Lossless Segmentation of Brain Tumors from MRI Images using 3D U-Net

by

Ramisa Farha
18101406
Nigar Sultana Nuha
18101143
Syed Nazmus Sakib
18101160
Sowat Hossain Rafi
18101140
Md Sabbir Khan
18101274

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degrees of Bachelor of Science in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
January 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing the undergraduate degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material that has been accepted or submitted for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Ramisa Farha

Ramisa Farha
18101406

Nigar Sultana Nuha.

Nigar Sultana Nuha
18101143

Syed Nazmus Sakib

Syed Nazmus Sakib
18101160

Sowat Hossain Rafi
18101140

Md. Sabbir Khan

Md Sabbir Khan
18101274

Approval

The thesis/project titled "Lossless Segmentation of Brain Tumors from MRI Images Using 3D U-Net" submitted by

1. Ramisa Farha (18101406)
2. Nigar Sultana Nuha (18101143)
3. Syed Nazmus Sakib (18101160)
4. Sowat Hossain Rafi (18101140)
5. Md Sabbir Khan (18101274)

Of Fall 2021 has been accepted by the Department of Computer Science and Engineering as satisfactory partial fulfillment of the requirements for the degrees of Bachelor of Science in Computer Science and Engineering on January 2022.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Co-Supervisor:
(Member)

Tanzim

Md Tanzim Reza
Lecturer
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Program Coordinator:
(Member)

Dr. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Abstract

2D computer vision and activities related to medical image analysis are remarkably guided with the help of Convolutional Neural networks (CNNs) in recent years. Since a chief portion in the available clinical imaging data is in 3D, we are inspired to further develop 3D CNNs for seeking the advantage of greater spatial context. Despite the fact that many FCNs are previously worked on and built by using various approaches, current 3D approaches still rely on patch processing due to the utilization of GPU memory, which limits the incorporation of bigger context information for improved performance. Using efficient 3D FCNs in MRI images without any data loss would result in more efficient disease detections. In this paper, we propose an approach to an efficient 3D U-net segmentation technique for MRI Images using a lossless preprocessing of an MRI image dataset. Our proposal has the advantage of an impressive reduction of the required GPU memory for 3D Medical Image processing activities and that too, with an enhanced performance which is evaluated by the IoU (Intersection over Union) evaluation metric. Comprehensive experiment results performed with MICCAI BraTS'20 exhibit the viability of the presented strategy.

Keywords: 3D CNN, FCNs. 3D-Unet, segmentation, volumetric medical images, 3D medical image processing, Brain 3D MRI.

Dedication

We dedicate our research work for developing various medical image analysis in the future as well as to aid in the field of computer vision. Our study findings may be valuable in the future medical field. The code of the model will be made publicly available for further uses in case of semantic segmentation.

Acknowledgement

First and importantly, we acknowledge and praise Almighty Allah for allowing us to complete our thesis without major impediments.

Secondly, we would like to express our heartfelt gratitude, Dr. Md. Ashraful Alam Sir, for his excellent guidance and assistance during our research. Also, Md Tanzim Reza Sir, our Co-Supervisor. Their advice helped us all the way from the beginning to the conclusion.

Thirdly, we owe our parents our deepest appreciation; without their selfless encouragement, we would not have progressed as far as we have.

Most importantly, our remarkable teamwork has aided us cross the finish line.

Last but not least our best regards to the institution, Center for Biomedical Image Computing Analytics for the Data Set MICCAI 2020 BraTS for making this dataset public for research purposes. Also thanking Dr. Sreenivas Bhattiprolu for his tutorials for helping us in developing the Unet model.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

$ADMN$ Advanced Design And Modeling

$BraTS$ Brain Tumor Image Segmentation

$CIFAR$ Canadian Institute for Advanced Research.

CNN Convolutional Neural Network

CRF Chronic Renal Failure

CT Computed Tomography

$DCNN$ Deep convolutional neural network

DSC Dice score coefficient

DSN Data Source Name

DUC Digital Up Converter

FCN Fully-Connected Network

GDL Gradient Difference Loss

GPU Graphics Processing Unit

HGG High Grade Glioma

IoU Intersection over Union

$KITTI$ Karlsruhe Institute of Technology and Toyota Technological Institute

LGG Low Grade Glioma

$MICCAI$ Medical Image Computing and Computer Assisted Intervention Society

MR Magnetic Resonance

MRI Magnetic Resonance Imaging

NCR Necrotic Core

NLP Natural language processing

$NORB$ NYU Object Recognition Benchmark

$P-NET$ Primitive Neuro-Ectodermal Tumors

RAM Random access memory

$TCGA-GBM$ The Cancer Genome Atlas Glioblastoma Multiforme

$TCGA-LGG$ The Cancer Genome Atlas Low Grade Glioma

$TICA$ The Cancer Imaging Archive

VGG Visual Geometry Group

VOI Volume-Of-Interest

Chapter 1

Introduction

1.1 Overview

Convolutional Neural Networks or CNNs, are recently taking advantage in solving problems within the fields of Computer-Vision along with Medical Image Analysis [1]. In the arena of medical image computing, CNNs have been used in an increasing number of applications during the past few years. Despite such acceptance, most of the techniques are just capable of processing 2D images, in spite of the fact that the great majority of medical imaging data obtained is in 3D. Data obtained from 3D imaging modalities such as (micro-CT or X-ray) Micro-Computed Tomography , (CT) Computed Tomography , or (MRI) Magnetic Resonance Imaging scanners are labeled with 3D image segmentation to extract regions of interest. Thus, they can profit from greater geographical context, by increasingly developed 3D CNNs [2] [3] [1]. (CNNs) Convolutional Neural Networks can be considered as a strong tool in terms for discovering visual representations from pictures, often consisting of several layers of nonlinear processes along with a huge portion of parameters that are trainable [4]. Despite the fact that immediate measurement as well as analysis of 3D pictures are possible for some cases, segmented-images form the foundation for the majority of 3D Image analysis. The extraction of a region of interest's geometry by 3D picture segmentations allow 3D model conversions, allowing viewing and measurement of the topics that are scanned . Further, the virtual examinations of these model, such as through simulations in computer, or even creating a physically represented subject via 3D printing, all necessitate the completion of segmentation on the 3D pictures.

A prejudiced hierarchical characteristic is noted when CNNs were trained [3].[5] depicts a survey done on GPU-based Medical Image computation techniques like segmentation, visualization and registration. It discussed how the GPU's computation speed has significantly grown, allowing the provision of a significant acceleration for many computationally-intensive tasks as compared to the traditional GPU-based computing frameworks. Because of its huge processing capabilities, GPU has recently emerged as a competitive platform for higher performance in computing [5]. Due to the constraints from GPU memory as a result of a complete conversion to 3D, the state-of-the-art approaches rely on patch-processing / sub volume. The patch to be inputted is usually smaller in size if no specialized version of hardware is used along with plenty of GPU memory, which limits the synthesized large context information for a reasonable performance [2]. In the following paper, we are proposing an approach to an efficient segmentation technique for MRI Images using an efficient 3D-unet model after a thorough data preprocessing to minimize the data loss. Our approach is beneficial for proportionally reducing the memory of GPU for 3D Medical Image processing activities without losing any kind of informative data.

1.2 Motivation

Artificial intelligence (AI) advances, more individuals are attempting to use AI approaches to the automated-segmentation of the brain tumors in MRI image datasets. A (CNN) convolutional neural network is a kind of Artificial Neural Network used to interpret visual images in deep learning. Contrary to popular belief, most convolutional neural networks are merely equivariant to translation, rather than invariant. They're used in recognition of image and video, recommendation systems, image processing,[6] image classification and segmentation, medical image analysis, NLP, brain-computer interfaces and financial series data.CNNs (convolutional neural networks) are a sort of deep learning model [7][8] that may operate on original dataset directly. [9] In some past years, (CNNs) Convolutional Neural Networks have been used to drive 2D computer-vision and activities connected to medical image analysis. [10] We are inspired to continue developing 3D CNNs in order to gain the positive result of more spatial context, because a large amount of the accessible clinical imaging data is in 3D. Two-dimensional convolutional neural networks or CNNs that predict the segmentation maps for MRI images in a single unique anatomical plane. Whereas, 3D-CNNs solve this problem by predicting segmentation for a volumetric region of a scan using 3D convolutional kernels. Fully convolutional networks are regarded as some higher class models addressing many pixel wise tasks. [11] FCNs are reviewed as the networks that do not contain dense layers like CNNs, rather they contain 1 x 1 convolutions and work like fully connected layers. In recent days, FCNs for semantic segmentation highly improved the accuracy by transfers of pre-trained classifier weights, layer representations, fusing different as well as learning end-to-end on entire data images models.Using efficient 3D FCNs in MRI images without any data loss would result in more efficient disease detections. Therefore, We present an approach to an efficient 3D U-net segmentation technique for MRI images employing a lossless preprocessing of the MRI image dataset in this paper. Our aim is to offer the benefit of a significant decrease in the amount of GPU RAM required for 3D Medical Image processing operations, as well as improved performance.

1.3 Problem Statement

In the medical field it is greatly important to be accurate when diagnosing or detecting something unusual from the volumetric images of different organs of a patient. Segmentation of volumetric images in [12] like 3D CT (Computed Tomography) images and MR (Magnetic Resonance) images are greatly important in case of diagnosis and treatment procedure. Medical Image analysis is generally the science to solve or analyze medical problems that are on the basis of several imaging modalities and digital images analyzing procedures. If not properly segmented everything might go wrong in the treatment of the patient and the patient might even lose their life. On the other hand, when the segmentation is highly accurate it helps in precise diagnosis and surgical planning as well as ensuring a fast and better surgical procedure[2].That is, it ensures a better treatment. Automated volumetric medical image segmentation is not something easy at all, because various organs have various shapes, sizes, and structures which we call inter-patient anatomical variability. Keeping the accuracy problem in our obvious consideration, we required something that would save time as well, causing no delay of treatment of any patient. That's why the segmentation technique needed to be efficient as well, and which can be done better using an efficient 3D U-net model. We already know convolutional neural networks or CNN already play a great role in [12], especially in 2D medical image segmentations. While 3D images are still a problem for segmentation due to organ structure and shape variance and for the process being complicated in case of 3D ones.

Therefore, 3D Fully Convolutional Networks could be used for an efficient segmentation of the volumetric images to solve these problems. This is generally a sequential down shuffling operation by 3D convolutions with lower resolution. As stated, our approach also has an advantage which will cause the 3D image processing task to significantly reduce the data loss. Since the loss of data is a great disadvantage in detecting a disease. During our research and experimentation, we are also going to get answers on our own for questions like how effective the performance of our approach would be using our approach in 3D medical image segmentation and if it is more accurate and better compared to most other previously introduced methods.

1.4 Aims and Objectives

The aim for our research is to propose an efficient segmentation technique for MRI images. Making the use of a 3D U-net model, to meet the purpose by systematically lowering the amount of GPU memory required for processing tasks on 3D medical images. Reducing the data loss through preprocessing, then using it as an input to make it go through subsequent FCN and so on for the volumetric image analysis. The objectives of the research are-

- To process a highly accurate segmentation technique.
- To reduce the GPU memory that is generally demanded for 3D medical images.
- To compare the experiment results with the previously done research papers.
- To declare an efficient procedure for 3D image analysis with an improved performance.

Chapter 2

Background And Related Works

Glioma, a form of brain tumor that is highly common and makes for around 33% of all brain tumors. They mainly affect the central nervous system. And it is a deadly disease, a kind of cancer with a very terrible prognosis along a survival rate much less than about 2 years.Segmenting, or can be said defining the pixels that correspond to the tumor is an important diagnostic procedure. It's usually done manually by a professional radiologist, but it takes a long time, and manual segmentation is impractical if we have a large number of patients and want to do anything like a meta analysis or something. So, in order to supplement a radiologists' efforts, a Volumetric image segmentation, like 3D CT (Computed Tomography) and MR (Magnetic Resonance) images, that can successfully separate gliomas can be quite beneficial to us. Magnetic resonance imaging, often known as MRI, Magnetic Resonance, or MR imaging, is commonly used to monitor patients of Brain Tumours. Medical image analysis is a science that focuses on using a variety of imaging techniques and digital image analysis technologies to solve or assess medical diagnosis. As a result, rather than acquiring a single image, volumes are generally recorded. MR image volumes, also known as MRI image volumes, are routinely acquired. MRI is a non-invasive imaging technology that does not employ ionizing radiation like most other imaging techniques. If medical images are not properly segmented anything might go wrong in the treatment of the patient and the patient might even lose their life. On the other hand, when segmentation is very accurate, it aids in precise diagnosis and surgical planning, as well as assuring a faster and more effective surgical operation. In other words, it ensures a better outcome.

[13] proposed a deep network that has learned to do volumetric dense segmentation from sparsely annotated volumetric images. It requires 2 Dimensional annotated slices for training. It can relevantly be applied to densify a sparsely annotated dataset, also to learn from sparsely annotated datasets to generalize to a new one. It is supported by an earlier U-net architecture which was 2 Dimensional in nature, while the network proposed here only carries 3 Dimensional volumes because the input and processes them with adapting 3 Dimensional procedures, specifically 3 Dimensional convolutions, 3 Dimensional max-pooling and 3 Dimensional up-convolutional layers. Since each picture contains repetitious structures with the related difference, this biomedical application needs only a few images to coach. The information for the structure is $132 \times 132 \times 116$ voxel tiles of a picture with three channels and output is $44 \times 44 \times 28$ voxels respectively in X, Y and Z axial

paths. The dataset used to teach the structure is Xenopus kidney, tiles stitched using XuvTools, used Slicer3D for manual annotation of orthogonal slices in each volume. The experiments were done on down-sampled editions of the initial resolution by an element of two in every dimension. So, the info sizes utilized in the trials are 248 × 244 × 64, 245 × 244 × 56 246 × 244 × 59 in X, Y, Z volumes. 70,000 training iterations were done on an NVIDIA TitanX GPU and it carried 3 days for training. This network used an end-to-end technique for understanding the semi-automatically and the fully-automatically done segmentation of a 3 Dimensioned volume from a limited annotation. Even for highly variable structures within the given image, it gives an accurate segmentation. The common IoU of 0.863 was achieved in three-fold cross-validation experimentation for the semi-automated format and within the fully-automated setup, there was an operation increase of the 3 Dimensioned architecture to constant 2 Dimensional implementations. The structure here is unoptimized and coaching from the beginning with no pre-training.

[14] introduce a three dimensional (CNN) for segmenting the brain tumors from the multimodal MR datasets of the brain. The model is a revised form of the well-known three dimensional U-net design, which accepts multi-modal MR data of the brain as input, evaluates those data at different scales then outputs a complete accurate multiclass tumor segmentation. A compensated (CCE) loss of function has been used to train the model from start to finish using the BraTS 2018 Training dataset. They use BraTS 2018 datasets to construct a revised version of the classic three dimensional U-net design for segmentation of tumor of the brain. Several medical diagnostic segmentation operations, including segmentation of organ and lesion, segmentation of retinal surface , and so on, have been effectively implemented using the U-net design. The three dimensional U-net is trained on BraTS 2018 training sample using the (CCE) loss of function, and a training on the class weights that is used to rectify unbalance problem. On the BraTS' 18 test and data validation, authors obtained successful performance with the Dice values for the test data and validation dataset for expanding tumor and entire tumor area. Researchers selected reversed convolution over regular upsampling because it permits the model to learn an appropriate approximation function rather than a predefined approximation function. After that, instance leveling and Dropout with a probability of 0.05 are applied once again. Finally, Two three dimensional convolutions with k multiply 2n filters of size 3x3x3 are performed. Each convolution layer uses the Rectified linear component as a non-linearity parameter. C filters make up the last layer, with C denoting the entire classifier. SoftMax nonlinearity comes after that.In the pre-processing stage, Isotropic, co-registered and skull-stripped MR volumes are included in the BraTS test. Then the author next proceeds to several pre-processing processes. The luminosity of the dimensions was re-scaled from 0 to 1 divided by the standard deviation, using mean subtraction, and cropped to 184x 200 x152. The findings show that the suggested technique operates well on entire tumors and tumor cores, but has a lesser performance for augmenting tumors. That is to be anticipated, as enhancing tumors depend significantly on T1c imaging and appear in the same way as other enhancements upon these pictures. Different methods aid in the segmentation of other tumor subtypes. However, their technique performed poorly on the test dataset of Enhancing Tumors Tumor Core categories.

[15] In this paper the author discusses all about the U-net. U-net is part of the neural network system for an image segmentation. They can do 2 ways, one is the analysis path and another one is the expansion path . Author say Ronneberger create a huge impact on U-Net improvement .author inform that U-Net is faster for train other segmentation because of context based learning.Firstly Auther state Base U-net architectures in this U-Net architectures there are two stage of U-Net segmentation .first one is contracting path and second stage is expansive path. In stage one it uses 3 x 3 convolutions by the Rei.U activation unit with max-pooling level . They duplicate those processes as much as the process needs. And in stage two U-Net uses 2 x 2 up-convolution for up samples.then it crops and adjusts by double 3 x3 convolution and Relu activation. After that 1 x 1 convolution is applied to maintain balance and give expected image outcome.after that writer gives the knowledge of 3D U-Net . Here all the 2D structures are removed and 3D structures are introduced. This 3D model mostly applies in the field of medicine to find out the illness of a patient. Then the author writes about Attention U-Net . In this architecture they use attention gates to focus on a specific region . They use an attention gate to find out what part of the image is needed.additive attention gate can give the best accuracy rate. In the list the author states about Inception U-Net. Most of the structure uses an exact size of image filter . but in most images inner objects have different dimensions . for this model to need high level information for the image process. In this method they use 3 x 3 convolution two times to get more detailed information . In the queue Author next explains about Residual U-Net model . if more layers are present it can give more accurate results but overload of layers also can damage the data and more valuable data can be lost . But ResNet can reduce this problem by bypassing some connections where feature maps take data from one layer to the next layer. They skip connections in between their convolutional layers. Author also state that for complex image residual U-Nets will give good accuracy .Writer also discuss about RECURRENT U-NET, -NET++,,ADVERSARIAL U-NET, ENSEMBLE U-NET, . he also add About MRI,CT,RETINAL FUNDS IMAGING, DERMOSCOPY,ULTRASOUND, X-RAY,MICROSCOPY,DERMOSCOPY and other image mode.then they brief us about U-Net network statistics and its limitations. In this paper the writer wanted to show the way for those people who are willing to discover U-Net.

[16] suggested DeepCut which is a technique for deriving pixel wise object segmentations from a dataset of image with poor labeling, in this context bounding boxes. This incorporates machine learning into the famous GrabCut technique by building a neural network classifier from the annotation of the bounding box . To get pixel wise object segmentations, authors structure the task as a power reduction problem over a fully linked unsupervised field and repeatedly modify the goal of training. In addition, they suggest DeepCut variations and compare them to a simple approach to convolutional neural network training with little supervision. They put it to the test on a difficult resonance dataset of fetal magnetism to see if it can tackle the segmentation of brain and lung challenges, and the findings are promising in terms of effectiveness.Consumer given bounding boxes are a basic common annotation type which has been widely utilized to initialize object segmentation systems in the computer vision. Bounding boxes provide a benefit over other types of annotations in that they allow the problem to be properly confined (i.e., the item should be

unique to the bounding box region completely enclosed inside it). Bounding boxes may be constructed in practice using two corner positions, allowing for quick placing (about fifteen times quicker than pixel wise segmentations and minimal storage of data. DeepCut is part of a group of recurrent optimization algorithms that includes a famous GrabCut method. Both techniques have two main stages such as label updating and model estimation.Researchers purposefully picked a database with a lot of diversity in the imaged structure to see if a basic network design is enough for the segmentation of bounding box object challenges. They prevent learning attributes for the full picture space by limited learning background regions from the halo H, allowing for quicker training. They found that the results of segmentation of the brain are more accurate than lung segmentation results for all internally compared approaches. There are various reasons that contribute to these discrepancies, and they affect all of the compared approaches in the same way. The brain's regular form can be better represented with a bounding box than the lungs. This reduces the number of initial false-positive targets, making the CNN training easier.Second, as the brain is often surrounded by hypo-intense muscular tissue, the context contrast is higher in the brain. DeepCut can segment both the lung and brain from an image data with a wide range of anatomy and is easily consistent with similar problems on medical data. In terms of accuracy, the proposed technique outperforms the competition to a model being trained under strict observation and at the same time considerably decreases the amount of annotation time needed for analysis.

In [17], the researchers used a network for doing semantic segmentation, for the brain tumor segmentation of 3D MRIs. They used an NVIDIA Tesla V100 32GB GPU for the model. They used the BraTS 2018 dataset, which has 285 cases. They cropped the size to a 160x192x128 size to make sure that it contained all the necessary contents. They divided the data into 2 parts and validated with 66 cases and the testing with 191 cases. They used CNN architecture with an encoder-decoder-based approach, which had an asymmetrical encoder for extracting the features of the image. As the dataset size was limited, the auto-encoder branch was used for more guidance and regulations. The encoder part used ResNet blocks with an additive identity skip connection. As the batch size was small, group normalization was used. All the input images were normalized so that it can achieve zero mean with a std unit. A random intensity shift and scale were applied per channel for the images too. A random axis mirror flip was also used. Using the common CNN approach of progressive downsampling and also increasing the feature size by 2 at the same time. The encoder endpoint was 256x20x24x16, which is much smaller than the input image. The decoder structure was built similarly to the encoder, with minor exceptions. The decoder levels began by reducing the feature by 2's factor and making the spatial dimension double. The decoder ended up with an image of the same size as the pre-processed image. They used the Adam optimizer to progressively decrease it. L2 norm regularization with a weight of 1e 5 was used on the kernel parameters. Also, a random mirror flip on all 3 axes was applied with a probability of 0.5. To draw input images in random orders, they used a batch size of 1. Using the high gpu size of 32 GB, the number of features were doubled in the process. The VAE branch which regulated shared encoding improved the performance of the method and helped it have good accuracy for any random initializations.

Firstly, in [18] the writer describes brain tumors and how they grow in the human body . The objective of this writer is to create a 3D (CNN) for brain tumor division from 3D MRIs and give an instability measure to evaluate the certainty of the model's future predictions. brain cells are tiny in size so they show in graphs of ET,NCR and ED. In image pre-processing writers apply Random Flip of fifty percent and they use 90 degree angle on 2 axes. Author uses patch for sampling strategy . The author uses a dual way for patch size. First one is Binary related and another is random tumor related . After that they used the Dice score coefficient and modified that they took the Number of N number of voxels . The writer uses 3 networks in this paper and they use ADAM optimizer. If writer's model validation loss is not satisfied within 30 epochs then they get the help of GDL loss.author discusses the V-Net for 4 output paths. For training writers use a 96 x 96 x 96 size patch and shuffle the tumor strategy which is called random tumor distribution strategy. Writer also explains that the U-NET symmetric Encoder has been used here for network architecture. Author changed individual normalization into group normalization.writer use an extra step so that even smallest part can be detect .writer is sharing that they wanted to run the prediction model more than 100 times and in final output their average and they will be put it in model for prediction.The last expectation and vulnerability maps are processed after similar systems as in the epistemic vulnerability.. Writer claiming they work with data set in pytorch. Finally When an ensemble of the suggested models is made , the simplest leads to the validation set are achieved. Writers can make use of every model's biases, but still an extended way from this state . Results could be a consequence of poor training approach during which the sampling technique fails to represent the proper label distribution, leading to a better number of erroneous detections. And the author states that there are more ways are open to develop this model .

[3] addresses the Compact nature, Efficient nature, and about Representation of 3D Convolutional Networks and planned a highly minimized network design for the division of fine constructions in volumetric pictures. Thus the basic and adaptable components of modern convolutional networks needed to be contemplated, such as enlarged convolution and remaining association. The majority of the current network designs followed a completely convolutional downsample-upsample pathway. So, a 3D engineering that consolidates high spatial goal highlight maps all through the layers, that could be prepared with a wide scope of responsive fields was suggested in this approach. In this paper, the network got authenticated with the difficult errand of computerized brain division into one fifty five constructions from T1-weighted MR pictures, and the proposed network, with numerous occasions less boundaries, accomplished cutthroat division execution contrasted and best in class structures. The notable AlexNet and VGG net were prepared on the ImageNet dataset. The prepared network was the initial move towards a broadly useful volumetric image representation and it conceivably gave an underlying model to move learning other volumetric image division undertakings. In this work, the strategy required around sixty seconds to foresee an average volume with 192×256×256 voxels. To reach better division results and scale the weakness, ten Monte Carlo trials of the misfit model were required. The whole interaction required somewhat over ten minutes, altogether. Nonetheless, the run time of Monte Carlo examination at trial time, just the misfit layer and the last expectation layer were randomized in the

mentioned paper. The attainability of voxel-level vulnerability assessment utilizing Monte Carlo trials of the proposed network with misfit on trial time was exhibited. Contrasted with the current volumetric division organizations, the compact network has less boundary communications, and in this manner possibly accomplished better vulnerability gauges with less samples.

In [19] paper the authors are using Lesion Priority to Optimize three-dimensional U-Net in case of tumor segmentation of the brain. They are using a lesion prior and a three-dimensional U-Net to optimize the segmentation of a brain tumor using a viable method. To begin, they create heatmaps of different sorts of lesions using ground-truth of the brain tumor lesions from a sample of medical patients. The volume-of-interest (VOI) map, which incorporates previous data on brain tumor lesions, is created using these heatmaps. In this paper they mentioned that there are two phases in the described lesion before fusion method: First, From the ground-truth of the tumor of the brain lesions, they construct a volume-of-interest (VOI) map, which is then combined with the several different modes of MR images and fed into a three-dimensional U-Net for the brain tumor segmentation. The incorporation of the lesion prior toward a three-dimensional U-Net design, which increases the segmentation of tumor performance of the three-dimensional U-Net, is the paper's key contribution.They used a Multimodal Brain Tumor Image Segmentation Benchmark (BraTS) 2017 [2–4, 14] dataset which includes 285 training subjects and 46 validation subjects.For the purpose of data pre-processing. Normalizing intensity is a way of transferring data from various MR images onto a common scale, and this is a key step in eliminating early biases and enhancing the network's performance. This paper's suggested network architecture is built on three-dimensional u-nets with a three-dimensional convolutional layer.Spontaneously sliced regions of size 128 128 128 blocks and packet size Two are used to train the suggested structure. Expanded input region collects more from the brain's relevant data. A clipped region is arbitrarily taken from each participant at the end of each period. The model was trained for 300 iterations in total. To get the anticipated lesion mask, they fed the complete picture of size 240 240 155 blocks inside the trained three dimensional U-Net for every patient during testing. Both training and testing do not need data augmentation.Their suggested lesion before fusion technique increases the performance of three dimensional U-Net, especially for ET's DSC which is 3.5 percent, for H95 which is 2.56, and entire tumor is 2.39. Additionally, their developed lesion before fusion strategy increases the segmentation of tumor effectiveness of the assemblage of five three dimensional U-Nets, especially for the DSC of ET which is 2.1 percent and tumor core segmentation which is 1.7 percent . The metrics of the ET heatmap have had the most importance when we build the VOI map, which is why the VOI map has the biggest boost on the ET.Furthermore, the suggested lesion prior fusion approach may be simply combined with various network designs to improve segmentation efficiency even further.

[20] introduce us to a Computed Tomography image method .Firstly they talk about two image methods. First one is Carvalho et al. and second one is that discussion is based on 2D convolutional neural networks and 3D-based methods.their network can learn more feature information than two-dimensional networks by fully using three-dimensional spatial data. They restrict the size while losing the effectiveness

by using 3D-UNet. They face some issues when they bring more layers when trying to bring better results.They solve that problem by using residual network (Resnet). Author added more channels in Res2Net to make it 3D-Res2Net. It is the quantity of component features with scale = 4 so the element map shipped off this design is changed over to eight channels after the $1 \times 1 \times 1$ convolution.They are announcing a training method that if we group the channels and then train them we can bring out better result than non group train data set. They alter the weight of every channel after every convolution. They also consider the speed of the network . In the dataset they divided the dataset into three sections .Seventy percent of the whole data was utilized for the training part, twenty percent for the testing part, and ten percent for verification. author's data set including two files with the suffix zraw and mhd. The U-net guarantees the size of the element map while adequately fixing the vanished important information . The strategy of their paper was tried on the LUNA16 public dataset,The analysis initially looked at the capacity of 3D-Res2UNet and the first organization to section and fit lung knobs. where the dice coefficient record arrived at 95.30% and the review rate came to 99.1%. Few pictures whose segmentation isn't exceptionally precise. This is on the grounds that the state of such lung nodules is not the same as most lung nodules.The suggested network performed well in the segmentation of lung nodules and made significant progress in the segmentation of tiny nodules.

[21] proposed a segmentation framework which is based on deep learning that incorporates CNN. They are fine-tuning either supervised or unsupervised images specifically to make an adaptive CNN model for them. The training was done on one node of Emerald cluster3 with two 8-core E5-2623v3 Intel Haswell's and a 128GB memory K80 NVIDIA GPU. The dataset used for the training is BRATS2015 which included 274 scans from many different patients. For different organs and modalities, a bounding box was used which takes standard deviation of the region and mean value as input values. Fine-tuning-based Segmentation is used. Their proposed CNN takes the content of the bounding box as inputs and outputs for binary segmentation. For the testing phase, a bounding box and their segmentation technique extract the specified region in the bounding box and feed it to the CNN to obtain an initial segmentation. For the 2D images, they adopt P-NET to get the bounding-box-based segmentation. The proposed network contains six blocks with a 181×181 receptive field. The first five boxes contain dialative parameters which capture features at different scales. These features are then fed into a block6 classifier. For the 3D images, the network is extended. It has an $85 \times 85 \times 9$ anisotropic receptive field. In the testing stage, the trained model was implemented on a MacBook with 16 GB RAM and an Intel Core i7 CPU running at 2.5GHz. For user interactions on images, they used Matlab GUI and a PyQt GUI. For the evaluation, The segmentation results by an Obstetrician and a Radiologist were used. For quantitative evaluations, they used the ground truth and the Dice score between the segmentations. Through their framework can be applied to different CNNs, this research mainly focuses on the interactive segmentation where the memory efficiency of the network and short inference. This also helps to work on low GPU machines. The end results show that the segmentation performance is improved by image-specific fine-tuning. Though the fine-tuning helps correct some mis-segmentations, the performance of the model is satisfactory. It leads the model to under-perform when

dealing with various complex cases.

Classical deep CNNs for completely autonomous brain tumor segmentation include limitations such as spatial data lost due by both repetitive pooling and inadequate capacity of multi-scale lesion processing, according to authors of [22] . Authors employ a three dimensional atrous-convolution with a singular stride to substitute pooling establish the backbone for feature learning to solve their first difficulty. A three dimensional atrous-convolution feature triangle is built appended to the end of the backbone for the second challenge. This structure increases the overall model's discriminating capacity to separate tumors of varied sizes by adding contextual data. Authors employ every hierarchical feature map created by the backbone system to accurately forecast tumors, taking into consideration the multi-scale fundamental feature of hierarchy in DCNNs. They construct a feature triangle linked to the backbone system for combining multi-scale context characteristics with lesions in order to deal with malignancies of various sizes. They claim that their system can sparsely relay complicated data flows from several hierarchies without sacrificing any feature data. The studies show that the suggested technique just not fixes the issue of data loss that is caused by typical DCNNs' pooling operations, moreover accurately segments malignant lesions of various sizes.In this article, they choose a size of the data that is greater than the receptive areas of DCNNs, causing the final softmax layer to yield many predictions at the same time . All forecasts are equally reliable as DCNNs' receptive area can encompass entire data without padding. This can prevent repeating convolving in the overlapping patches of the very same voxels , lowering the computational cost and storage burden significantly.Authors evaluate the segmentation performance obtained by the proposed model to the state-of-the-art methodologies on the BRATS 2013, 2015, and 2018 benchmarks to legitimize the efficacy and reliability of their method. They can see that the proposed approach has a superior impact in segmenting tumor's components, for example the enhancing core and tumor core . Their observations on various data demonstrates that the suggested system is capable of distinguishing lesion features from several other tissues, particularly when equipped with the three dimensional atrous-convolution feature triangle. The key reason is that the network's planned structure, which is formed using three dimensional atrous-convolution and the atrous rate for every hierarchy, allows it to hierarchically collect characteristics without losing any data during data flow propagation.

In [23], the authors worked on a 3D U-net-based architecture with encoder-decoder. They tried to segment the tumor. They tried measuring the overlap region between the two regions. For that purpose, the function of soft dice loss was used. Then the focal loss was used to balance the samples from negative and positive, which is done by tuning the weighted parameters. NVIDIA Quadro K5200 and Quadro P5000 GPUs were used for this research. They used the Brats 2019 dataset for their research, which contains 293 HGG and 76 LGG pre-operative scans. The network inputs are done in patches of 64x64x64 from four modalities of the dataset. For the preprocessing of the data by de-noising and Z-score normalizing the individual MR sequences, Then the data is augmented by flipping around the vertical axis of the patches. Images in patches of size 64x64x64 are then trained on the focal loss and dice loss functions. Then the dataset is run on 610 epochs of this dataset with a batch size of 1. The post-processing is done by analysing the connected component to remove the tumor which has less than a thousand voxels and enhance the tumor in the surrounding necrosis. By applying this technique in the post-processing, the achieved false-positive voxels of the segmentation are removed. The proposed network fails while segmenting the tumor from some LGG and HGG images. The failure is observed by the small size of the tumor, necrosis and the absence of enhancing tumor. When the proposed network fails to segment necrosis , all the features are marked by the feature extractor except age as zero considering the absence of necrosis. The paper also informs us about the fact that the gender of the patient from whom the image is coming can improve the accuracy of the system in the cases of post-operative treatment. Though the network outperforms some of the ensemble approaches, it fails if the tumor size is too small or too big.

In [1], the volumetric CNN executes segmentation on some prostate volumes of MRI. They augmented the image dataset through a randomized variation in different training iterations. Using V-net on MRI volumes, the prostate was depicted. They evaluated the performance of the applied process in regards to the Dice coefficient, Hausdorff distance of the predicted description to the original annotation. By using the Caffe framework in python, training and testing were run. The method was trained on fifty MRI volumes and their related various true annotations. The results were the optimization of the training. The model's dice loss layer was highly improved during the imbalance of the background and the foreground pixels. Then improvement of both the results and convergence time were done.

Segmentation for medical images that are very small are quite difficult, because they have a low contrast and are anatomically variable in size and shape. Focusing on the small organs and accuracy of their segmentation the recurrent saliency transformation network is described in [24]. It is based on a previous paper on the same coarse-to-fine strategy containing 2 stages that are instructed exclusively, but with some advanced strategies consisting of multi-staged visual cues in optimization. Here, 2 Dimensional networks are instructed for 3 Dimensional segmentation by slicing the 3 Dimensional volume along the 3 axes and each axis is sent to train in a particular 2D-Fully Convolutional Network on a 16-layered VGGNet. Two FCNs are trained, firstly for the segmentation of coarse-scaled and secondly for the segmentation of fine-scaled and these two are optimized jointly as a consequence of the key innovation of the proposed approach, the module for saliency transformation that often transforms the segmentation's probability-map from the previously done iterations as spatial weights and applies those weights to the recent most iteration. The DSC loss term is computed on each probability map for minimization of the overall loss. The training phase mainly aims to minimize the loss function. The entire NIH pancreas segment dataset is filled with eighty two enhanced contrast abdomen CT volumes and are split and trained using standard cross validation strategy and the DSC value for everyone of the sample is measured along the average and standard deviation over all of those eighty two cases. It gives an average accuracy of 84.50% which is greater than the previously introduced fixed-point model where two stages were calculated individually without joint optimization. Again, a case-by-case study shows it reports a higher accuracy number of 67 out of 82 cases with the largest advantage of +17.60% and largest inadequacy of a simple -3.85%. The model is reoccurring, allowing fine and scaled segmentation that's updated iteratively, as well as cropping the input images for focusing on the prominent regions, resulting in improved segmentation accuracy. The output achieved shows 78.23% accuracy in the coarse stage, 82.73% just after the first sequence and after completion of the entire testing phase the accuracy report scales towards 84.80%. This approach, when applied to 2 datasets at once for the pancreas the segmentation method and the multi organ segmentation method, crucially outperformed in the- state-of-the-art.

Tiled CNNs used tiling schemes to concurrently appreciate the advantage of fundamentally lessening the quantity of absorbable boundaries while furnishing with the calculation adaptability to learn different uniforms.In this paper, unaided pre-training was introduced to be specifically utilizing an adjustment of Topographic ICA (TICA) to master arranging highlights in a geographical guide by pooling groups of related features together and [25] showed that TICA could be used productively to pretrain Tiled CNNs using nearby symmetry. The subsequent Tiled CNNs referred in the paper with TICA were undoubtedly ready to learn uniform representations, with pooling units that were strong to both scaling and turn. Tracking these down improved order execution, empowered Tiled CNNs to be serious with recently distributed outcomes on the NORB and CIFAR-10 datasets. In the mentioned paper, CNNs also had been effectively applied to numerous acknowledgment tasks. Such as digit acknowledgment from MNIST dataset, object acknowledgment from NORB dataset and common language preparing. The outcomes showed that the loosening loads are valuable for order execution. As a result of variable outcomes for a simple change in exceptional instance, k a completely tied model, a naive approach to deal with getting familiar with the channel loads was to straightforwardly train the main layer channels utilizing little fixes (e.g., 8x8) arbitrarily tested from the dataset mentioned in the paper, with a technique like ICA. In this paper, The naıve approach brought about fundamentally diminished arrangement precision acquiring 51.54% on the demonstrated set, while pre-training with TICA accomplished 58.66%. These outcomes affirmed that improving for paltriness of the pooling units brought preferred highlights over na¨ıvely approximating the main layer loads. A characteristic decision of the tile size k is to set it to the size of the merging locale p, which for this situation was three. Here, each pooling unit consistently joined straightforward units which were not tied. The Tiled CNN just required unlabeled information for preparation, which could be acquired efficiently and fundamental outcomes on networks pre-trained utilizing 2,50,000 unlabeled images from the small image dataset showed that presentation increments as k goes from one to three, leveling out at k = four. This prompted that when there was adequate information to keep away from overfitting, mounting k = p can be a generally excellent decision. Moreover, in the paper it was presented that Tiled CNNs as an expansion of CNNs that help both unaided pretraining and weight tiling.
[26] states the primary Convolutional Neural Network qualifies for real-time SR of 1080p tapes on one K2 GPU. Here, a CNN architecture is introduced, which extracts the characteristic diagrams within the LR space and a profitable sub-pixel convolution coating that is competent of learning a collection of upscaling filters to upscale the ultimate LR form, mapping to the HR outcome. This approach suggests increasing the result from LR to HR only at the very end of the structure and super-resolve HR data from LR characteristic maps which removes the necessity

to execute maximum of the SR procedure within the far enormous HR resolution. For this goal, a further profitable sub-pixel convolution layer is applied to find out the upscaling strategy for picture and video superresolution. During this structure, upscaling is dealt with by the previous layer of the hierarchy which suggests each LR image is immediately employed to the structure and extraction arises through the nonlinear convolutions in LR space. This hierarchy is prepared for discovering reasonable LR to HR mapping distinguished to single cured filter upscaling at the very early layer which ends in more increases within the reconstruction exactness of the classification. The primary dataset used for this method is the Timofte dataset which is widely employed in SISR, it contains 91 training pictures and two trial datasets named Set5 Set14 with 5 14 images respectively. Then, Berkeley segmentation dataset BSD300 and BSD500 with 100 and 200 pictures for testing are additionally used including the super textured dataset that comes with 136 texture pictures. Lastly, for training ultimate prototypes, 50000 random images are selected from ImageNet. For every upscaling factor, a particular network is trained. For video operations, 1080p HD videos are used from the Xiph database1, which contains 8HD videos together with the Ultra Video Group database that contains seven videos. Within the training phase, sub-images are taken out from the training basis reality images I HR. The training takes approximately 3 hours on a K2 GPU on 91 images, and seven days on images from ImageNet for an upscaling component of three. PSNR is employed as the execution metric to gauge the models. This method accomplished a surprising regular speed of 4.7ms for super-resolving 1 single image from Set14 on a K2 GPU and achieved state-of-art performance because the results are better compared to the simplest SRCNN 9-5-5 ImageNet model. It's approximately a sequence of magnitude quicker than earlier announced methods on images videos at the time of publishing the paper.

For attaining pixel-level accuracy in semantic segmentation, [27] presents the design of a dense upsampling convolution (DUC) which is conversant with capturing and decoding very particularized information as well as a hybrid dilated convolution framework or a HDC framework for the encoding phase. The proposed model is named ResNet-DUC-HDC and it is trained and tested using three datasets; Cityscapes dataset, KITTI dataset and the PASCAL VOC2012 dataset. For starting the ResNet-101 or ResNet-152 networks are primed on the ImageNet dataset. Here, MXNet is used to train the models as well as to evaluate them on NVIDIA TITAN X GPUs. Cityscapes is a large dataset containing about 5000 images annotated with 30 categories. DeepLab-V2 ResNet-101 framework is utilized in terms of training the model's base line. The network is then primely trained, for 20 number of epochs and achieved a mIoU of 72.3% on the experimental validation set. The efficacy from DUC on the baseline network is examined by only changing the shape of the topmost convolution layer. The combined ResNet-DUC network is trained for twenty number of epochs and achieved a mIoU of 74.3% that is, a 2% increasing validation compared to the baseline network. The ResNet-DUC-101 model is then experimented with several HDC modules and then replacing a deeper model, ResNet-152 with the ResNet-101, a further better performance was acquired. The output here is, the ResNet-DUC-HDC model achieved 77.6% mIoU using the fine datas only. Adding the coarse data helped achieve 78.5% mIoU, which after updating the network after retraining reached to 80.1% mIoU. The KITTI road segmentation activity having 289 number of training images and 290 number of testing images was also used for experimenting on the ResNet-101-DUC model without any application of CRFs for post processing. The result was highly impressive, as the average precision was 93.88% and highest maximum F1-measure was 96.41%. Again, PASCAL VOC2012 segmentation benchmark dataset was used for experimentation with the ResNet-DUC-152 model. That dataset contains about 1464 numbers of images for training, 1449 numbers for validation and 1456 numbers for testing. Then after completion of the experiment, the model ended up achieving mIoU of 83.1% on the test set just using a simplified sole model without any type of model ensembles or multiscale testing application, which was considered the best and effective method at the time being.

An effective strategy for volumetric clinical picture segmenting is introduced in [12], which is a notable testing issue in clinical picture analysis. In their paper, they described a 3 Dimensional profoundly regulated system for computerized segmentation of volumetric clinical pictures. One principal difficulty of saddling CNNs for clinical picture analysis tasks was, contrasted and common image applications, medical applications as a rule had restricted preparing information. Albeit the clinical picture datasets mentioned in the paper are utilized in the two solicitations in the approach were not extremely huge in subject-level, they accomplished cutthroat execution to the best in class strategies. In the mentioned paper they offered 3 Dimensional DSN with two testing division tasks; here was broadly approved by the authors, i.e., liver segmentations from midsection 3 Dimensional CT sweeps and heart segmentations from 3 Dimensional MR pictures, by partaking two notable difficulties held related to MICCAI. To approve the proposed strategy on the utilization of liver segmentation, in the paper the SLiver07 (Heimann et al., 2009b) dataset was utilized, which was from the segmentations of the Liver Competition held related to MICCAI' 07, and the terrific test stayed open subsequently. The dataset completely comprised thirty CT filters with twenty preparings and ten testings. Furthermore, To approve the proposed technique on using the heart segmentation, the dataset of MICCAI' 16 Challenge was utilized on Whole-Heart and Great Vessel division from 3 Dimensioned Cardiovascular MRI in Congenital Heart Diseases, also for short, the HVSMR challenge what's mentioned in the paper. The dataset generally consisted of twenty axial, cropped pictures with ten training and ten testings. Next, envisioning the middle-of-the-road consequences of the neural networks that are trained on the liver segmentation dataset, to approve the viability of the profound oversight on rapid layers in the training interaction. The test results depicted, the CRF was accommodating to improve the liver segmentation results. Notwithstanding, the heart picture segmentation didn't acquire incredible upgrades from this post-handling. It was hard to sort out a universal arrangement of frameworks that make enhancements for both myocardium and blood pool. At long last, based on the top-notch score volumes acquired from 3 Dimensional DSN, in the paper they utilized a CRF model to refine the segmentation results and broadly approved the proposed 3 Dimensional DSN on two unmistakable solicitations also, the outcomes exhibit the adequacy and speculation of the offered network.

A dual pathway, 11-layers deep, three-dimensional Convolutional Neural Network for the challenging task of brain lesion segmentation is introduced in [28]. Firstly, an architecture was established in the field of in-depth experiments with the boundary of a network related to it. According to the belief that expert intelligence and experience are very much needed to get accurate results in segmentation, the proposed method has included two main points. First one was to get light version segmentation maps with an FC (fully connected) 3 Dimensional CRF and the other one was final hard segmentation labels. A baseline was set and the dense training was done on image segmentations as well as class balance. It was claimed that the method in that stage was a merged idea between mostly used training on individual patches and the heavy training idea on a full image. The proposed method was compared with two other commonly used methods, of which, one was a common scheme that trains on $17\hat{3}$ patches that's extracted uniformly from the brain regions and the other one was a scheme that samples the patches equally from the lesion as well as the background class. The model was created by doubling the nine layers of 'deep+' and merging two concealed layers. Also introduced "BigDeep+", which has eleven layers and "DeepMedic", which has the same numbers of the parameters. "DeepMedic", changed over to 2 Dimensional by reshuffle 3rd measurement of every single bit to one and a productive and successful thick preparation conspiracy was formulated which joined the preparation of adjoining picture patches into one pass through the organize whereas consequently adjusting to the inalienable course awkwardness displaying within the information. In order to consolidate both neighborhood and bigger relevant data, a double pathway design was utilized that formed the input pictures at numerous scales at the same time. For the next stage of the process of the network's light version segmentation which is also known by soft segmentation, a fully 3 Dimensional connected CRF environment was used which viably expelled wrong positives. The Our pipeline is broadly assessed on three issue errands of injury division in the multi channel MRIs understanding information with the brain tumors, traumatic brain wounds, and ischemic stroke. The approach made strides on the-state-of-the-art for all of the three applications, with the beat positioning execution on the open benchmarks, BRATS 2015 and ISLES 2015.

Chapter 3

Methodology

3.1 Model Architecture

3.1.1 CNN

A convolutional neural network (CNN) is a pattern of artificial neural network that collects information employing perceptions, which is a machine learning unit method. Image processing, NLP, and other specific components can all be applied to CNNs [28]. CNNs are multilayered perceptions that have been standardized and made up of many layers of artificial neurons. Multi - layered perceptions often refer to networks that are completely linked, meaning that each neuron inside one layer is linked to all neurons in the next layer.The Convolutional Neural Network (CNN) [29] is used for image recognition and detection. CNNs are incredibly efficient, and picture detection requires just a little amount of computer resources. CNN's working procedure is quite simple, and it uses layers of neurons to perform its detection.

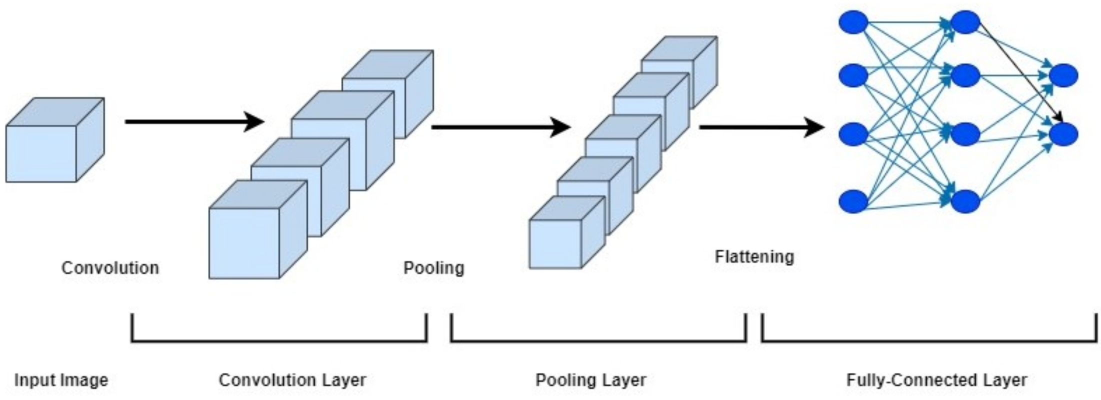


Figure 3.1: Convolutional Neural Network (CNN)

3.1.2 FCN

Fully convolutional networks are regarded as some higher class models addressing many pixel wise tasks. [11] FCNs are reviewed as the networks that do not contain dense layers like CNNs, rather they contain 1 x 1 convolutions and work like fully connected layers. In recent days,[30] The semantic segmentation highly improved the accuracy by the transfer of pre-trained classifier-weights, fusing different layers of representations, as well as learning the end-to-end on entire images [16]. FCNs generally have a primary architecture like below, though the shape varies for different models.

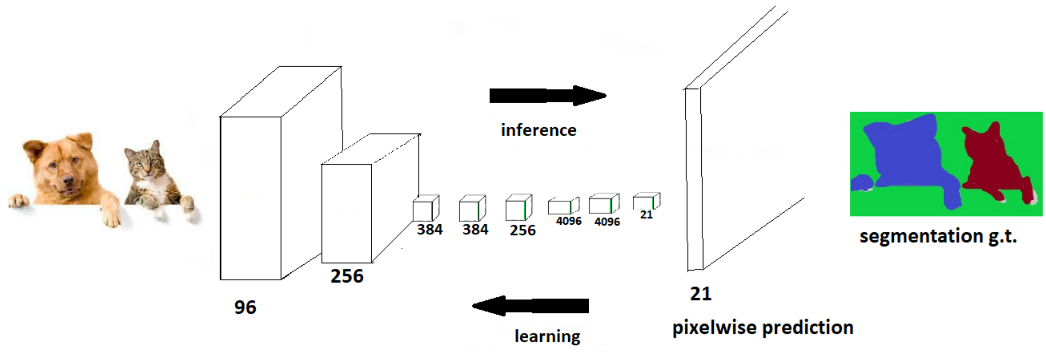


Figure 3.2: FCN Model Architecture

Fully convolutional networks are efficient and can learn to make denser predictions for per-pixel approaches. The FCNs used in this network are-
1. 3D U-net
2. 3D V-net

3.1.3 3D U-Net

Depending on the work type and use models are created and used. U-net is designed for semantic segmentation. It is simply called U-net because of its architecture. It has excelled in a number of tasks, and it is still one of the most widely used end-to-end structures in the domain of semantic segmentation [18].

O. Ronneberger, P. Fischer, and T. Brox initially proposed 3D-UNet in their article [13] [31]. We host a 3D-UNet version modified for the brain tumor segmentations done by Fabian I. et al. 3D-UNet enables seamless segmentation of 3D volumes with excellent accuracy and performance, and it may be modified to tackle a wide range of segmentation challenges.The first portion of the U shape is a contraction or a down-sampling operation the path is called the encoder path and the other half is the expansion or the up-sampling operation while the path is called the decoder path.The operations or path is concatenated which is a reason we get localized information making it possible for us to get a semantic segmentation. Here, The input image dimension we provide is 128 x 128 x 128. We perform a standard combination of the 3D convolution operation, followed by 3D max pooling operations.

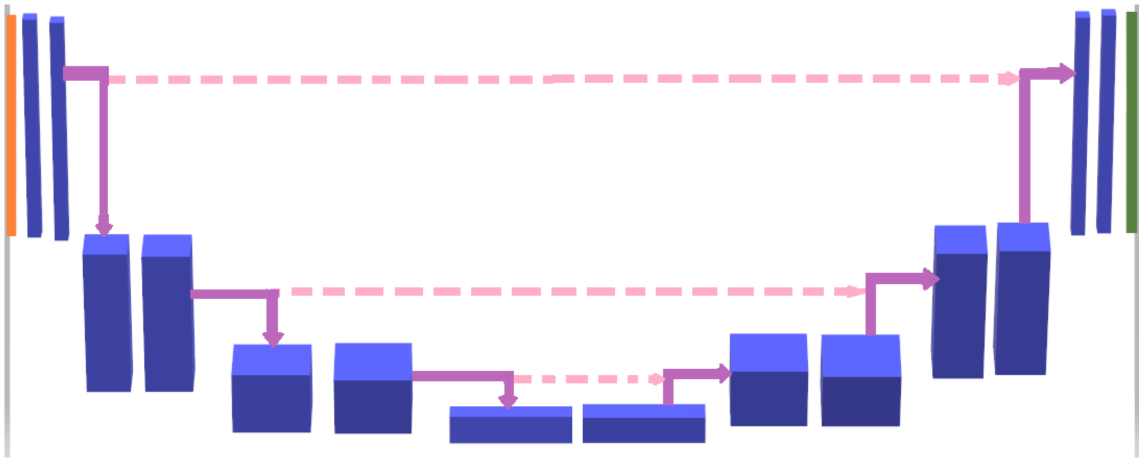


Figure 3.3: 3D U-Net Model Architecture

In the architecture, each layer contains two 3×3×3, 3D convolutions each followed by a ReLU, and afterward a 2×2×2, 3D max pooling with steps of two in each aspect. In the blend way, each layer comprises an up-convolution of 2×2×2 by steps of two in each aspect, trailed by two 3×3×3 convolutions each followed by a ReLU. Easy route associations from layers of equivalent goal in the investigation way give the fundamental high-goal elements to the amalgamation way. In the last layer, a 1×1×1 convolution diminishes the quantity of result channels to the quantity of marks. The group standardization is done before each ReLU.

3.1.4 3D V-Net

F. Milletari, N. Navab, and S. Ahmadi originally introduced V-Net in [1] . V-Net provides for seamless 3D image segmentation with great accuracy and performance, and it may be modified to tackle a wide range of segmentation challenges. V-Net is made up of a contractive and expanding path that seeks to create a bottleneck in its center using a mix of convolution and down sampling. Following this bottleneck, the picture is rebuilt using convolutions and up sampling. Skip connections are inserted to aid the backward flow of gradients in order to improve training.

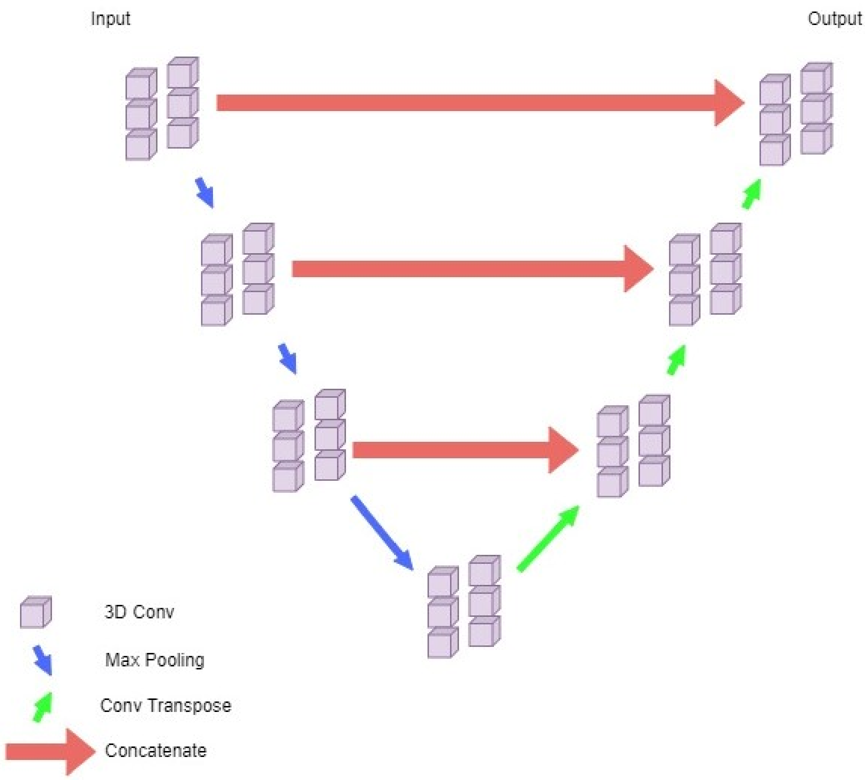


Figure 3.4: 3D V-Net Model Architecture

It is like U-Net, yet with certain distinctions. The left half of the organization is partitioned in various stages that work at various goals. Each stage includes one to three convolutional layers. The contribution of each stage is utilized in the convolutional layers and handled through the non-linearities and added to the result of the last convolutional layer of that stage to empower learning a remaining capacity. The convolutions acted in each stage while down sampling utilized volumetric pieces having size of 5×5×5 voxels. The size of the subsequent element maps is split, with comparative reason as pooling layers. Furthermore, the quantity of element channels pairs at each phase of the pressure way of the V-Net. Supplanting pooling activities with convolutional ones assists with having a more modest memory impression during preparing, because of the way that no switches planning the result of pooling layers back to their bits of feedback are required for back-spreading. Down sampling assists with expanding the open field. PReLU is utilized as non-linearity actuation work. (PReLU is proposed in PReLU-Net.) The organization separates, includes and grows the spatial help of the lower goal highlight maps to accumulate and gather the essential data to yield a two-channel volumetric division. At each stage, a deconvolution activity is utilized all together to increment the size of the information sources followed by one to three convolutional layers, including a large portion of the quantity of 5×5×5 bits utilized in the past layer.

3.2 System Architecture Of Our Proposed Model

We start the work through data collection, and we find the data of 3D MR Images. For our data pre-processing we focus on our data being lossless. Then the pre-processed data is then divided into two portions keeping the largest portion to train on the 3D Unet model for an efficient segmentation and the other for validation. The data to be trained is then trained in the model, data is then extracted and segmented for model testing with the help of the testing data. The result obtained is then analyzed and compared to other procedures of approaches.

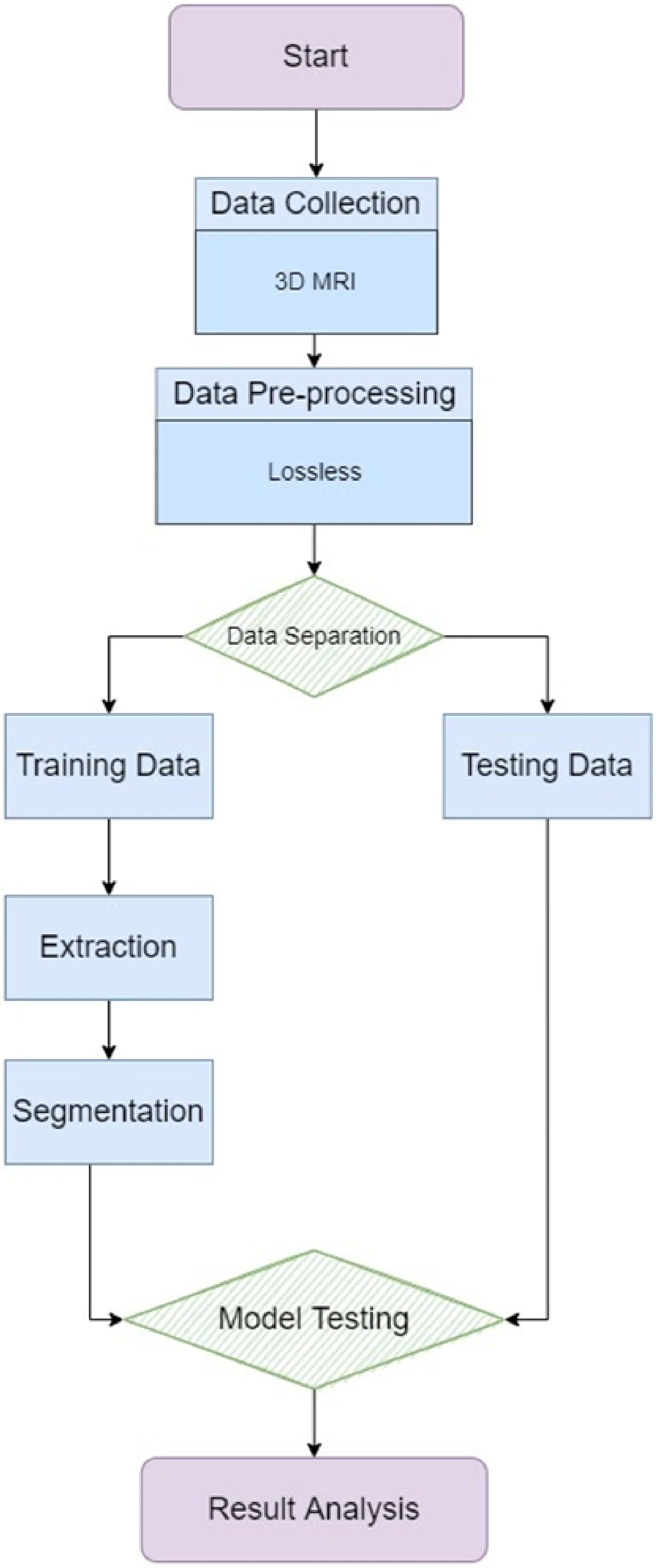


Figure 3.5: Workflow Diagram Of Our Proposed Model

To summarize, the methodology proceeds by these following steps:

1. Data Collection (3D MRI)
2. Data Pre-processing (Lossless)
3. Data Separation
   (a) Data Training
      i. Data Extraction
      ii. Data Segmentation
   (b) Data Testing
4. Model Testing
5. Result Analysis

3.3 Dataset and Preprocessing

3.3.1 Dataset Description

The dataset used is BraTS'20 MICCAI [32] [33] [34] [35]. The data provided went through a great journey before reaching BraTS'20. The previous datasets prepared by this source differed in each publication of data. The BraTS'17 was very different from the previous ones and so was the others. BraTS'17-'20 had similarities with the BraTS'12-'13 images and annotations as they were manually annotated by the clinical experts. The data used in the BraTS'14-'16 which were from the The Cancer Imaging Archive were discarded for a mixture of pre and post operative scan descriptions and also because their ground truth labels were annotated using fusion of segmentation results from algorithms that ranked high during BraTS'12 and '13. Expert neurologists are said to have radiologically assessed the complete original The Cancer Imaging Archive glioma collections as well as categorized scans of pre- or post- operative scans in a great precision. The Cancer Imaging Archive glioma collections had a TCGA-GBM of n=262 and TCGA-LGG with n=199. All the pre-operative The Cancer Imaging Archive scans with 135 GBM and 108 LGG were then annotated subsequently through a manual process for various glioma subregions and included in this year's BRaTS datasets, that is BRaTS'20.

The multimodal scans found here are saved as '.nii.gz' files, that is in NIFTI mode. In the dataset there are annotated brain tumor images. It contains 369 folders of images for training and 125 image folders for validation purposes well arranged. The training images describe 5 '.nii' files each, namely T1 which are explained as native, T1Gd explained as post contrast T1=weighted images, T2-weighted, T2-Flair referred as Fluid Attenuated Inversion Recovery volumes and seg while the Validation folders include the same except the 'seg.nii' files which a training model should come up with. It also includes information data sheets of 'name mapping' and 'survival info'. Here the TCGA-GBM and TCGA-LGG collections are made available through The Cancer Imaging Archive. The date-defined overall survival (OS) data is contained in a comma-separated values (.csv) file and the result matches the pseudo-identifier of the image data. The age of patients as well as their resection status were noted in the following file as well. It also provides information about the age of people whose MRI images were shown. All of this information was collected from different sources. The source information was collected through different clinical protocols and scanners, from about 19 institutions. Some images of the '.nii' files from the data are demonstrated below.

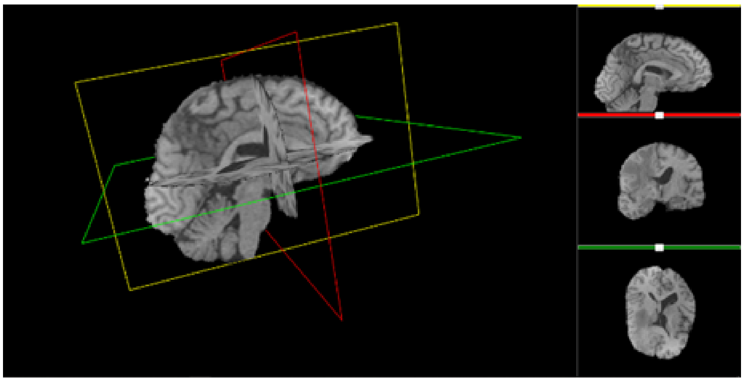


Figure 3.6: T1 Image from the BraTS'20 MICCAI dataset

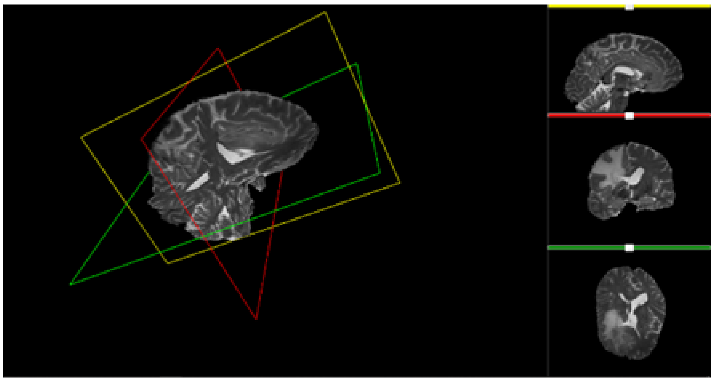


Figure 3.7: T2 Image from the BraTS'20 MICCAI dataset

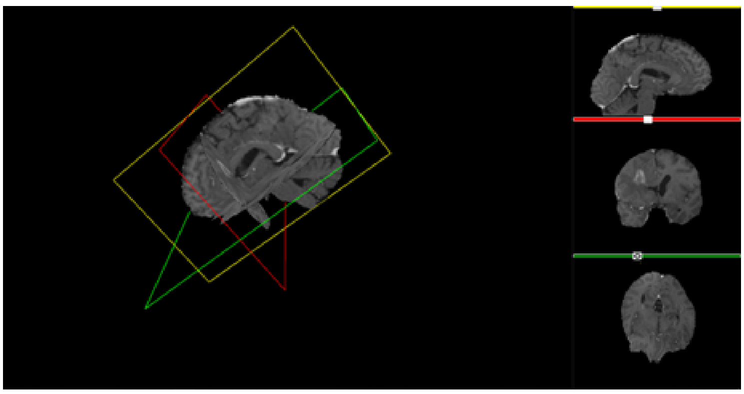


Figure 3.8: T1ce Image from the BraTS'20 MICCAI dataset

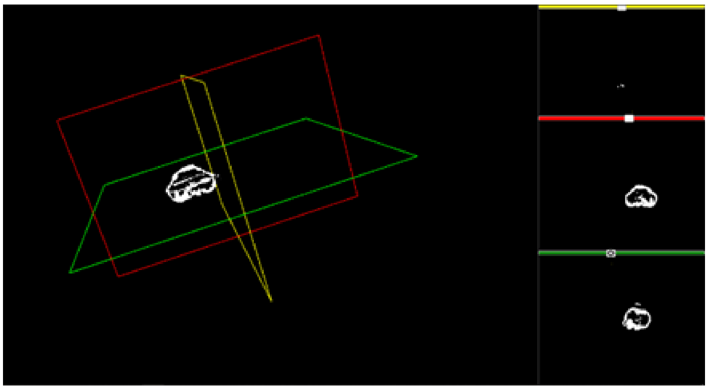


Figure 3.9: Segmented 3D Image from the BraTS'20 MICCAI dataset

The image dimensions are 240 × 240 × 155 for x× y × z. The voxel dimensions are 1× 1×1 mm. The images have an axial, a sagittal and a coronal view and here we include sample images from our dataset where every image has their corresponding tumor labeled with ground truth. Following the same annotation protocol, about one to four raters segmented the data manually for training purposes. The annotations were then approved by the neuro-radiologists. The annotations done shows the GD-enhancing tumor (ET — label 4), then the peritumoral edema (ED — label 2) as well as the necrotic and non-enhancing tumor core (NCR/NET — label 1) accordingly as described in the 2012-2013 TMI papers and latest summarizing papers from the same data source. All of them are co-roster for the same anatomical templates that are inflated in the same resolution after skull-stripping.

3.3.2 Data Preprocessing

The data we used was BraTS'20 MICCAI and we needed to pre-process our data for a lossless segmentation. In the dataset we got 4 channels of information, or could be said four different volumes of the same region. T1 which is a bit bright and can be easily detected, T1ce is the T1-weighted image which has a high contrast to make it clear, T2-weighted and T2 flair in which we can see the structures clearer. We took the T1, T1ce and the T2-weighted images. T1 images are greater in contrast so we get more information from that. Other than the images inside the slices we see a huge black region or area which is a waste of space and is not required. If that is removed then we get a better view to train the data as well as an optimized space. But since the black unlabeled volumes are different for different slices of the 3D image, we needed to deduct or crop the image in a way where there is no loss of image information after cropping the image data. So, we made 5 list for the five channels of information, passed and looped to get their cropped versions. Thus, there was no data lost and the cropped slices were resized to the shape (128 x 128 x 128) we needed for our model, then were iterated further through a condition where we separated and saved the slices which would be used and the ones that were not sufficient with annotated data, less than a certain percentage or could not be used were discarded. According to the dataset information, the pixels of value needed to be 4 and so was then reassigned to 3 because of it being missing from the original labels. We saved the data combining the T1, T2 and T1ce images into a single-multi channel volume from each folder into numpy arrays (npy) to use as our 3D input for the proposed model in use. Because using single volumes would be less beneficial, the combining procedure enhances the data more. Thus, our dataset was ready to train and validate our proposed model. Since, fixed region cropping was not the best lossless solution in this case so we chose to let our data be cropped by a bounding box algorithm where the black portions of the image were bound outside of a box letting us to clip it out. So that we do not lose even a single data from one of the many slices, we decided to pre- process the data further, to keep all the parts of the data except the label 0 or the blank region of the data. Thus, we plan to get more accurate segmentation validation through the post proposed preprocessing of the data.

Here are several cropped versions of the data, as well as their prior versions, using the bounding box cropping algorithm.

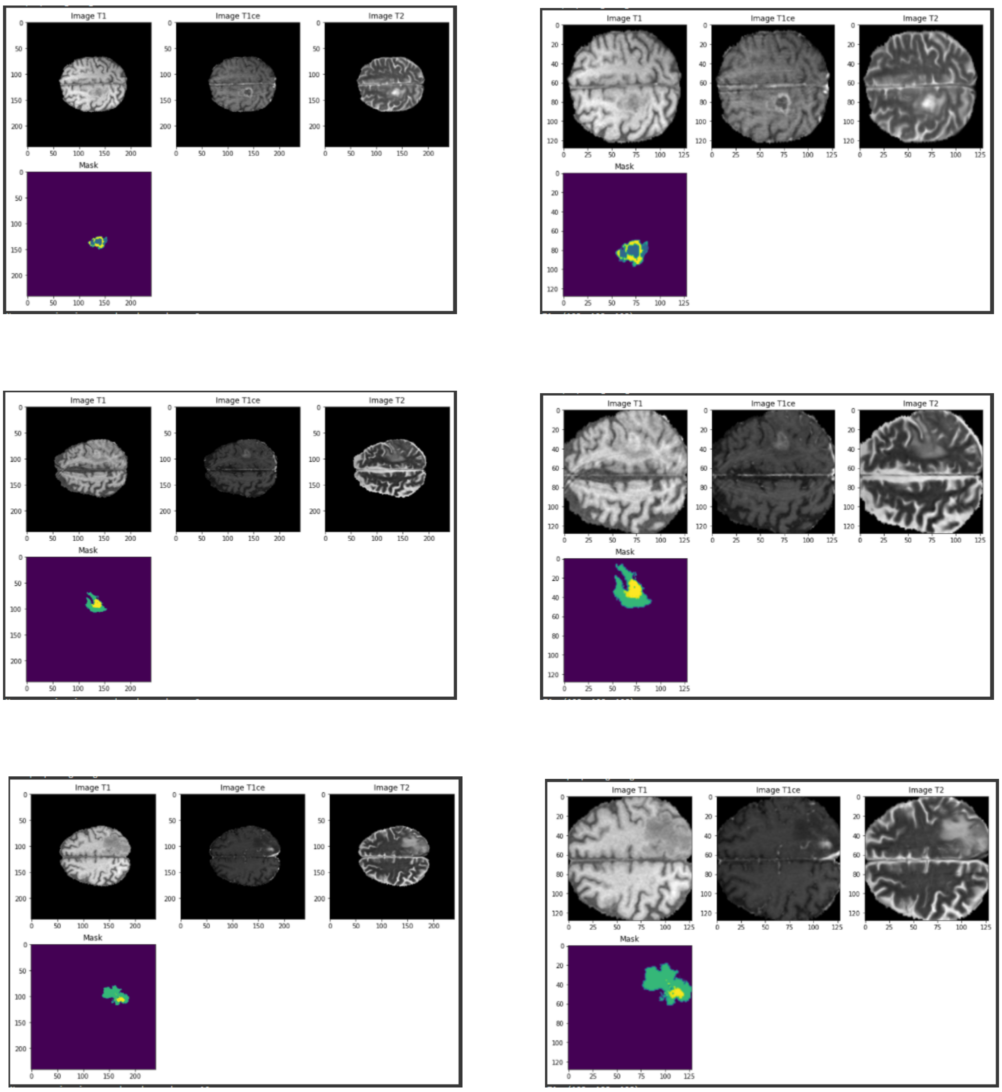


Figure 3.10: Cropped versions of Data using Bounding Box Cropping algorithm

3.4 Model Implementation

For our implementation we selected python and used the TensorFlow framework. The training was performed in a desktop with 16GB ram DDR5 10TH Gen core i7 8GB RTX3060 Ti. The dataset we used to train and validate our proposed model is BraTS'20 MICCAI with a dimension of $240 \times 240 \times 155$ for $x \times y \times z$. After preprocessing we got cropped data to feed our model as input. The cropped data generated was of 128 x 128 x 128 dimension.

The architecture we chose to implement was a simplified 3D Unet model for an enhanced semantic segmentation of our data. The model takes four parameters namely, height, width, depth and no. of channel and gives us the segmented output of the same dimension with its number of labels. The systematic 3D convolution operations were done on the input images, followed by a drop operation and a 3D max pooling operation for its down sampling part. The padding is kept the same as the rectified linear unit to add extra pixels to the edges so that the final output image dimension would remain the same as the input dimension. The convolution operation is like a 3D matrix multiplication which in our case is 3 x 3 x 3 and as it is a down sampling operation the dimension of the image will be multiplied by half resulting it to be 64 from 128. We give a 10 percent drop for the data size reduction. Then comes the 3d Max Pooling operation where the matrix of the input data is compared to a specific matrix of 2 x 2 x 2 stride and the maximum selected is taken. The same steps are continued till the image dimension reaches 16. Then we started the up-sampling operation or the expansion part. Transpose operation of 3D convolution is done multiplying the dimension to be 32 as well as concatenating the layer with the previous path layer before going for a 3D convolution operation followed by a drop out and another convolution operation. Thus, we keep repeating the steps until we get the original dimension back. The final output we get is 128 x 128 x 128 for x, y and z. We train the model on our preprocessed data and predict the data for validating it.

3.5 3D U-Net Model Training

The preprocessed data is loaded to the directory for training the data that we defined and saved as npy files. For this we customized an image loader to list out all the training images and masks saved in npy method. The data dimension used to train is 128 x 128 x 128 dimension so we select a lower batch size to load less volume. With a batch size of two we iterate our image generator which would filter the data image and load 2 x 128 x 128 x 128 x 3 with a mask of 2 x 128 x 128 x 128 x 4. We further implanted the dice loss and IoU score functions inside to evaluate the accuracy of the segmented data we come up with, through our proposed model. We take the steps per epoch and validation per epoch into consideration and train the data accordingly for 100 epochs.

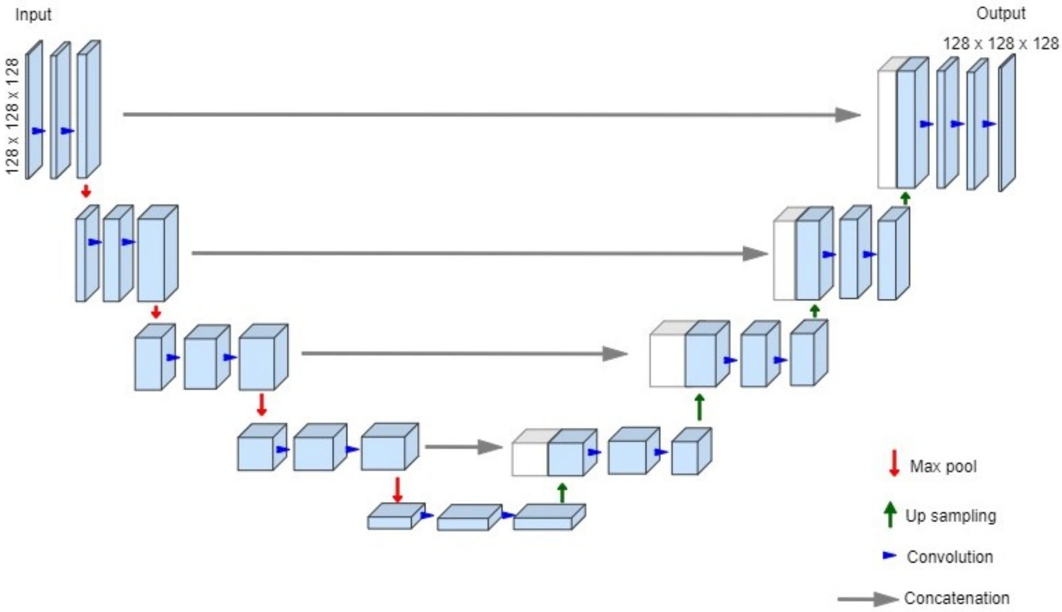


Figure 3.11: 3D U-Net Model Training

Chapter 4

Result Analysis and Comparison

4.1 Result Analysis

The mean IoU score and data loss we find through the model surpass many other models we came across that were previously done. The validation accuracy we received on average was a mean IoU value of 80.63%, with a maximum average IoU score of 83.96% and with a minimum of 64.60%. That means we have more maximum IoU scores for the validation purpose since the minimum number was 64.60% but yet we received a mean score of 80.63%, which is pretty near to the maximum value. The accuracy metrics were calculated using the IoU score calculation and the Dice coefficient calculation method. Here the IoU or the Intersection over Union score was calculated to be the quotient of the Overlapping Area over the Union Area i.e.,

IoU = The Overlapping Area/The Union Area

While the Dice coefficient was formulated as, two times the area of overlapping over the total number of pixels in length in both the images. i.e.,

Dice coefficient = (2 * (the area of overlapping/ the total no. of pixels in length in both the images))

Through these we can achieve our accuracy metrics and IoU score metrics. Now we can see the training and validation comparisons based on the accuracy metric we used and explained here.

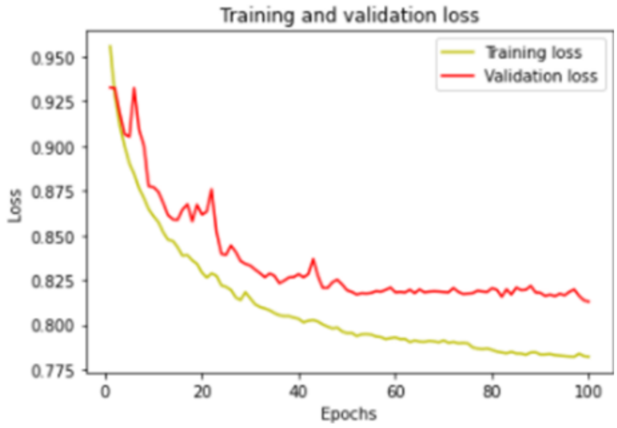


Figure 4.1: The Training vs Validation Loss Graph

Figure 4.2: The Training vs Validation Accuracy Graph

Here, in the first image we can see the validation loss is quite greater than the training loss, also in the second image we see the training accuracy quite larger compared to the validation accuracy. These perimeters prove that our training procedure is a lot efficient. For the bounding box implementation in our data preprocessing part we get a very optimized data set prepared to pass our model for more accuracy. The validation accuracy after rises to 94% as well as the IoU score changes in a good direction. The data preprocessing using a bounding box results in the following optimization as we saw in the data preprocessing.

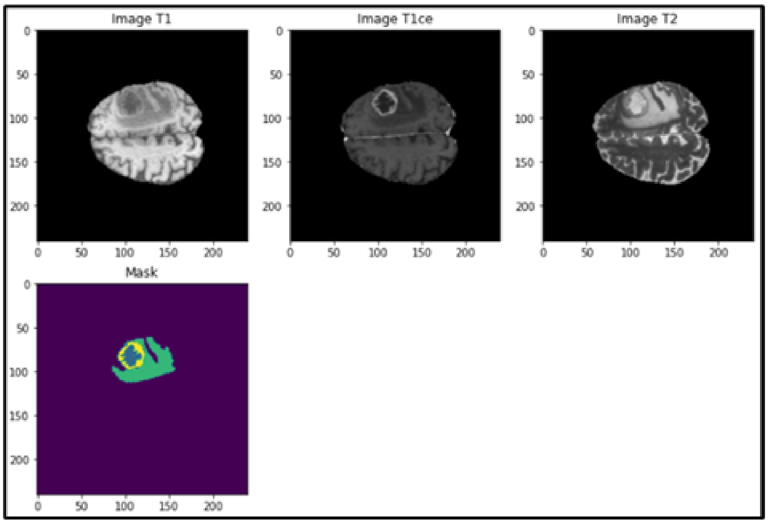


Figure 4.3: The BraTS'20 dataset Data

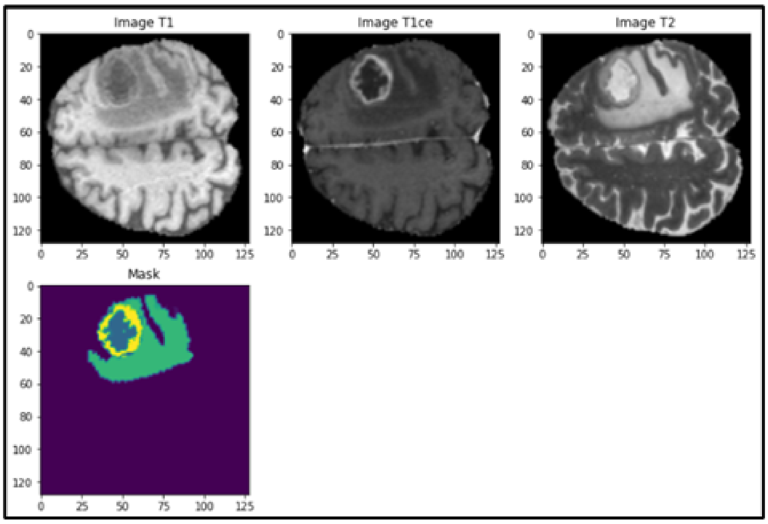
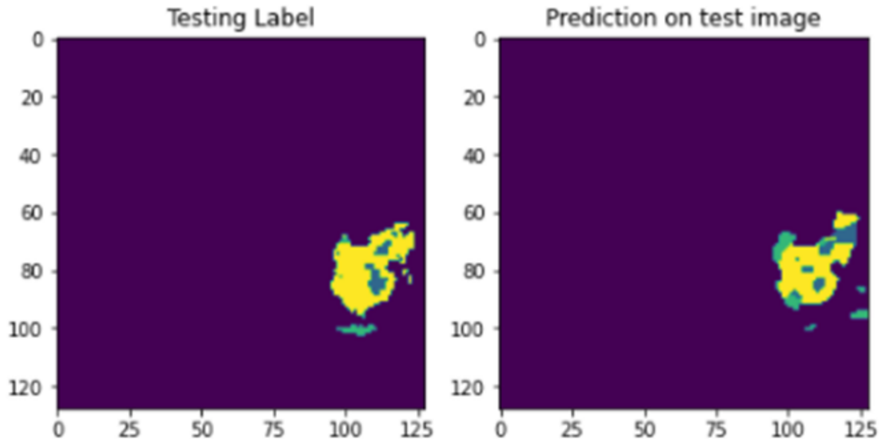
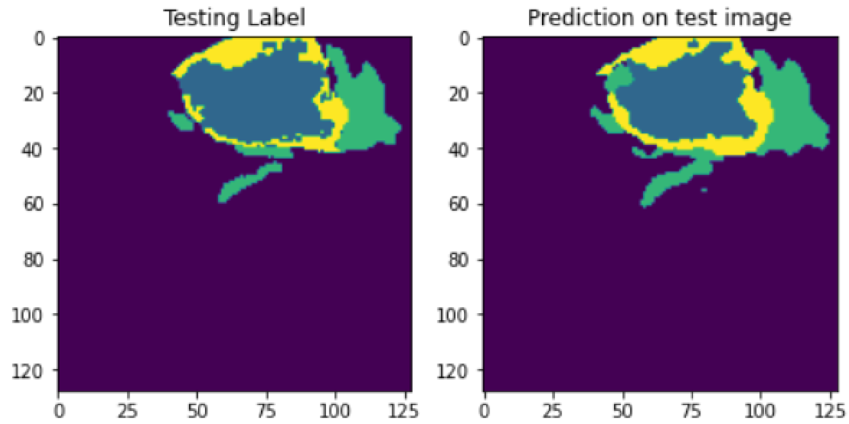


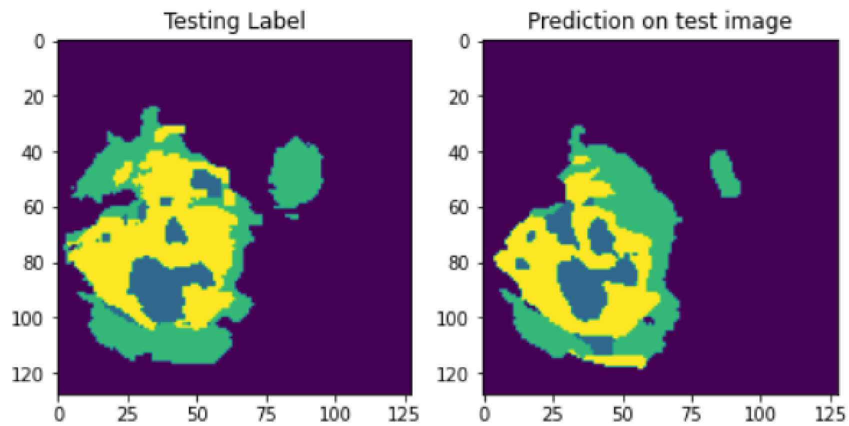
Figure 4.4: The Pre-processed Data

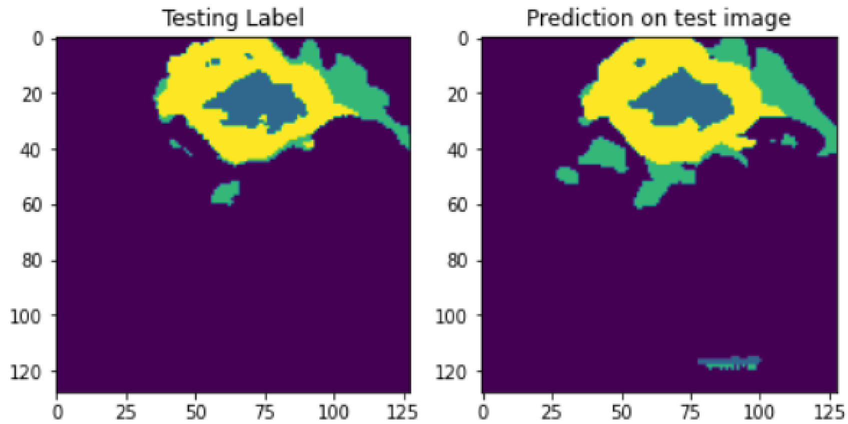
4.1.1 Observation

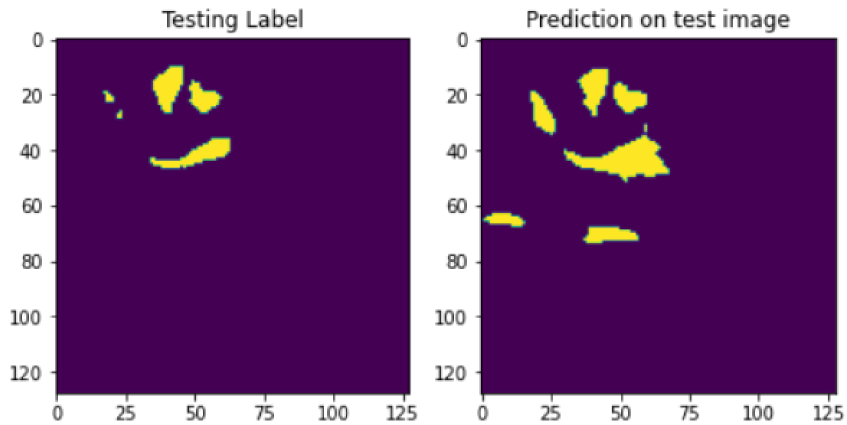
In our approach we highly focused on the data loss reduction and tried building a model more efficiently. We achieved more than half of the goal and compared to many previous models got an efficient model to train and validate. We show a few plotted prediction accuracies for some image files after the prediction with comparison as follows.

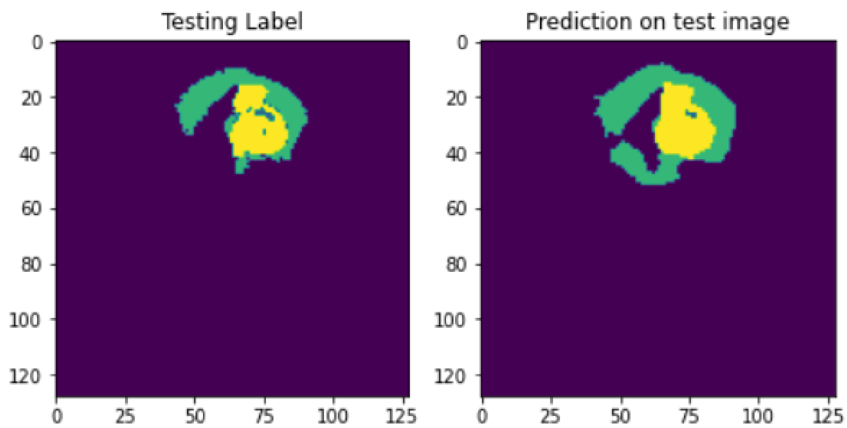

Testing Label
Prediction on test image
0 20 40 60 80 100 120
0 25 50 75 100 125


Testing Label
Prediction on test image
0 20 40 60 80 100 120
0 25 50 75 100 125


Testing Label
Prediction on test image
0 20 40 60 80 100 120
0 25 50 75 100 125


Testing Label
Prediction on test image
0
20
40
60
80
100
120
0
25
50
75
100
125


Testing Label
Prediction on test image
0
20
40
60
80
100
120
0
25
50
75
100
125


Testing Label
Prediction on test image
0
20
40
60
80
100
120
0
25
50
75
100
125

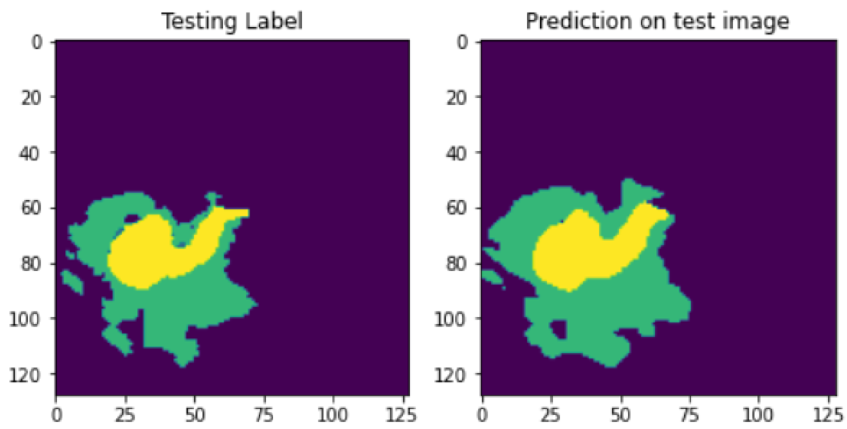


Figure 4.5: Comparison between Testing Label and Prediction on Test Image

We clearly can see here in few consecutive outcomes, the testing label and prediction on test image are quite similar with each other along with minimal difference in between. And here we can also see very few data loss in the segmentation outcomes.

4.1.2 Result Comparison

If we observe our result and compare it with other papers and approaches we get the following table.

Table 4.1: Comparison of Results

Approach Name	Accuracy (Mean Value)	Validation Accuracy
Mehta, R., & Arbel, T. [14]	78.8%	82.5%
Kao, P.-Y., Shailja, [19]	74.9%	89.7%
Rajchl, M., Lee,[16]	74.3%	82.9%
Zhou, Z., He, Z., & Jia, Y.[22]	78.5%	86.5%
Our Approach	80.6%	83.96%

From the table above we can see the following comparison where the mean average value of our approach was maximum. Again, the validation accuracy can be seen as a bit less compared to few approaches. Right below we include the Mean IoU score graphical demonstrations along with the validation accuracy graph.

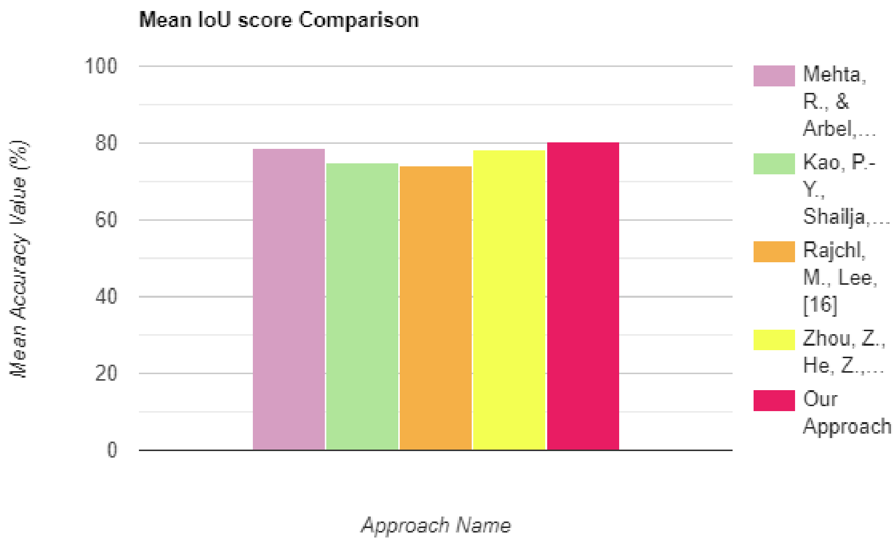


Figure 4.6: Result Comparison of IoU score (Mean Value)

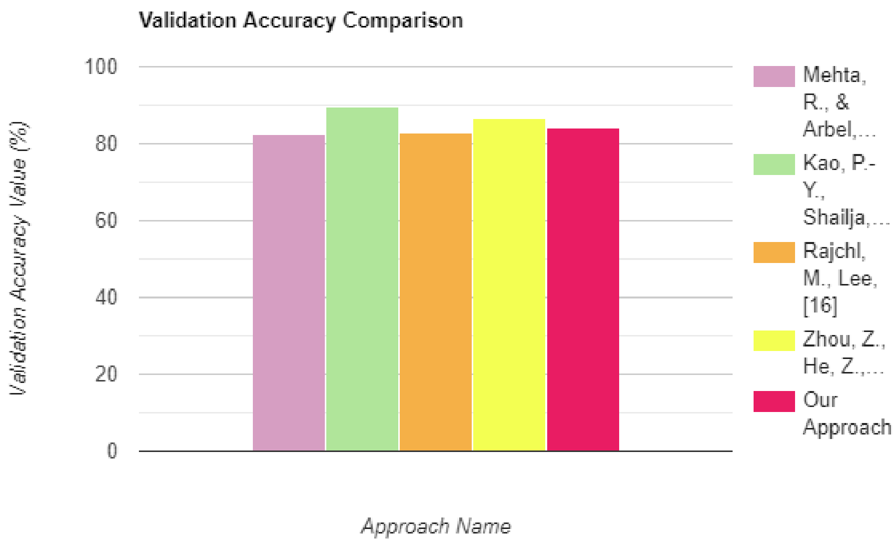


Figure 4.7: Result Comparison of Validation Accuracy

The graphical representations depict a clear difference among the values we achieved and the other approaches achieved. Our validation accuracy was a bit lower compared to the other approaches, but the mean accuracy of ours is much more compared to every other paper. That is, from our approach evaluation we got a validation accuracy of 83.96%. Though it is not better than [19], [22]. But our approach has a better IoU mean value than the others at 80.6%. It implies that our method is efficient at generating better results on average than many other proposed methods. Still to gain the highest of it we need more work to do on it. Also, it demonstrates quite well that, by applying our approach we will have no loss of information about the data.

Chapter 5

Conclusion and Future work

5.1 Future work:

For our future work, we want to make a more efficient version of this model where we can both optimize the data loss and get more accurate segmentation. The training model needs more work to maximize our validation accuracy, decreasing the amount of validation loss. We want to work further with 3D Vnet using our dataset for the semantic segmentation with a different approach. We also are looking forward to working further on GPU reduction while segmentation of data, making a large patch network augmenting our 3D Unet model with other networks and models for a proper segmentation. We want to see if the 3D tiling feature learning approach can be augmented with our approach making a dense downsampling operation. We would also like to add a dense upsampling feature to that, making sure of a proper and balanced downsampling and upsampling operation for helping in segmentation procedures. We plan to experiment more with 3D data and their segmentation in the time ahead.

5.2 Conclusion:

In our approach we used an optimizing FCN model for semantic segmentation and the FCN processed here is a 3D Unet model. Due to organ structure and shape and space variance and also for the process being complicated for the 3D models, it is a critical process to segment the 3D images. In our model, our focus was to increase the accuracy and the optimization by applying a bounding box algorithm to augment with Fully Convolutional Networks. With evidence and strong support from the data that has been gathered and arranged through observations from our recently done experimentation by the researchers we were required to conduct our experimentation. Therefore, we worked to validate and increase the accuracy of our model by using a large dataset like MICCAI BraTS'20 through the use of an advanced and powerful GPU. We already achieved more than half from what we wanted to implement but our model requires further work on it to achieve the maximum result.

Bibliography

[1] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, IEEE, 2016, pp. 565–571.

[2] G. Zeng and G. Zheng, "3d tiled convolution for effective segmentation of volumetric medical images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 146–154.

[3] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "On the compactness, efficiency, and representation of 3d convolutional networks: Brain parcellation as a pretext task," in *International conference on information processing in medical imaging*, Springer, 2017, pp. 348–360.

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

[5] E. Smistad, T. L. Falch, M. Bozorgi, A. C. Elster, and F. Lindseth, "Medical image segmentation on gpus–a comprehensive review," *Medical image analysis*, vol. 20, no. 1, pp. 1–18, 2015.

[6] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[7] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[8] D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, "Segmentation of brain tumor tissues with convolutional neural networks," *Proceedings MICCAI-BRATS*, vol. 36, pp. 36–39, 2014.

[9] Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, "Recent advances in convolutional neural network acceleration," *Neurocomputing*, vol. 323, pp. 37–51, 2019.

[10] K. Kanagaraj and G. L. Priya, "A new 3d convolutional neural network (3d-cnn) framework for multimedia event detection," *Signal, Image and Video Processing*, vol. 15, no. 4, pp. 779–787, 2021.

[11] A. Jesson and T. Arbel, "Brain tumor segmentation using a 3d fcn with multi-scale loss," in *International MICCAI Brainlesion Workshop*, Springer, 2017, pp. 392–402.

[12] Q. Dou, L. Yu, H. Chen, *et al.*, "3d deeply supervised network for automated segmentation of volumetric medical images," *Medical image analysis*, vol. 41, pp. 40–54, 2017.

[13] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, Springer, 2016, pp. 424–432.

[14] R. Mehta and T. Arbel, "3d u-net for brain tumour segmentation," in *International MICCAI Brainlesion Workshop*, Springer, 2018, pp. 254–266.

[15] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, 2021.

[16] M. Rajchl, M. C. Lee, O. Oktay, *et al.*, "Deepcut: Object segmentation from bounding box annotations using convolutional neural networks," *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 674–683, 2016.

[17] A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*, Springer, 2018, pp. 311–320.

[18] L. M. Ballestar and V. Vilaplana, "Mri brain tumor segmentation and uncertainty estimation using 3d-unet architectures," *arXiv preprint arXiv:2012.15294*, 2020.

[19] P.-Y. Kao, S. Shailja, J. Jiang, *et al.*, "Improving patch-based convolutional neural networks for mri brain tumor segmentation by leveraging location information," *Frontiers in neuroscience*, vol. 13, p. 1449, 2020.

[20] Z. Xiao, B. Liu, L. Geng, F. Zhang, and Y. Liu, "Segmentation of lung nodules using improved 3d-unet neural network," *Symmetry*, vol. 12, no. 11, p. 1787, 2020.

[21] G. Wang, W. Li, M. A. Zuluaga, *et al.*, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1562–1573, 2018.

[22] Z. Zhou, Z. He, and Y. Jia, "Afpnet: A 3d fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via mri images," *Neurocomputing*, vol. 402, pp. 235–244, 2020.

[23] R. Agravat and M. S. Raval, "3d semantic segmentation of brain tumor for overall survival prediction," *arXiv preprint arXiv:2008.11576*, 2020.

[24] Q. Yu, L. Xie, Y. Wang, Y. Zhou, E. K. Fishman, and A. L. Yuille, "Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8280–8289.

[25] J. Ngiam, Z. Chen, D. Chia, P. Koh, Q. Le, and A. Ng, "Tiled convolutional neural networks," *Advances in neural information processing systems*, vol. 23, 2010.

[26] W. Shi, J. Caballero, F. Huszár, *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1874–1883.

[27] P. Wang, P. Chen, Y. Yuan, *et al.*, "Understanding convolution for semantic segmentation," in *2018 IEEE winter conference on applications of computer vision (WACV)*, IEEE, 2018, pp. 1451–1460.

[28] K. Kamnitsas, C. Ledig, V. F. Newcombe, *et al.*, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[29] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[30] H. Chen, Q. Dou, X. Wang, J. Qin, J. C. Cheng, and P.-A. Heng, "3d fully convolutional networks for intervertebral disc localization and segmentation," in *International Conference on Medical Imaging and Augmented Reality*, Springer, 2016, pp. 375–382.

[31] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis mri," *arXiv preprint arXiv:1604.00494*, 2016.

[32] B. H. Menze, A. Jakab, S. Bauer, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[33] S. Bakas, H. Akbari, A. Sotiras, *et al.*, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, no. 1, pp. 1–13, 2017.

[34] S. Bakas, M. Reyes, A. Jakab, *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv:1811.02629*, 2018.

[35] S. Bakas, H. Akbari, A. Sotiras, *et al.*, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. the cancer imaging archive," *Nat Sci Data*, vol. 4, p. 170 117, 2017.