STUDY OF ATTENTION NETWORK AND ITS IMPLEMENTATION IN BRAIN TUMOR SEGMENTATION

A Project Report

submitted by

ANAMIKA SINHA

in partial fulfilment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

MAY 2019

THESIS CERTIFICATE

This is to certify that the thesis titled **STUDY OF ATTENTION NETWORK AND ITS IMPLEMENTATION IN BRAIN TUMOR SEGMENTATION**, submitted by **Anamika Sinha**, to the Indian Institute of Technology, Madras, for the award of the degree of **Master of Technology**, is a bona fide record of the research work done by her under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. 1
Dr. Ganpathy Krishnamurthi
Associate Professor
Dept. of Engineering Design
IIT Madras, 600 036

Prof. 2
Dr. Bharath Bhikkaji
Associate Professor
Dept. of Electrical Engineering
IIT Madras, 600 036

Place: Chennai

Date: 5/5/2019

ACKNOWLEDGEMENTS

At the outset, I express my deep sense of gratitude to my Project Guide Prof. Dr. Ganpathy Krishnamurthi for his guidance, support, encouragement and help throughout the period of the project work. I am highly indebted to him for devoting his valuable time to help me complete the work in time. I would also like to thank Dr. Bharath Bhikkaji for being my co-guide in this project. Also, I would like to thank the Department of Engineering Design and Department of Electrical Engineering, IIT Madras for providing me the opportunity to work in this challenging field.

ABSTRACT

Artificial Intelligence has gain a tremendous amount of attention in recent years in field of medical imaging because of substantial improvements in image recognition performance, especially based on class of algorithms known as deep learning. Medical image analysis involves measurements in medical images i.e. extraction of relevant quantitative information from images. Manual measurements by human experts in large 3D medical imaging datasets are not only tedious and time consuming but also impractical clinical routine. Thus deep learning technique has gain a lot of importance in medical image classification, segmentation and registration purpose. In large medical images there are lot of regions which are irrelevant. In order to focus on relevant regions of Images, researchers have introduced attention networks. There are many applications in which these networks have gained importance and improved efficiency. One of the application in medical image analysis is brain tumor segmentation. Many research work has been done on several deep learning models to segment tumor effectively and successful in getting good accuracy and dice score. This thesis aims at studying brain tumor MRI data, various attention networks and implementation of soft attention network for brain tumor segmentation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CNN	Convolution neural network
FCN	Fully convolution network
FC	Fully connected
AG	Attention gate
MRI	Magnetic resonance imaging

CHAPTER 1

INTRODUCTION

Artificial intelligence and deep learning have captivate the healthcare industry as these innovative analytics strategies become more accurate and applicable to a variety of tasks. With the advancement in technology large database of well documented imaging data and other medical information of patients are being built up. Novel imaging modalities techniques such as multi-slice, multi-frame (MRI) , multi-dimensional, multi-modal techniques are being introduced for medical image acquisition. With availability of large well documented data, it is possible to introduce deep learning techniques in medical field also. One type of deep learning, kown as convolutional neural network is well suited to analyzing images such as MRI images or x-rays. CNNs automatically learn representative complex features directly from the data itself. Medical image analysis involves image classification, image segementation and image registration. One of the application of medical image analysis is brain tumor segmentation from 3D MRI images. Accurate 3D segmentation of complex shaped objects in medical images is usually complicated by the limited resolution of the images and by the fact that resolution is often not isotropic (mostly multi-slice 2D instead of accurate 3D acquisitions). Many researchers have used deep learning networks such as Fully convolutional network, attention U-Net network for 3D segmentation of medical images. Advantage of attention network is that it focus on relevant regions of images ignoring noisy data hence increasing the accuracy and performance.

CHAPTER 2

PREVIOUS WORK

Several research work has been done in field of brain tumor segmentation. Early work on brain tumor segmentation used methods based on generative models which requires domain specific prior knowledge about healthy and tumorous tissue. In one of the generative model proposed by Prastawa *et al.* (2004), they used registered brain atlas to detect abnormal regions and then computes the posterior probability of different tissue types. Tumorous region is detected by localising voxels whose posterior probabilities is below threshold. In 2009, Khotanlou *et al.* (2009) proposed brain tumor segmentation based on selecting asymmetric areas with respect to the approximate brain symmetry plane and fuzzy classification technique.

Later many discriminative models such as SVMs and decision forest have been proposed which does not require much prior knowledge of brain's anatomy and rely on low level image features. One disadvantage with this model is the use of large number of hand designed features in order to be accurate which increases the computation time and also these features exploit very generic edge related information with no specific adaptation to domain of brain tumors.

Inspired by the recent success of convolutional neural networks, number of deep learning based approaches have been proposed. Zikic *et al.* (2014) used 3D input patches that are interpreted into 2D input patches to train a CNN. Dice score for this model was 83.7 percent for whole tumor, 73.6 percent for core tumor and 69 percent for active tumor. Havaei *et al.* (2017) proposed cascaded two pathways convolutional neural network for capturing local and global features. Dice scores for these models ranges from 85 percent to 88 percent for whole tumor, 74 to 79 percent for core tumor and 68 to 72 percent for active tumor. Kayalibay *et al.* (2017) developed very successful adaptation of the popular U-Net architecture and achieved state of the art results for the BraTS 2015 dataset. Now-a-days neural networks are extensively used for detecting and recognising organs or tumors in medical images.

Apart brain tumor segmentation, several other deep learning models have been used

in other medical image classification and segmentaion problems such as detection of pancreas, classification of blood type, etc. Oktay *et al.* (2018) used attention U-Net model to detect pancreas. They proposed a novel attention gate (AG) model for medical imaging that automatically learns to focus on target structures of varying shapes and sizes and also suppress the irrelevant regions in an input image. Currently many research works have been done on use of attention networks and its advantages. Attention networks focus attention selectively on parts of the data to acquire information when and where it is needed, and combine information from different time stamps to build up an internal representation of the data and to make decisions. Focusing the computational resources on parts of a image saves memory and time complexity. It also reduces the task complexity as the object of interest can be placed in the center of the fixation and irrelevant features are ignored. Attention model have been used both in classification and segmentation.

In papers such as Mnih *et al.* (2014) and Momeni *et al.* (2018), they have used recurrent attention model for classification purpose which is mainly composed of recurrent neural network where each time step, it processes the sensor data, integrates information over time, and chooses how to act and how to deploy its sensor at next time step. Since network is not fully differentiable because of hard attention (glimpses), reinforcement learning is used. Further in order to make network fully differentiable, Ablavatski *et al.* (2017) proposed model which uses spatial transformer attention instead of getting glimpses by cropping images. In this model, all parts were fully differentiable that enables the optimization of the whole model end-to-end by using gradient descent and trained with standard back propagation. In papers such as Shen *et al.* (2018) and Oktay *et al.* (2018), they used soft attention mechanisms for segmentation purpose where attention gates are included in decoder network to focus on relevant part of image and ignore the irrelevant data.

CHAPTER 3

DEEP LEARNING AND ITS UNDERSTANDING

What is Deep Learning?

Deep learning is a Neural Network consisting of a hierarchy of layers, whereby each layer transforms the input data into more abstract representations. These series of layers, between input and output, identify the input features and create a series of new features based on the data, just as our brain. In deep learning the more layers a network has, the higher the level of features it will learn. The output layer combines all these features and makes a prediction.

A basic neural network

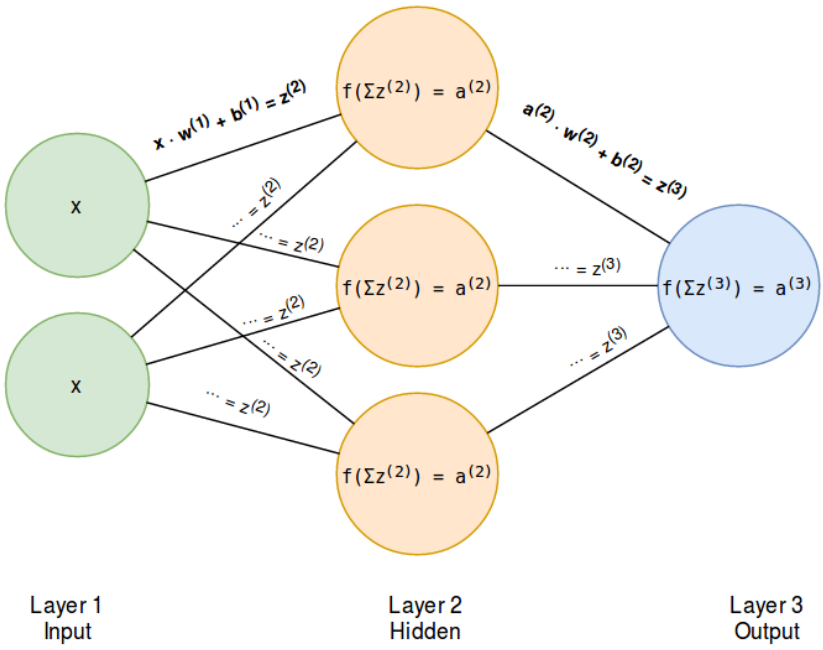

Figure 3.1: Neural network

A basic Neural Network mostly consists of 3 layers (input, hidden and output) and each layer consists of neuron/nodes that perform numerical computations and other operations. Each node in a layer is inter connected to other nodes present in consecutive layers. There are weights assigned to each interconnection and a bias assigned to each layer. These weights and biases are called parameters of the network.

The input layer in the network is responsible for receiving large volumes of data as inputs. The hidden layer is where all the calculation, computation and feature extraction take place. The output layer is responsible for generating the desired output. The hidden layers use several activation functions like ReLU, Sigmoid, Step function, etc, to calculate the weighted sum of inputs plus a bias is added. The output of the computation will decide which nodes to fire. A Softmax function is usually applied to get the proper output. The predicted output of the network is compared with the actual output and the difference in the output (i.e. the error) is back propagated through the network and the weights are adjusted accordingly. By using a cost function, the error in the network is calculated. The same process follows again to get the desired output.

Deep neural networks

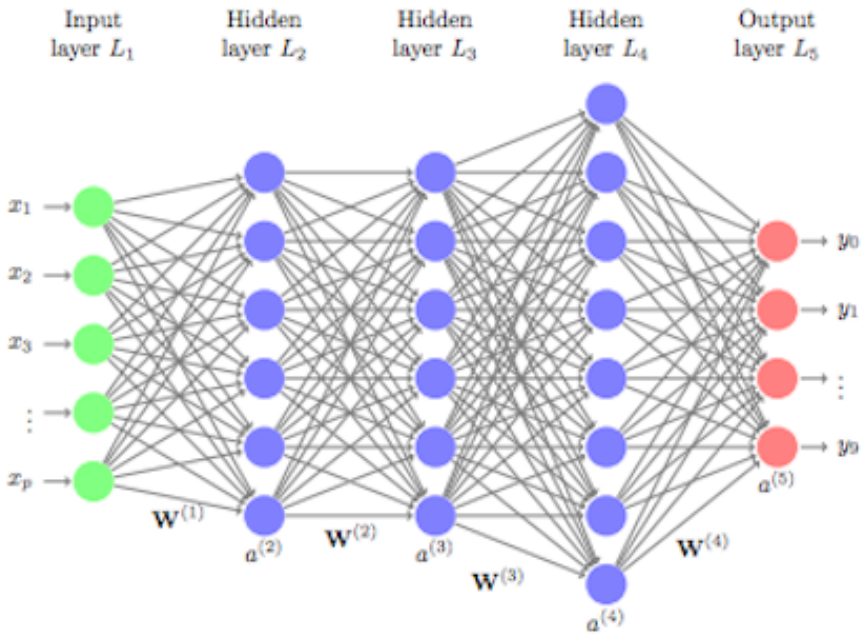

Figure 3.2: Deep Neural network

Deep neural networks consist of several hidden layers. Multiple DNN models exist

and, as interest and investment in this area have increased, expansions of DNN models have flourished. For example, convolutional neural networks (CNN or ConvNet) have wide applications in image and video recognition, recurrent neural networks (RNN) are used with speech recognition, and long short-term memory neural networks (LTSM) are advancing automated robotics and machine translation.

CHAPTER 4

CONVOLUTIONAL NEURAL NETWORK

Convolution neural network

A convolutional neural network is a class of deep neural network, which consist of several convolutional layers most commonly applied to visual imagery. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. As compared to feed forward neural network ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.

Building Blocks of ConvNet:

- Convolution layer
- Pooling layer
- Relu layer
- Batchnormalisation layer
- Dropout
- Fully connected layer

The most common form of a ConvNet architecture stacks a few CONV-RELU layers, follows them with POOL layers, and repeats this pattern until the image has been merged spatially to a small size. At some point, it is common to transition to fully-connected layers. The last fully-connected layer holds the output, such as the class scores.

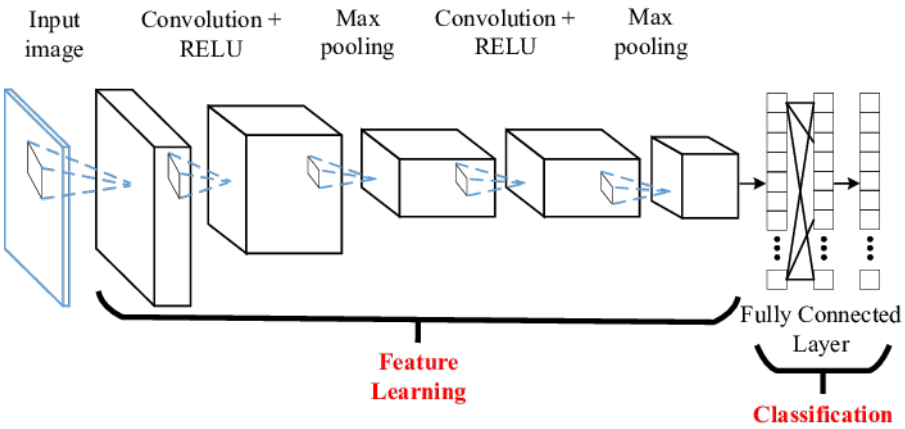

Figure 4.1: CNN architecture

Convolutional layer

The convolutional layer is the core building block of a CNN. The layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field, but extend through the full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot product between the entries of the filter and the input and producing a 2-dimensional activation map of that filter. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input.

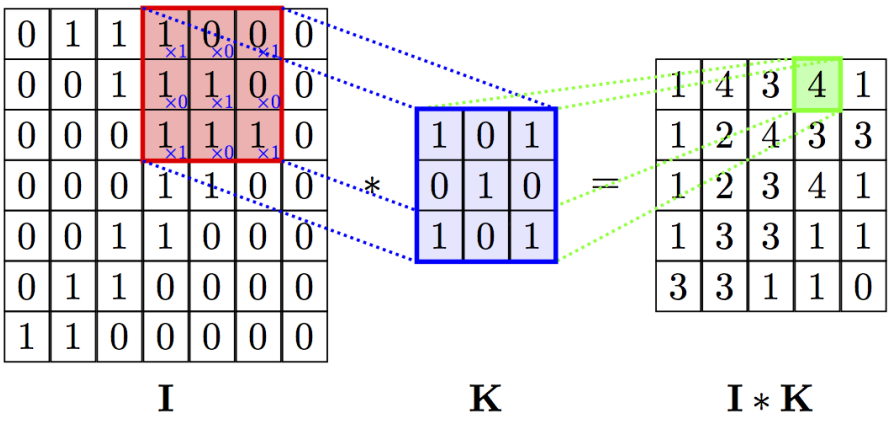

Figure 4.2: Convolution

Stacking the activation maps for all filters along the depth dimension forms the full output volume of the convolution layer. Every entry in the output volume can thus also be interpreted as an output of a neuron that looks at a small region in the input and shares parameters with neurons in the same activation map. The objective of the

Convolution Operation is to extract the high-level features such as edges, from the input image. ConvNets need not be limited to only one Convolutional Layer. Conventionally, the first ConvLayer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network which has the wholesome understanding of images in the dataset.

Pooling layer

The pooling or downsampling layer is responsible for reducing the spatial size of the activation maps. In general, they are used after multiple stages of other layers (i.e. convolutional and non-linearity layers) in order to reduce the computational requirements progressively through the network as well as minimizing the likelihood of overfitting. Spatial pooling can be of different types:

1 Max pooling

2 Average pooling

3 Sum pooling

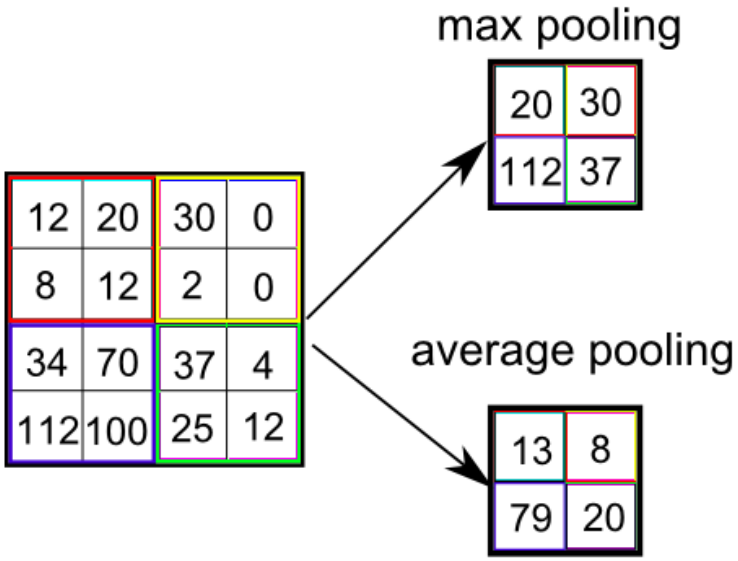

Figure 4.3: Pooling

Max pooling take the largest element from the rectified feature map. Taking the largest element could also take the average pooling. Sum of all elements in the feature

map call as sum pooling. The key concept of the pooling layer is to provide translational invariance since particularly in image recognition tasks, the feature detection is more important compared to the feature's exact location. Therefore the pooling operation aims to preserve the detected features in a smaller representation and does so, by discarding less significant data at the cost of spatial resolution.

ReLU layer

ReLU stands for Rectified Linear Unit for a non-linear operation and is a type of activation function. The output is max(0,x). The purpose of ReLU layer is to introduce non-linearity in convNet. Other activation functions are tanh and sigmoid. ReLU is most commonly used activation function because it performs better than two.

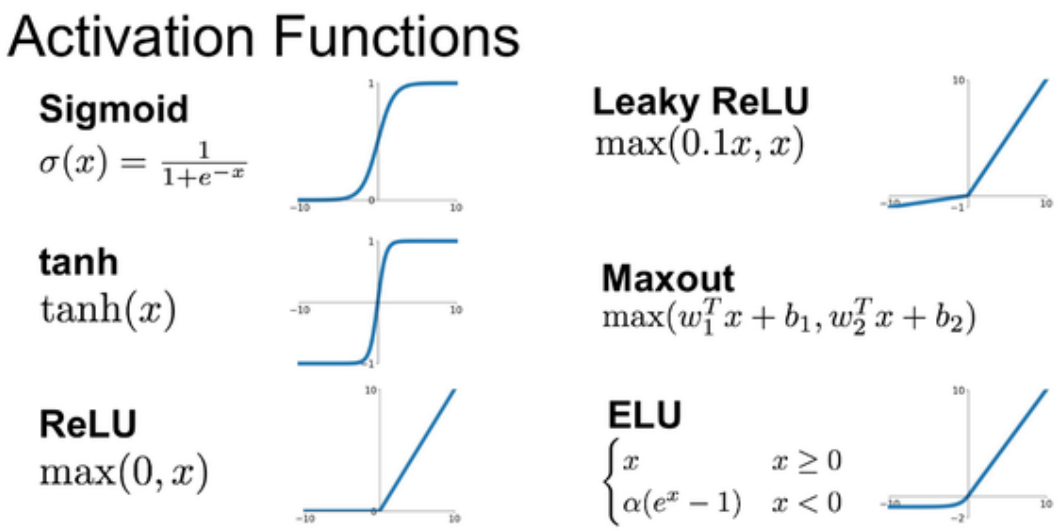

Figure 4.4: activation functions

Batchnormalization layer

Batch normalization is a technique for improving the speed, performance, and stability of artificial neural networks. It is used to normalize the input layer by adjusting and scaling the activations.

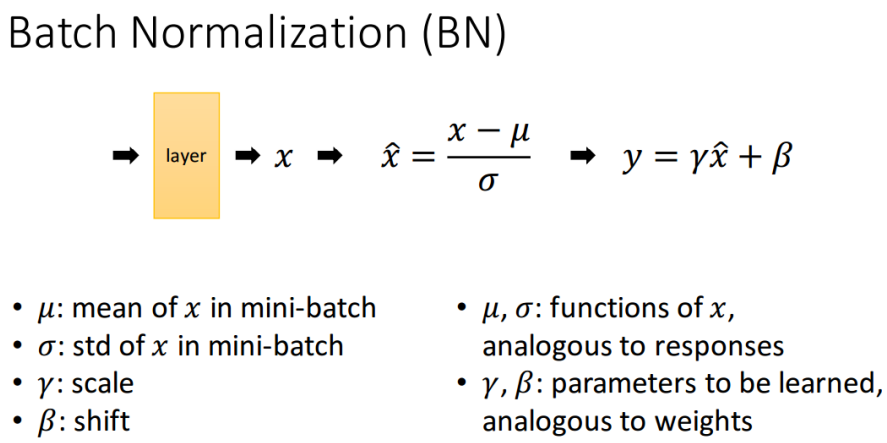

Figure 4.5: Batchnormalization

Dropout

Dropout prevents overfitting due to a layer's over-reliance on a few of its inputs. Because these inputs aren't always present during training (i.e. they are dropped at random), the layer learns to use all of its inputs, improving generalization.

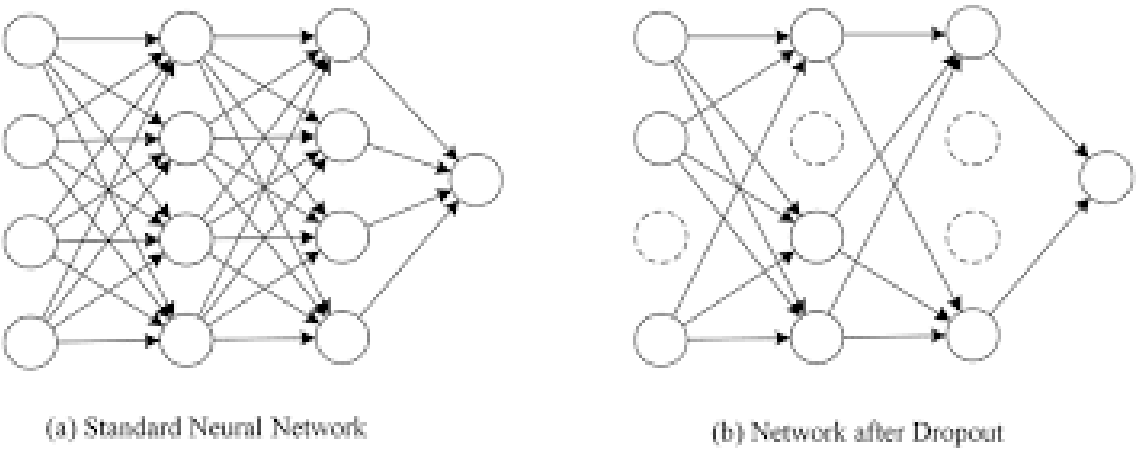

Figure 4.6: Dropout

Fully Connected layer

Fully connected layers connect every neuron in one layer to every neuron in another layer. The flattened matrix goes through a fully connected layer to classify the images. FC layers are used to detect specific global configurations of the features detected by the lower layers in the net. They usually sit at the top of the network hierarchy, at a

point when the input has been reduced (by the previous, usually convolutional layers) to a compact representation of features.

Fully convolution network

For segmentaion pupose fully connected layer is replaced with decoder network which consists of several deconvolution and concatenation layers and is known as Fully Convolution neural network.

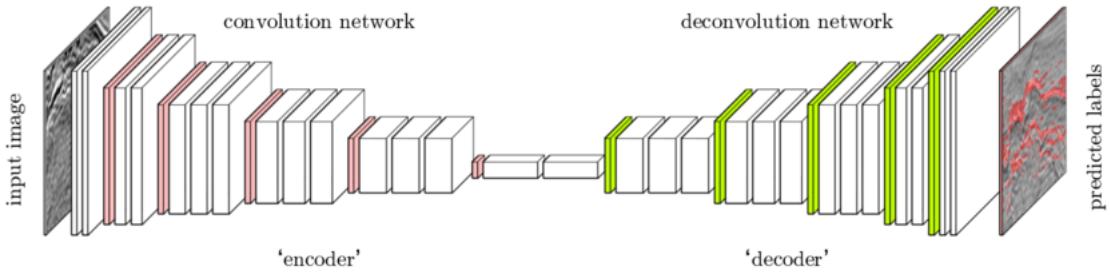

Figure 4.7: Fully convolution network

Deconvolution layer

Deconvolution layer is also known as transposed convolution layer. It upsample the down-sampled feature maps from CNN, producing the feature map that can be used to predict class labels at each pixel level.

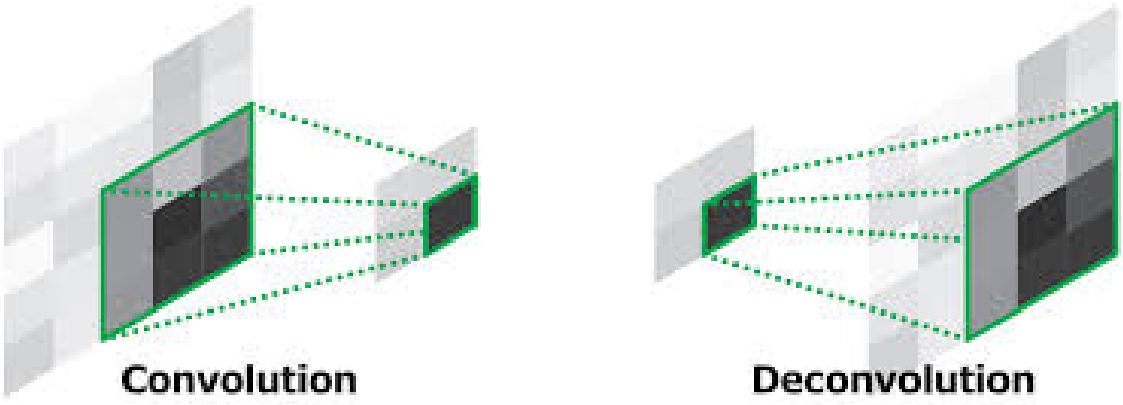

Figure 4.8: Convolution and Deconvolution

Concatenation layer

A concatenation layer takes inputs and concatenates them along a specified dimension. The inputs must have the same size in all dimensions except the concatenation dimension. Specify the number of inputs to the layer when you create it.

CHAPTER 5

ATTENTION NETWORKS

What is Attention?

Attention describes the ability to allocate consideration unevenly across a field of sensation, thought and proprioception, to focus and bring certain inputs to the fore, while ignoring or diminishing the importance of others. Use of attention mechanism in neural network can help in understanding what it is referring to and how to disregard the noise and focus on what is relevant. Attention networks are a kind of short-term memory that allocates attention over input features they have recently seen. Attention modules are mostly combined with rnn and cnn network. These type of networks are helpful when we are dealing with large image or other kind of data and need to focus on only certain part and rest can be ignored. Attention models can relieve computational burden.

Types of attention networks:

- Hard attention network
- Soft attention network

Hard Attention Network

In hard attention, we crop region of interest on which we want network to focus and only process the selected regions at high resolution. Hard attention network is non-differentiable. Therefore in order to train hard attention network, reinforcement learning method is used. Recurrent attention model is a kind of hard attention model.

Recurrent attention model

This model is built around recurrent neural network. At each time step, it processes the sensor data, integrates information over time, and chooses how to act and how to deploy its sensor at next time step.

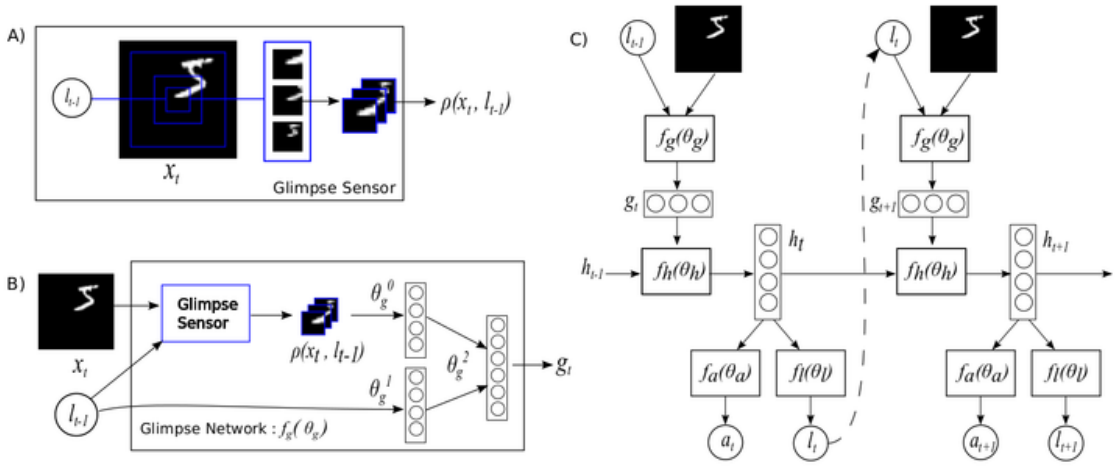

Figure 5.1: Recurrent attention model (Mnih *et al.*, 2014)

Glimpse sensor

At each time step, given the coordinates of the glimpse and an input image, the sensor extracts a retina-like representation centered at given location that contains multiple resolution patches.

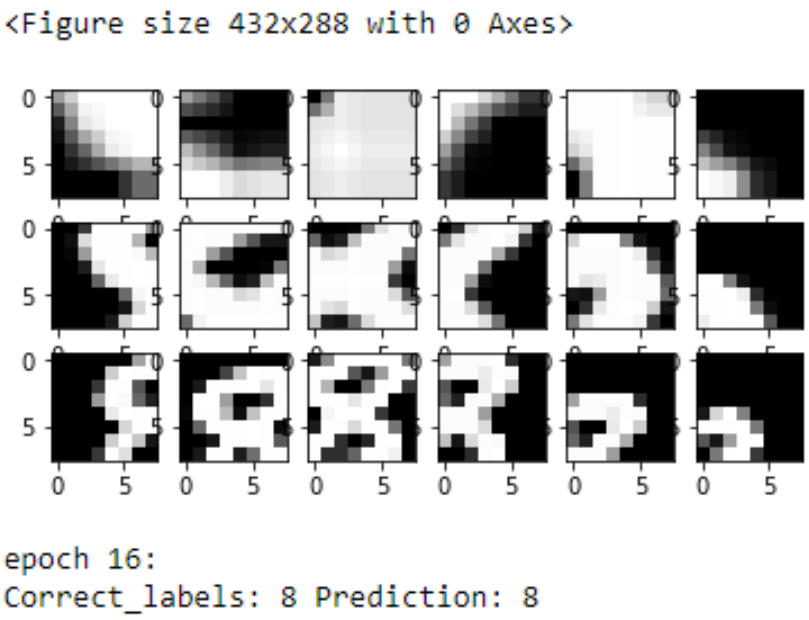

Figure 5.2: 6 Glimppses of 3 different scales taken in one iteration

Glimpse network

Given the location computed by location network and the input image, glimpse network uses glimpse sensor to extract patches at given location. The glimpses and location obtained are further encoded into feature vectors using several hidden layers and further both feature vectors are combined followed by another linear layer. It produces the glimpse representation which is then fed to RNN.

Overall model architecture

Overall, the model is based on RNN. The core network of the model takes the glimpse representation as input and combining with the internal representation at previous time step, produces the new internal state of the model. The location network and the action network use the internal state of the model to produce the next location to attend to and the action/classification at respectively. This basic RNN iteration is repeated for a variable number of steps.

Model training

The parameters of the reinforcement agent are given by the parameters of the glimpse network, the core network and the action network and these parameters are learnt in a way to maximise the total reward agent can expect when interacting with environment. In RL, everytime agent predicts correct label, +r reward is added while for every wrong decision 0 reward is given. Reinforcement learning in this case involves running the agent with its current policy to obtain samples of interaction sequences and then adjusting the parameters of our agent such that the log-probability of chosen actions that have led to high cumulative reward is increased, while that of actions having produced low reward is decreased.

RAM for classification

We trained MNIST dataset and blood cell dataset on recurrent attention model and test its performance.

Table 5.1: Test result of cluttered MNIST dataset and blood cell type classification

Dataset	*Size of image*	*size of glimpse*	*Glimpses*	*Accuracy*
Cluttered MNIST	$(64, 64)$	$(12, 12)$	6	98.2
Blood cell images	$(240, 320)$	$(30, 40)$	16	91.77

Soft attention network

soft attention, which multiplies features with a (soft) mask of values between zero and one. With soft attention, we multiply attention map over the image feature map (produce by feeding the image through a convolutional neural network) and sum it up. This makes features in the focused regions (the bright regions) dominate other irrelevant features in that time-step (the dark regions).

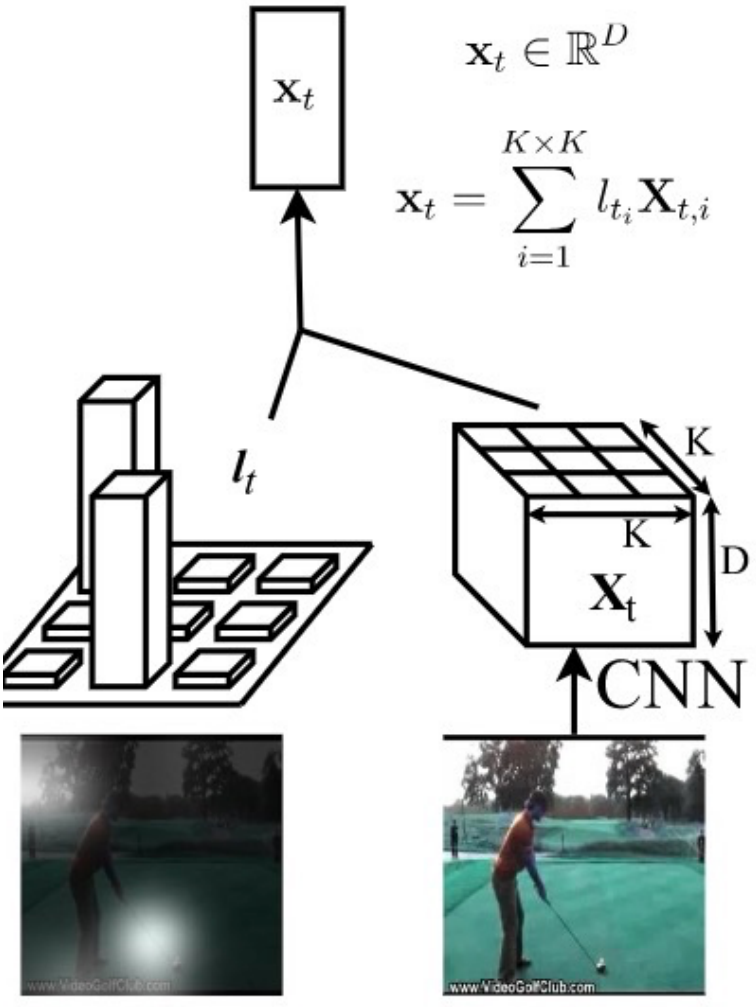

Figure 5.3: Soft attention mechanism

Unlike hard attention, soft attention is differentiable. In soft attention network, all parts were fully differentiable that enables the optimization of the whole model end-to-end by using gradient descent and trained with standard back propagation. Here in this thesis, we have used soft attention network for brain tumor segmentation which is explained later.

CHAPTER 6

MEDICAL IMAGE ANALYSIS APPLICATION: BRAIN TUMOR SEGMENTATION

Brain Tumor

A brain tumor is an abnormal mass of tissue in which cells grow and multiply uncontrollably, seemingly unchecked by the mechanisms that control normal cells. The two main groups of brain tumors are termed primary and metastatic. In primary ones, the origin of the cells are brain tissue cells, where in metastatic ones cells become cancerous at any other part of the body and spread into the brain. Gliomas are type of brain tumors that originate from glial cells. Brain tumor segmentation research mainly focuses on them. Gliomas can be low grade (slow growing) or high grade (fast growing).

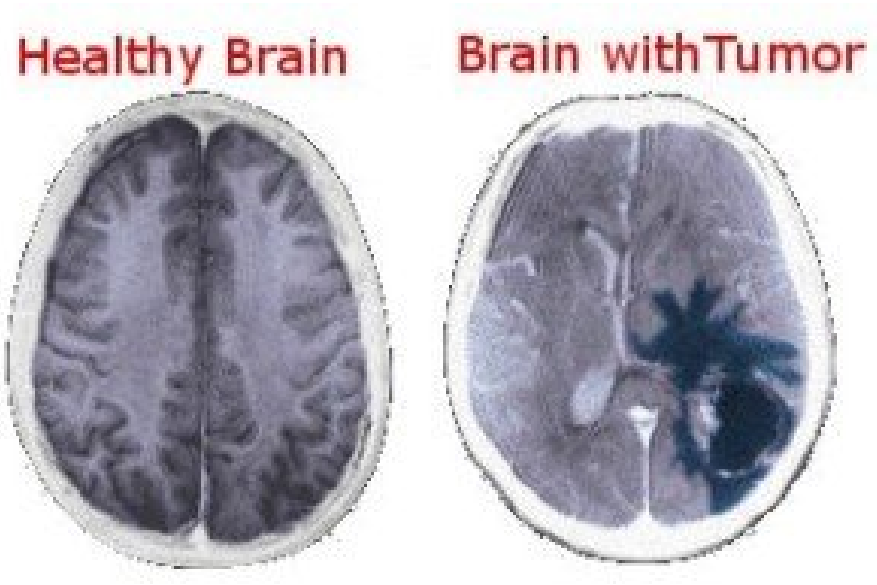

Figure 6.1: Healthy brain Vs Brain containing tumor

Medical Imaging Technique: MRI

Early diagnosis of brain tumor can improve treatment. There are many medical imaging techniques such as Computed Tomography (CT), Single-Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET), Magnetic Resonance Spectroscopy (MRS) and Magnetic Resonance Imaging (MRI) used to provide information about shape, size, location and metabolism of brain tumors. Among these techniques, MRI is considered as standard technique.

Magnetic resonance imaging is a medical imaging technique used in radiology to form pictures of the anatomy and the physiological processes of the body in both health and disease. MRI provides exquisite detail of brain, spinal cord and vascular anatomy, and has the advantage of being able to visualize anatomy in all three planes: axial, sagittal and coronal.

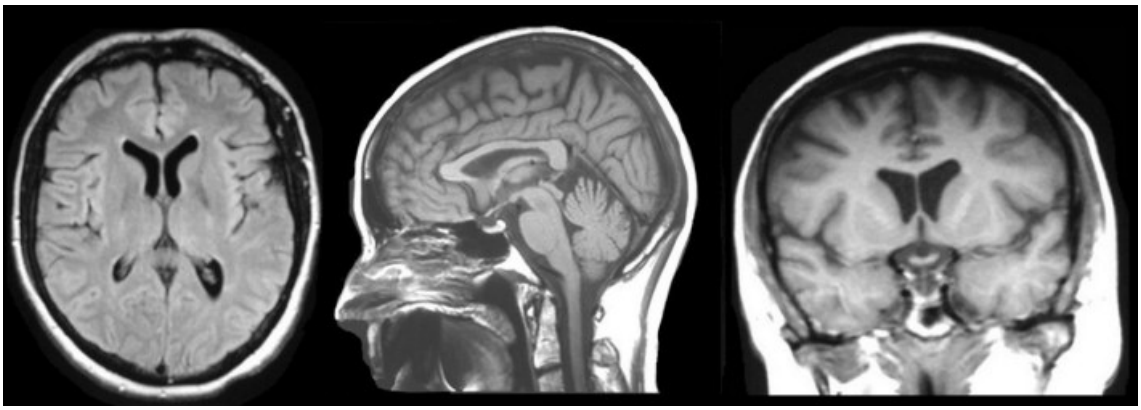

Figure 6.2: Brain MRI in axial, sagittal and coronal plane

MRI scanners use strong magnetic fields, magnetic field gradients, and radio waves to generate images of the organs in the body. Images of different MRI sequences are generated by altering excitation and repetition times during image acquisition. These different MRI modalities produce different types of tissue contrast images, thus providing valuable structural information and enabling diagnosis and segmentation of tumors along with their subregions. Four standard MRI modalities used for glioma diagnosis include T1-weighted MRI (T1), T2-weighted MRI (T2), T1-weighted MRI with gadolinium contrast enhancement (T1c) and Fluid Attenuated Inversion Recovery (FLAIR). During MRI acquisition, although can vary from device to device, around one hundred and fifty slices of 2D images are produced to represent the 3D brain volume.

Segmentation methods

Medical image segmentation for detection of brain tumor from the magnetic resonance images or from other medical imaging modalities is a very important process for deciding right therapy at the right time because the earlier the detection, the faster the treatment can be started. Healthy brains are typically made of 3 types of tissues: the white matter, the gray matter, and the cerebrospinal fluid.

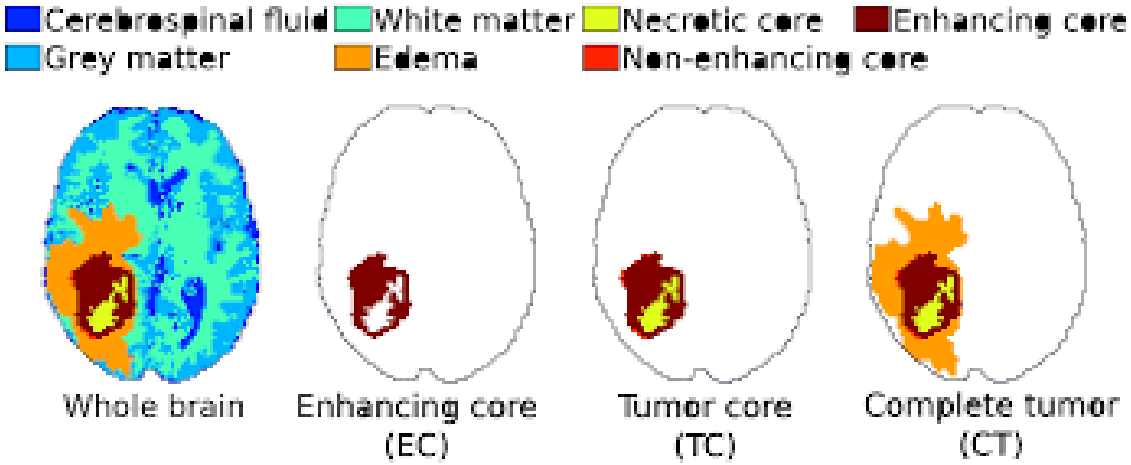

Figure 6.3: Left to right: whole brain, enhanced tumor, core tumor, complete tumor

The goal of brain tumor segmentation is to detect the location and extension of the tumor regions, namely active tumorous tissue (vascularized or not), necrotic tissue, and edema (swelling near the tumor). This is done by identifying abnormal areas when compared to normal tissue. Brain tumor segmentation can be done manualy, by using semi-automatic segmentation methods and fully automatic segmentation methods. The most effective method is fully automatic segmentation based on neural network. We are using 3D UNET and 3D UNET attention network for segmentation purpose.

Dataset

BRATS 2018 dataset is used for training the model. This dataset contains 210 HGG and 75 LGG images. All BraTS multimodal scans are available as NIfTI files (.nii.gz) and describe a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes, and were acquired

with different clinical protocols and various scanners from multiple (n=19) institutions. Annotations comprise the GD-enhancing tumor (ET label 4), the peritumoral edema (ED label 2), and the necrotic and non-enhancing tumor core (NCR/NET label 1). Size of BRATS 2018 dataset images are 240x240x155. The first preprocessing step we have done is to crop the images from sides to remove black regions which reduces the size to 192x192x144 and then resized images to 144x144x144. Further normalised the images to have values between 0 and 1. In BRATS 2018 dataset, for label image, we combined 1,2,and 4 for whole tumor, 1 and 4 for core tumor and 4 for enhanced tumor.

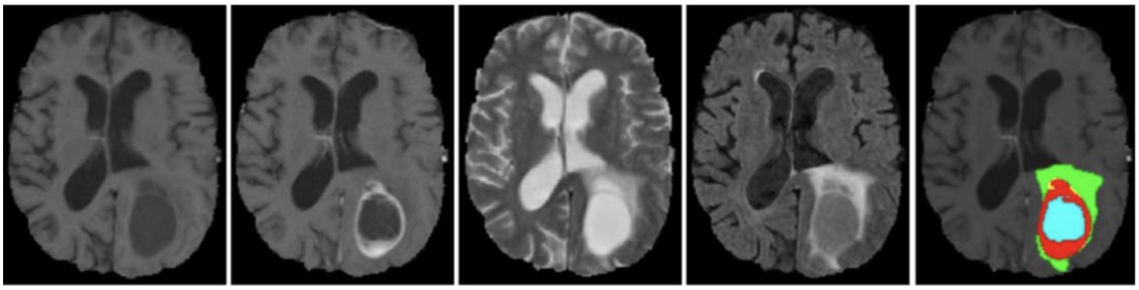

Figure 6.4: Left to right: T1, T1c, T2, Flair, Ground truth

CHAPTER 7

PROPOSED MODEL

Introduction

U-Nets are commonly used for image segmentation tasks because of their good performance and efficient use of GPU memory. The latter advantage is mainly linked to extraction of image features at multiple image scales. Coarse feature-maps capture contextual information and highlight the category and location of foreground objects. Feature-maps extracted at multiple scales are later merged through skip connections to combine coarseand fine-level dense prediction. U-Net architecture consists of encoder part to analyze the whole image and a successive decoder part to produce a full-resolution segmentation map.

3D unet model takes 3D input and further processes them with 3D convolution, max-pooling and deconvolution layers. Here we propose 3D U-Net attention model for brain tumor segmentation. In 3D U-Net attention model, we use attention gates in decoder part. AGs automatically learn to focus on target structures without additional supervision. At test time, these gates generate soft region proposals implicitly on-the-fly and highlight salient features useful for a specific task. In return, the proposed AGs improve model sensitivity and accuracy for global and dense label predictions by suppressing feature activations in irrelevant regions. AG parameters can be trained with the standard back-propagation updates without a need for sampling based optimisation methods as used in hard-attention. Attention gate module is shown in figure below:

Attention gate computes attention coefficients for each pixel wise feature vector of activation map of chosen layer and produce the feature vectors which are scaled by corresponding attention coefficients. In standard CNN architectures, to capture a sufficiently large receptive field and thus, semantic contextual information, the feature-map is gradually downsampled. The features on the coarse spatial grid level identify location of the target objects and model their relationship at global scale. Thus in order to

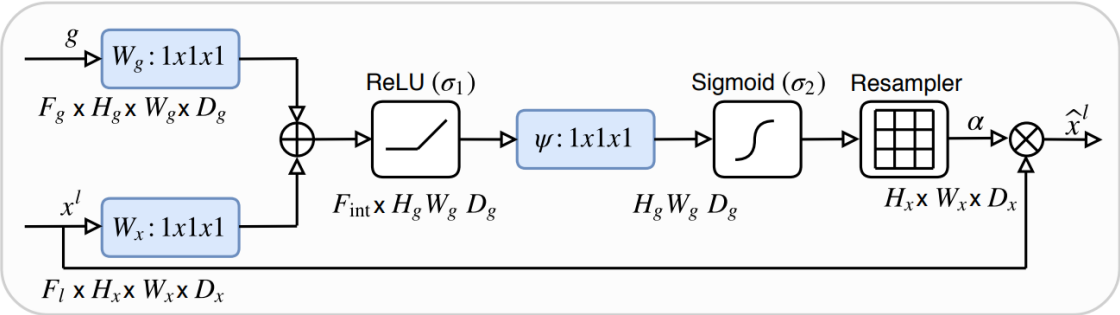

Figure 7.1: Attention gate block (Oktay *et al.*, 2018)

calculate attention coefficients, gating signals are used which are global feature vectors and provide information to AGs to disambiguate task-irrelevant feature content in activation map. AG module is used in standard U-Net architecture to highlight salient features that are passed through the skip connections. AGs filter the neuron activations during the forward pass as well as during the backward pass. Gradients originating from background regions are down weighted during the backward pass. This allows model parameters in shallower layers to be updated mostly based on spatial regions that are relevant to a given task. For AGs, we chose sigmoid activation function for normalisation.

Network Architecture

3D attention U-Net network consists of encoder having 4 repeated layers of conv -> batch normalisation -> relu -> conv -> batch normalisation -> relu -> pooling with 16, 32, 64, 128 filters respectively. Decoder consists of 3 deconvolution block concatenated with 3 attention gate block and 1 simple deconvolution block. In the last layer a 1x1x1 convolution reduces the number of output channels to the number of labels. The input to the network is a 144x144x144 image with 4 channels. Our output in the final layer is 144x144x144 with channels equal to number of labels. Brats dataset is trained on both 3D U-Net model and 3D attention U-Net model and both results are compared to see how efficient is later.

3D U-Net attention model

3D U-Net attention model is given below:

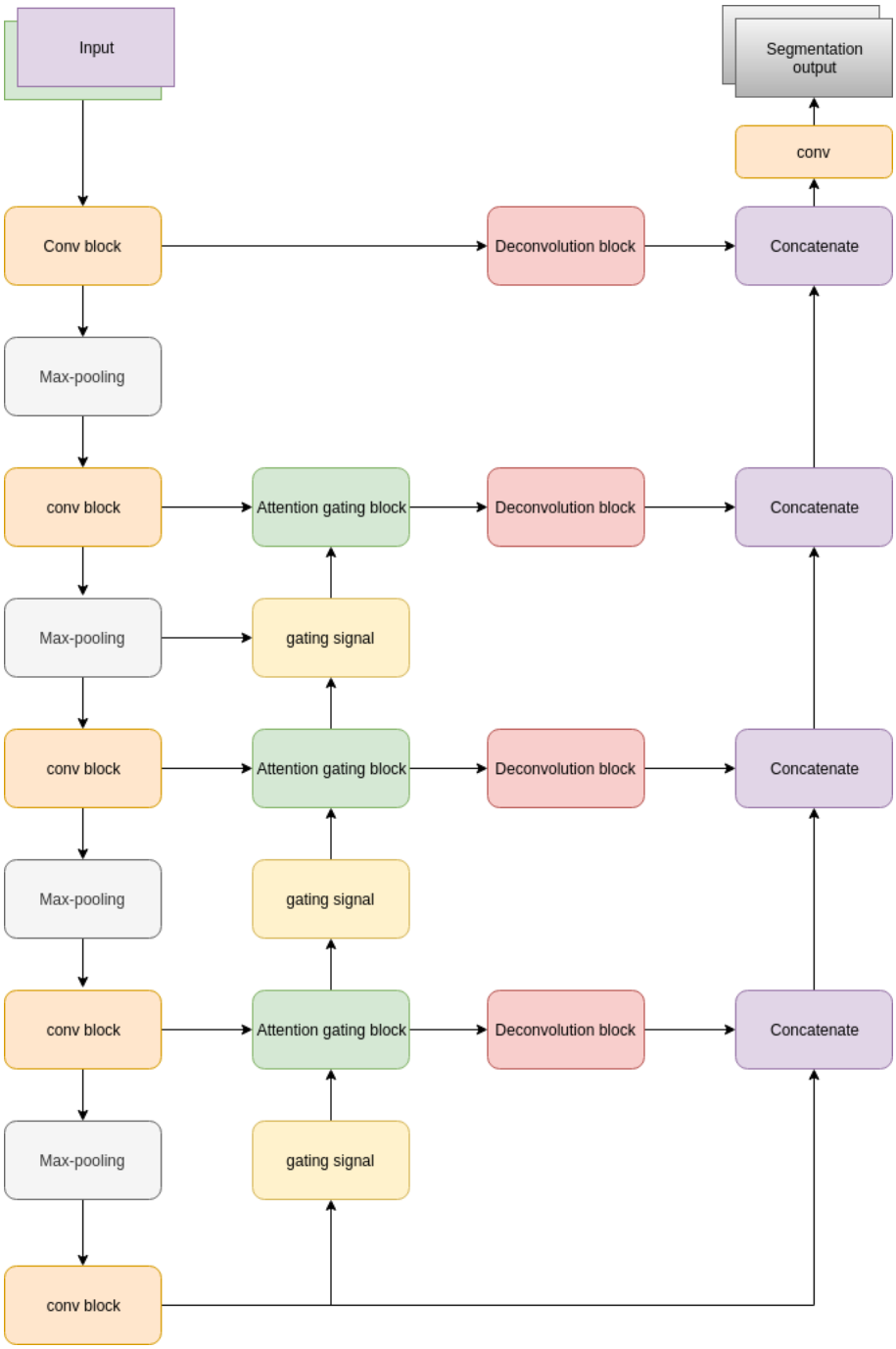

Figure 7.2: 3D U-Net attention network architecture

Training

3D attention U-Net model is trained on BRATS 2018 dataset. 80 percent data is used as training set and 20 percent is used as test dataset. Model is trained using adam optimizer with learning rate 0.0001 and decay rate of 0.00001. Loss is calculated using both cross entropy and dice-score. Since positives are very few in comparison to negatives, accuracy is not a correct performance metric. In this case F1-score which is same as dice-score is used as performance metric.

$$precision = \frac{tp}{tp + fp} \tag{7.1}$$

$$recall = \frac{tp}{tp + fn} \tag{7.2}$$

$$dicescore = \frac{2 * precision * recall + epsilon}{precision + recall + epsilon} \tag{7.3}$$

$$loss = 0.001 * binarycrossentropy(pred, label) - dicescore(pred, label) \tag{7.4}$$

CHAPTER 8

RESULTS

Training dataset : 190 HGG patients and 60 LGG patients

Test dataset : 20 HGG patients and 15 LGG patients

Table 8.1: Test result for segmentation of whole tumor using 3D U-Net, 2D attention U-Net network and 3D attention U-Net network (HGG)

$Model$	$Accuracy$	$Precision$	$Recall$	$Dice-score$
$3D\ U-Net$	99.5	82.99	81.28	82.26
$2D\ attention\ U-Net$	99.63	89.9	80.01	84.45
$3D\ attention\ U-Net$	99.52	84.09	85.8	84.94

Table 8.2: Test result for segmentation of whole tumor using 3D attention U-Net model

$Tumor$	$Accuracy$	$Precision$	$Recall$	$Dice-score$
HGG	99.52	84.09	85.8	84.94
$HGG+LGG$	99.5	83.49	80.29	81.86

Table 8.3: Test result for segmentation of core tumor using 3D attention U-Net model

$Tumor$	$Accuracy$	$Precision$	$Recall$	$Dice-score$
HGG	99.8	77.93	79.64	78.8
$HGG+LGG$	99.61	74.52	76.83	75.66

Table 8.4: Test result for segmentation of enhanced tumor using 3D attention U-Net model

$Tumor$	$Accuracy$	$Precision$	$Recall$	$Dice-score$
HGG	99.86	77.427	72.35	74.8
$HGG+LGG$	99.87	73.67	58.07	64.95

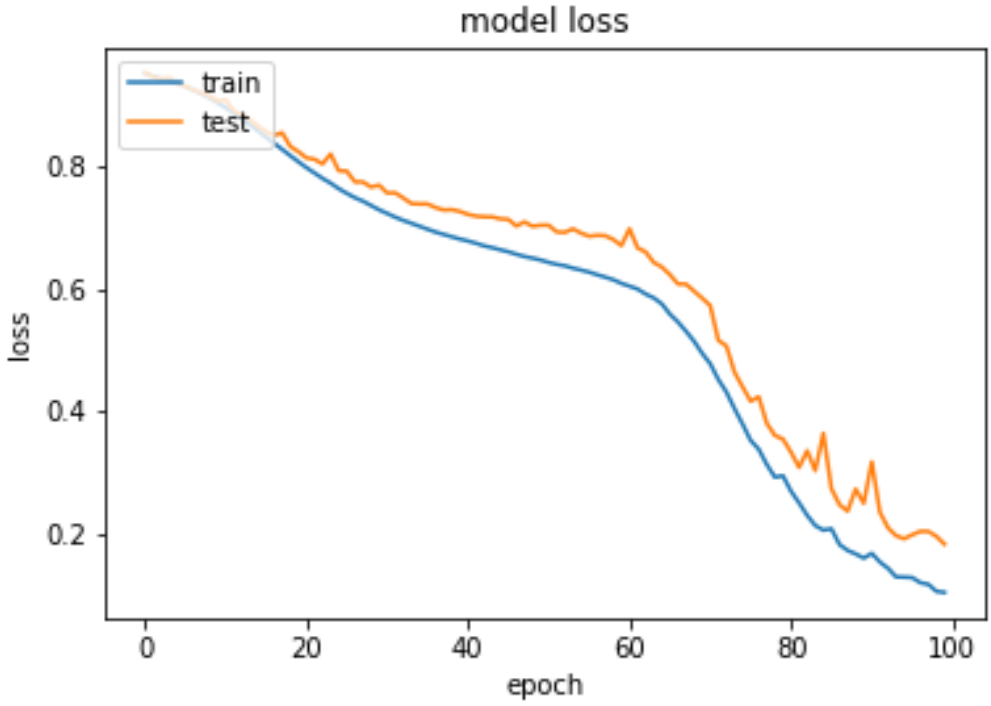

Figure 8.1: Plot of loss for whole tumor

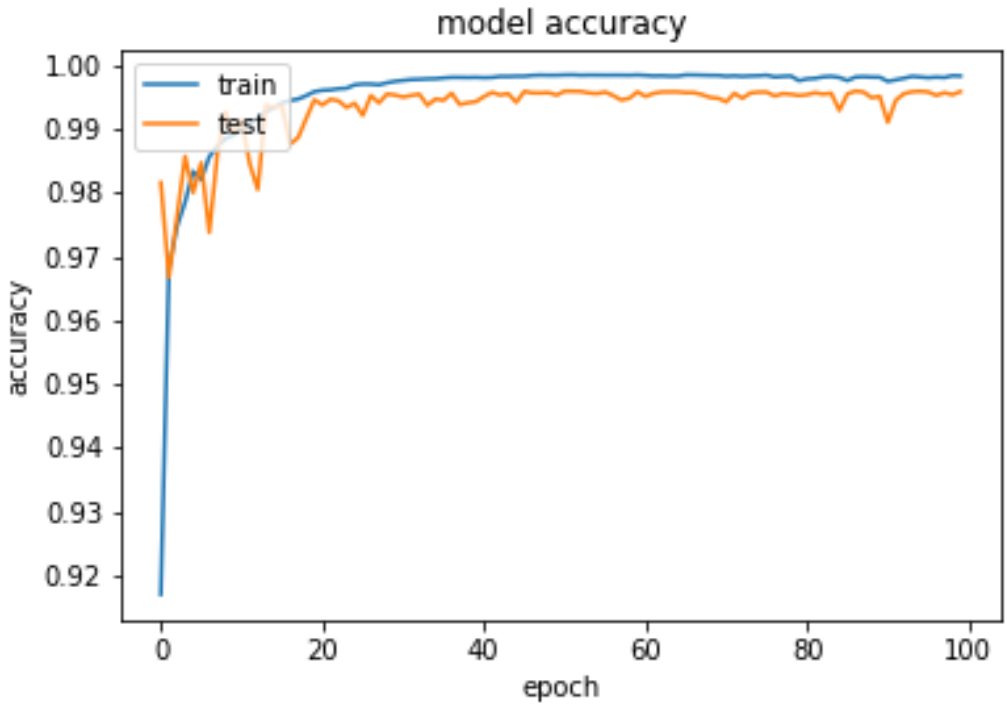

Figure 8.2: Plot of accuracy for whole tumor

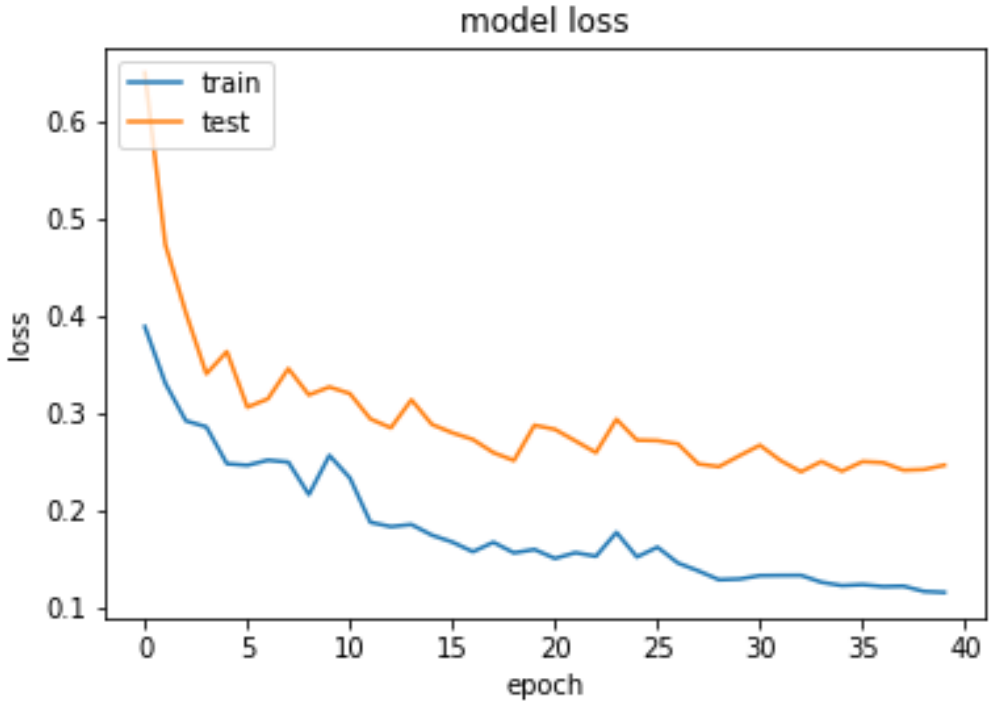

Figure 8.3: Plot of loss for core tumor

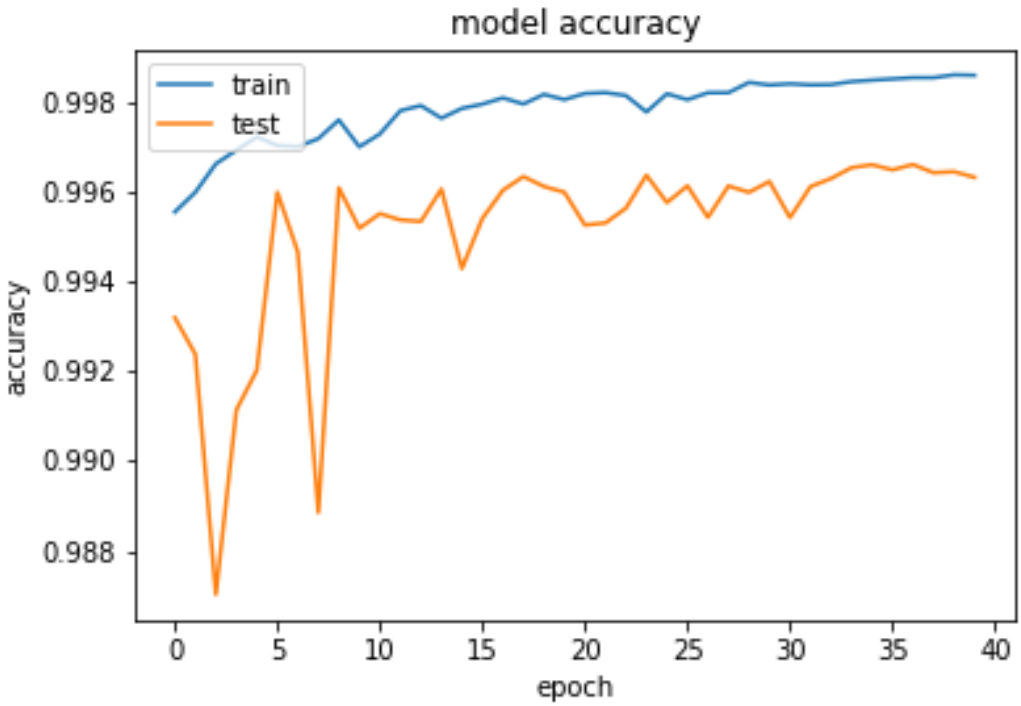

Figure 8.4: Plot of accuracy for core tumor

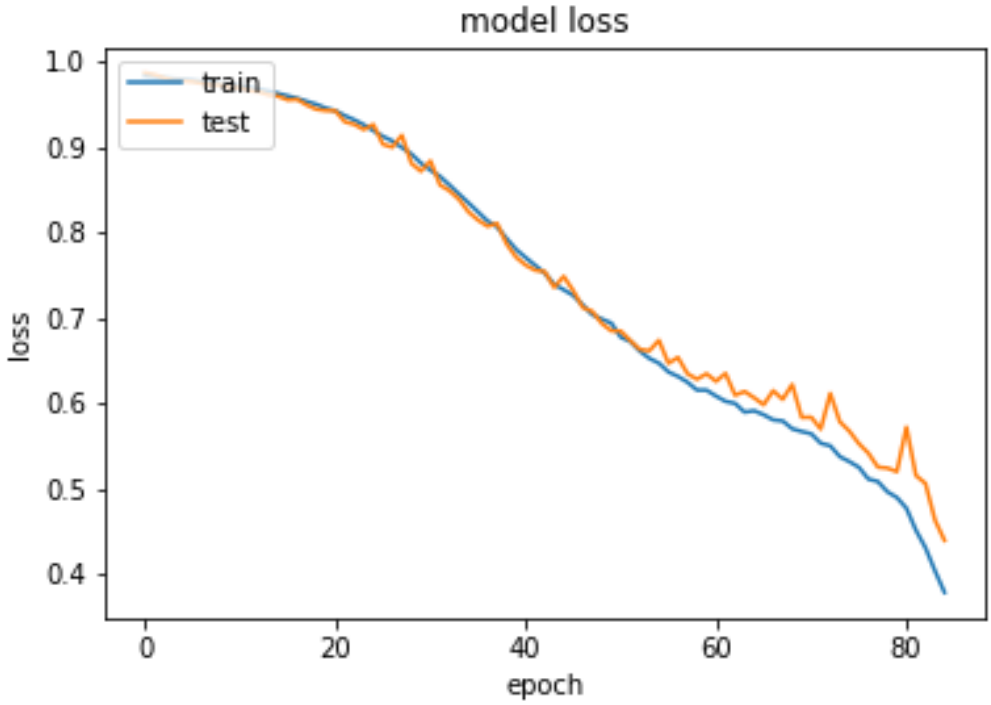

Figure 8.5: Plot of loss for enhanced tumor

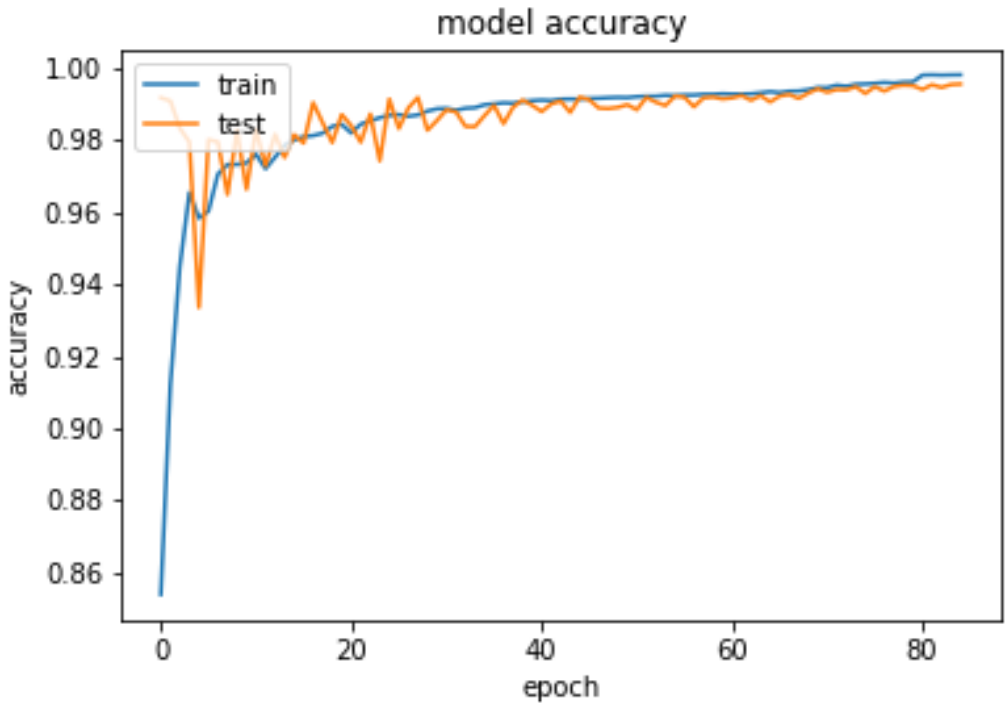

Figure 8.6: Plot of accuracy for enhanced tumor

Figure 8.7: segmentation result of whole tumor (HGG), left to right: actual label, pediction

Figure 8.8: segmentation result of whole tumor (LGG), left to right: actual label, pedic-
tion

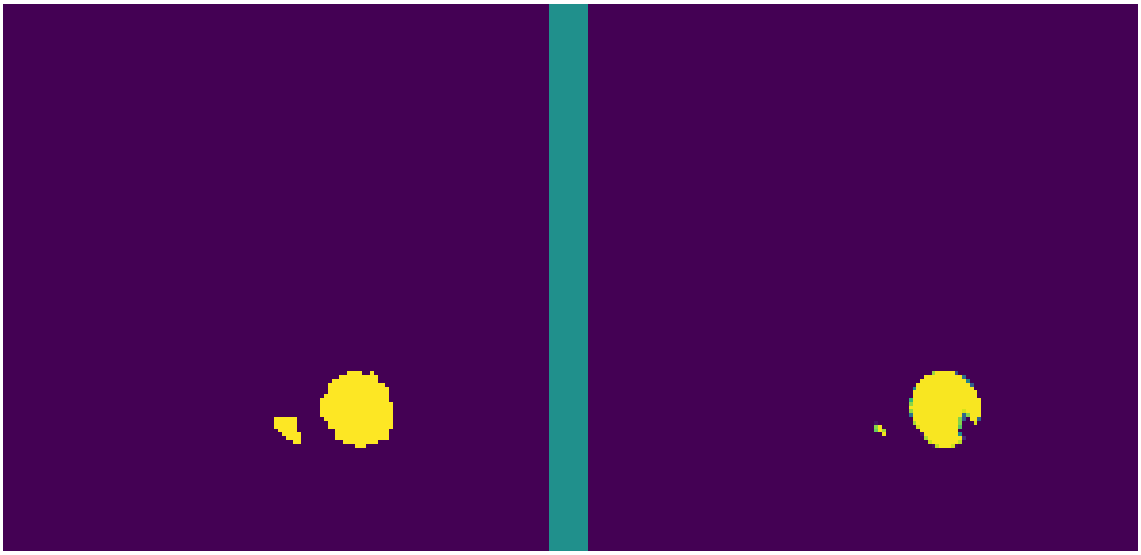

Figure 8.9: segmentation result of core tumor (HGG), left to right: actual label, pedic-
tion

Figure 8.10: segmentation result of core tumor (LGG), left to right: actual label, pedic-
tion

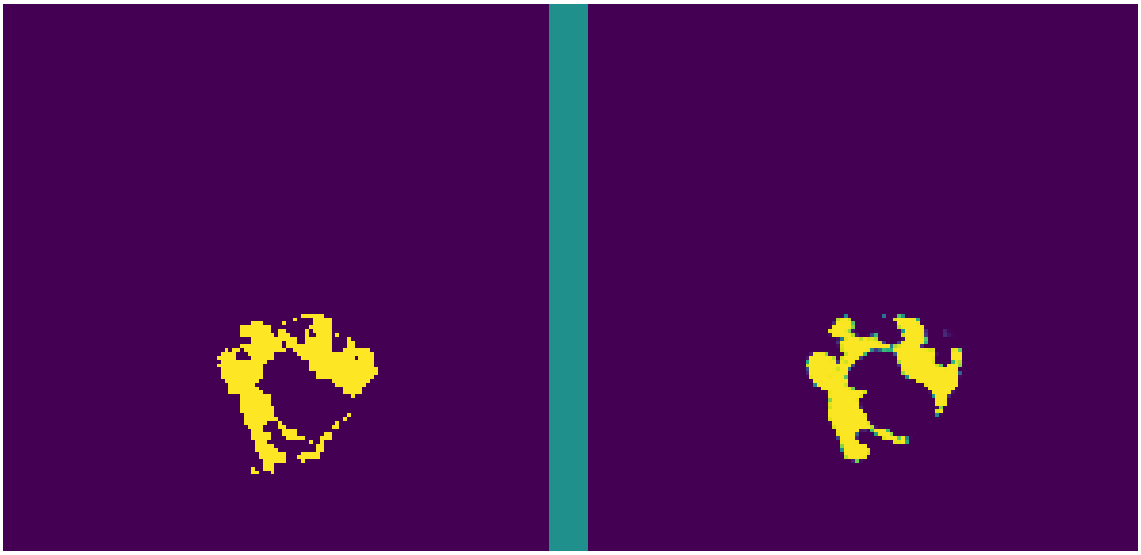

Figure 8.11: segmentation result of enhanced tumor (HGG), left to right: actual label,
pediction

CHAPTER 9

CONCLUSION

In this thesis, several types of attention networks have been studied and their implementation in medical image analysis especially in classification and segmentation. Further, an important field in medical imaging i.e. brain tumor segmentation is studied. Earlier brain tumor segmentation has been done using 2D networks and recently many researchers started using 3D networks. So 3D attention u-net network has been proposed in this thesis which is based on soft attention mechanism.Experimental results show that the proposed attention model is highly beneficial for object identification and localisation. This network performance is compared with 3D u-net network and 2D attention u-net network for whole tumor and it's dice-score is found to be better than others. Further 3D attention network is used for segmenting other type of tumor i.e. core tumor and enhanced tumor. Dice-scores and other performance parameters have reached state of art result..

CHAPTER 10

FUTURE SCOPE

One disadvantage of hard attention network is that it is non-differentiable. In soft attention, the blacked-out parts of the input do not contribute to the results but still need to be processed. It is also over-parametrised: sigmoid activations that implement the attention are independent of each other. Use of spatial transformer can solve above problems. Spatial Transformer (STN) allows differentiable image-cropping. It is made of two components: a grid generator and a sampler. The grid generator specifies a grid of points to be sampled from, while the sampler, well, samples. We can use spatial transformer to extract glimpses, encoder having several convolution layers to encode feature vector which can be passes to RNN and it's output is further decoded using deconvolution network and map it to size of label map using inverse spatial transformer which has been explored in paper. This network can be modified for 3D input and used for segmenting brain tumor from MRI images. We have to create 3D spatial transformer and 3D inverse spatial transformer.

REFERENCES

1. **Ablavatski, A.**, **S. Lu**, and **J. Cai**, Enriched deep recurrent visual attention model for multiple object recognition. *In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017.

2. **Havaei, M.**, **A. Davy**, **D. Warde-Farley**, **A. Biard**, **A. Courville**, **Y. Bengio**, **C. Pal**, **P.-M. Jodoin**, and **H. Larochelle** (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, **35**, 18–31.

3. **Kayalibay, B.**, **G. Jensen**, and **P. van der Smagt** (2017). Cnn-based segmentation of medical imaging data. *arXiv preprint arXiv:1701.03056*.

4. **Khotanlou, H.**, **O. Colliot**, **J. Atif**, and **I. Bloch** (2009). 3d brain tumor segmentation in mri using fuzzy classification, symmetry analysis and spatially constrained deformable models. *Fuzzy sets and systems*, **160**(10), 1457–1473.

5. **Mnih, V.**, **N. Heess**, **A. Graves**, *et al.*, Recurrent models of visual attention. *In Advances in neural information processing systems*. 2014.

6. **Momeni, A.**, **M. Thibault**, and **O. Gevaert** (2018). Deep recurrent attention models for histopathological image analysis. *BioRxiv*, 438341.

7. **Oktay, O.**, **J. Schlemper**, **L. L. Folgoc**, **M. Lee**, **M. Heinrich**, **K. Misawa**, **K. Mori**, **S. McDonagh**, **N. Y. Hammerla**, **B. Kainz**, *et al.* (2018). Attention u-net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

8. **Prastawa, M.**, **E. Bullitt**, **S. Ho**, and **G. Gerig** (2004). A brain tumor segmentation framework based on outlier detection. *Medical image analysis*, **8**(3), 275–283.

9. **Shen, C.**, **G.-J. Qi**, **R. Jiang**, **Z. Jin**, **H. Yong**, **Y. Chen**, and **X.-S. Hua** (2018). Sharp attention network via adaptive sampling for person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*.

10. **Zikic, D.**, **Y. Ioannou**, **M. Brown**, and **A. Criminisi** (2014). Segmentation of brain tumor tissues with convolutional neural networks. *Proceedings MICCAI-BRATS*, 36–39.