *Visualizing the Loss Landscape of Neural Nets*. 2018. arXiv: 1712.09913 [cs.LG].

[31] Weiyang Liu et al. *Large-Margin Softmax Loss for Convolutional Neural Networks*. 2016. arXiv: 1612.02295 [stat.ML].

[32] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 3431–3440.

[33] Warren S. McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4 (Dec. 1943), pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259. URL: https://doi.org/10.1007/BF02478259.

[34] Raphael Meier et al. "Appearance-and Context-sensitive Features for Brain Tumor Segmentation". In: Sept. 2014. DOI: 10.13140/2.1.3766.7846.

[35] W. Mengqiao et al. "The multimodal brain tumor image segmentation based on convolutional neural networks". In: *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)*. 2017, pp. 336–339.

[36] B. H. Menze et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)". In: *IEEE Transactions on Medical Imaging* 34.10 (2015), pp. 1993–2024.

[37] Bjoern H. Menze et al. "A generative model for brain tumor segmentation in multi-modal images". English. In: *Medical Image Computing and Computer-Assisted Intervention*. Vol. 6362. Lecture Notes in Computer Science. 13th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI2010 ; Conference date: 20-09-2010 Through 24-09-2010. Springer-verlag Berlin,

2010, pp. 151–159. ISBN: 978-3-642-15744-8. DOI: 10.1007/978-3-642-15745-5_19. URL: http://www.miccai2010.org/.

[38] Mehdi Mirza and Simon Osindero. *Conditional Generative Adversarial Nets*. 2014. arXiv: 1411.1784 [cs.LG].

[39] Tom M. Mitchell. *Machine Learning*. New York: McGraw-Hill, 1997. ISBN: 978-0-07-042807-2.

[40] S. Pereira et al. "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images". In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1240–1251.

[41] Sérgio Pereira et al. "Deep Convolutional Neural Networks for the Segmentation of Gliomas in Multi-sequence MRI". In: vol. 9556. Jan. 2016, pp. 131–143. ISBN: 978-3-319-30857-9. DOI: 10.1007/978-3-319-30858-6_12.

[42] Adriano Pinto et al. "Hierarchical Brain Tumour Segmentation using Extremely Randomized Trees". In: *Pattern Recognition* 82 (May 2018). DOI: 10.1016/j.patcog.2018.05.006.

[43] Adhish Prasoon et al. "Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network". In: *Advanced Information Systems Engineering*. Springer Berlin Heidelberg, 2013, pp. 246–253. DOI: 10.1007/978-3-642-40763-5_31.

[44] Alec Radford, Luke Metz, and Soumith Chintala. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2016. arXiv: 1511.06434 [cs.LG].

[45] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].

[46] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain." In: *Psychological Review* 65.6 (1958), pp. 386–408. ISSN: 0033-295X. DOI: 10.1037/h0042519. URL: http://dx.doi.org/10.1037/h0042519.

[47] Mousumi Roy et al. “Convolutional autoencoder based model HistoCAE for segmentation of viable tumor regions in liver whole-slide images”. In: *Scientific Reports* 11.1 (Dec. 2021), p. 139. ISSN: 2045-2322. DOI: 10.1038/s41598-020-80610-9.

[48] S. Ruan et al. “TUMOR SEGMENTATION FROM A MULTISPECTRAL MRI IMAGES BY USING SUPPORT VECTOR MACHINE CLASSIFICATION”. In: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2007, pp. 1236–1239.

[49] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning representations by back-propagating errors”. In: *Nature* 323.6088 (Oct. 1986), pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0. URL: https://doi.org/10.1038/323533a0.

[50] Tim Salimans et al. “Improved Techniques for Training GANs”. In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee et al. Curran Associates, Inc., 2016, pp. 2234–2242. URL: http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.pdf.

[51] Dominik Scherer, Andreas Müller, and Sven Behnke. “Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition”. In: *Artificial Neural Networks – ICANN 2010*. Ed. by Konstantinos Diamantaras, Wlodek Duch, and Lazaros S. Iliadis. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 92–101. ISBN: 978-3-642-15825-4.

[52] A. A. A. Setio et al. “Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks”. In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1160–1169.

[53] Mohammadreza Soltaninejad et al. “Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI”. In: *International Journal of Computer Assisted Radiology and Surgery* 12.2 (Feb. 2017), pp. 183–203. ISSN: 1861-6429. DOI: 10.1007/s11548-016-1483-3. URL: https://doi.org/10.1007/s11548-016-1483-3.

[54] Carole H. Sudre et al. “Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations”. In: *Lecture Notes in Computer Science* (2017), pp. 240–248. ISSN: 1611-3349. DOI: 10.1007/978-3-319-67558-9_28.

[55] Igor V. Tetko, David J. Livingstone, and Alexander I. Luik. “Neural network studies, 1. Comparison of overfitting and overtraining”. In: *J. Chem. Inf. Comput. Sci.* 35 (1995), pp. 826–833.

[56] Gregor Urban et al. “Multi-modal Brain Tumor Segmentation using Deep Convolutional Neural Networks”. In: 2014.

[57] Dongxian Wu et al. *Skip Connections Matter: On the Transferability of Adversarial Examples Generated with ResNets*. 2020. arXiv: 2002.05990 [cs.LG].

[58] Panpan Wu et al. “Classification of Lung Nodules Based on Deep Residual Networks and Migration Learning”. In: *Computational Intelligence and Neuroscience* 2020 (Mar. 2020), p. 8975078. ISSN: 1687-5265. DOI: 10.1155/2020/8975078. URL: https://doi.org/10.1155/2020/8975078.

[59] Zhao Xiaomei et al. “Brain Tumor Segmentation Using a Fully Convolutional Neural Network with Conditional Random Fields”. In: Apr. 2016, pp. 75–87. ISBN: 978-3-319-55523-2. DOI: 10.1007/978-3-319-55524-9_8.

[60] Zhilu Zhang and Mert R. Sabuncu. “Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels”. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS’18. Montréal, Canada: Curran Associates Inc., 2018, pp. 8792–8802.

[61] Darko Zikic et al. “Segmentation of Brain Tumor Tissues with Convolutional Neural Networks”. In: Sept. 2014.