Pyramid Pooling Enhanced ResUNet for Accurate 3D Brain Image Segmentation

by

Md. Shawon Mollah
19201103
Farhan Tanvir Ahmed
19201107
Mahjabin Chowdhury
19201110
Iftekhar Ahmed
19201097
S. M. Rakib Hasan
22241038

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

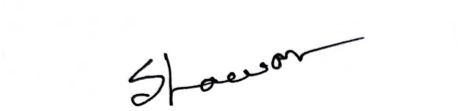
Department of Computer Science and Engineering
Brac University
September 2023

Declaration

It is hereby declared that

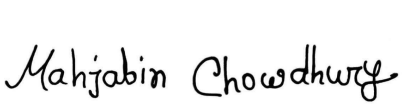
1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which 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:



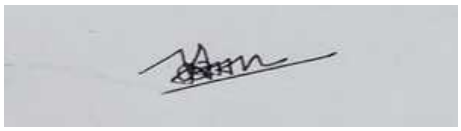
Md. Shawon Mollah
19201103

Farhan Tanvir Ahmed
19201107



Mahjabin Chowdhury
19201110

Iftekhar Ahmed
19201097



S. M. Rakib Hasan
22241038

Approval

The thesis/project titled "Pyramid Pooling Enhanced ResUNet for Accurate 3D Brain Image Segmentation" submitted by

1. Md. Shawon Mollah (19201103)
2. Farhan Tanvir Ahmed (19201107)
3. Mahjabin Chowdhury (19201110)
4. Iftekhar Ahmed (19201097)
5. S. M. Rakib Hasan (22241038)

Of Summer, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September 17, 2023.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Medical picture segmentation is important for clinical applications because it can offer valuable information on disease identification. With the inclusion of deep learning techniques, the original U-Net and ResUnet architecture has demonstrated excellent performance in 2D medical picture segmentation issues. However, it is still challenging to extend the U-Net to handle 3D volumetric medical images. This thesis proposed a redesigned ResUnet architecture with a hybrid model called pyramid pooling with enhanced ResUNet fusion with ResUnet and dialated spatial pyramid pooling from DeepLabV3+. Therefore, CNNs will effectively aid us in addressing the 3D segmentation problem.

Accurate segmentation of 3D brain pictures is critical in neuroimaging research because it allows for exact anatomical localization and quantitative analysis. In this paper, we introduce a novel framework for robust and high-fidelity 3D brain image segmentation that combines the capability of Dilated Spatial Pyramid Pooling (DSPP) with the Residual U-Net (ResUNet) architecture. The ResUNet's DSPP module improves multi-scale feature representation by aggregating information across several spatial resolutions, allowing the network which represent feature context. This integration tackles the issues given by complicated brain architecture as well as the unpredictability in picture quality that is frequent in real-world datasets in a synergistic manner. The model can comprehend complex patterns and recognize minute details in medical images thanks to attention processes, residual connections, and feature fusion methods. Brain tumors are divided in the research into medical images where the clinical data or benchmark datasets will be used to assess the proposed model. In order to assess segmentation accuracy and contrast it with cutting-edge techniques, The Dice similarity coefficient metrics will be used.

This paper will create a novel and efficient 3D image segmentation framework using a modified ResUNet architecture and enhanced pyramid pooling from DeeplabV3+.

Keywords: Segmentation; U-Net; ResUnet; Volumetric; CNN; Convolutions; tumors; 3D image; DeeplabV3+; Pyramid Pooling

Table of Contents

Chapter 1

Introduction

Medical image segmentation is a critical component of disease detection, management, and overall monitoring. Effective decision-making requires accurate segmentation of anatomical structures and diseased regions from 3D medical images. U-Net and ResUnet architecture, frequently utilized in medical research, has recently demonstrated amazing performance in picture segmentation using deep learning-based methodologies.
At the University of Freiburg in Germany in 2015, Olaf proposed an U-Net architecture which helps to segment biomedical images. It is currently one of the methods used the most frequently for semantic segmentation tasks. To fully extract the information contained in these photos, it is necessary to extend the U-Net to handle volumetric data because the majority of medical images are, as we all know, 3D in nature. As a result, a customized U-Net architecture that is designed exclusively for the segmentation of 3D medical pictures must be created. The primary goal of this thesis is to investigate the challenges of 3D medical image segmentation by developing a modified U-Net model using a hybrid convolutional neural network (CNN) deep learning model. We will attempt to increase the precision and effectiveness of the segmentation process by utilizing the strengths of both CNN and the U-Net.
Deep Residual UNET, which is often called RESUNET, is a specialized type of architecture that is generally used for tasks like semantic segmentation. The purpose of its original development is to extract roads from high-resolution aerial photos in remote sensing image analysis and was created by Zhengxin Zhang. By the change of time, by using it for tasks like identifying polyps, mapping brain tumors, and segmenting human images, among many other applications, researchers have found it to be quite versatile.

A larger architecture, equipped to analyze the 3D volumetric data, will be part of the planned improved U-Net. It will have a decoding path to recover spatial information and produce segmented maps in addition to an encoding path to extract hierarchical characteristics. The convolutional layers of the network will be modified to support 3D convolutions. As a result, the model will be able to accurately represent spatial relationships and volumetric context.

We will upgrade ResUnet and add new modified 3D deep-learning CNN modules by merging with ResUnet and DeeplabV3+ pyramid pooling layer to enhance segmentation performance. These modules may employ a variety of cutting-edge techniques, such as feature fusion techniques, attention mechanisms, or residual connections. In order to produce more accurate and reliable segmentations, we

want to improve the model's ability to spot subtle details and learn detailed patterns in medical images. The primary focus of the thesis will be the segmentation of specific anatomical structures or diseased regions, such as organs, tumours, or lesions, in medical images. The performance of the proposed improved U-Net with the hybrid modified fusion CNN model made by integrating ResUnet and DeepLabV3+ will be evaluated by comparison with publicly available benchmark datasets (BRAT-2020).

This thesis aims to develop the field of 3D medical image segmentation by developing an innovative and practical methodology based on a hybrid CNN model and a modified ResUNet architecture. The recommended method has the potential to advance medical picture analysis while also enhancing clinical judgment, patient care, and medical research.

1.1 Problem Statement

The technique of employing technology to see different components that make up the human body with the purpose of diagnosing, monitoring, and treating various medical pathogens is known as medical imaging. Medical image segmentation is the process of extracting regions of interest (ROIs) from three-dimensional image data, such as that generated by a computed tomography (CT) or magnetic resonance imaging (MRI) scan. The major objective of segmenting this data is to identify anatomical areas necessary for a particular investigation. For instance, the simulation of physical qualities and the virtual placement of CAD-designed implants inside a patient are both instances of research that need the use of particular anatomical locations. The process of segmenting medical images might be one that requires a lot of time [16]. Therefore, automated segmentation of medical images is essential for the immediate treatment of patients. Due to the tremendous unpredictability and complexity of medical pictures, as well as the fact that noise often contaminates these images, automatic segmentation of medical images is a job that is very difficult to accomplish. For disease diagnosis, medical image processing must be highly accurate; however, pixel-level or voxel-level segmentation makes it challenging to distinguish between cells and organs [1]. According to [9], FCN exhibits a significant advance in segmentation accuracy, but its segmentation of tiny target objects is inadequate. FCN problem exclusively highlight the high-level feature information which classifies pixels and aid in extracting low-level feature details, resulting in imprecise network segmentation results. In the realm of medical image analysis, some of the challenges that are faced include low accuracy of picture classification, limited segmentation resolution, and poor image attractiveness. To identify collections of image data required to train a segmentation model, the majority of existing works rely on manual selection, random sampling, or previously trained machine learning (ML) models [22].

1.2 Research Problem

3D image segmentation is not an easy task and needs several research where many problems will have to undergo for better accuracy. The segmentation will become tedious when there are irregular shape area portions of the MRI 3D images. This may increase the loss percentage of the particular model training and validation testing.

Data preprocessing will become inefficient when there are noises and distortion in the 3D MRI NIFTI formatted images. Therefore, it may need to annotate and find the distorted data which may resolve the issue for the data preprocessing part. Again, low-quality 3D images dataset will affect low performance which will cause a barrier in the segmentation.

Deep learning with the U-Net model always requires annotated datasets for training. However, it will become challenging if the dataset is not properly annotated because it is a supervised learning model that requires properly annotated data. Lack of annotated data or missing annotated will cause overall performance downfall. Moreover, improper way of data splitting will show unexpected training and validation test accuracy results.

Some image segmentations were needed in emergencies in a short period of time in medical scenarios. However, processing and analyzing large 3D MRI image data through TensorFlow are cost-effective and require tracing with high-performance GPU and CPU for acquiring and performing a large number of epochs and per steps in epochs. Therefore, it may not always be possible to serve the segmented data to the medical in a short period of time. Moreover, the performance decrease issue was also caused by not having enough 3D image datasets.

These research problems highlight the characteristic of 3D image segmentation which is essential for addressing future development in the medical sector by adding DSPP layer for enhancing ResUNet. Therefore, it upgrades the algorithm model into lightweight and advances the processing time as well.

1.3 Research Objectives

The main goal of the research objectives was to segmentation of the 3D MRI images. Segmentation of the 3D images is the hardest task because not all the models have given perfect and accurate segmentation results. It will give the dice coefficient result better but the validation test will not give the expected result. Therefore, we have classified the research objectives into subtasks which can help to break down the research objectives with ease mannerly way.

The main target is the accuracy of the models. The Mask R-CNN model can be helpful in segmenting images but will only be workable with the COCO dataset. So, all datasets will not be suitable for this model. With the Gated-SCNN model, it helps to identify the image shape but the real target of efficiently segmenting a particular image missing with this model. Therefore, the DeepLabV3+, Unet and ResUnet model shows some points of view segmentation on 3D MRI images.

Another objective is the efficiency in segmentation. Nowadays in the medical sector, the efficient way of segmenting 3D images are in not very good at numbers and results in some errors as well. This situation hits every patient and brings hassle to doctors. To overcome or maximize the efficiency, we are come up with a merged ResUNet and DeepLabV3+ model with an extra conv layer in the U architecture. This will not promisingly increase efficiency by 100% but will fulfill part of the research objective for 3D image segmentation by merging the ResUnet and DeepLabV3+ models.

- Segmenting 3D MRI images is difficult due to existing model errors. We divided the research objectives into subtasks to assist break them down.
- For picture segmentation, the MASK R-CNN model works well with COCO dataset but only for accuracy.
- The Gated-SCNN approach helps identify image shape but fails to efficiently segment an image. Thus, by merging DeepLabV3+ and ResUNet model gives the segmented 3D MRI pictures by point of view.
- ResUNet model with DeepLabV3+ promise 100% efficiency in medical image segmentation.

Chapter 2

Literature Review

Technologies in the medical sector have evolved with new advanced methodologies models which can detect and segment 3D magnetic resonance images. These 3D MRI images are acquired from ultrasonic which is later than pre-processed these collected information data and trained and analyzed in efficient compatible models for having better accuracy levels respectively. With this literature review, our aims provoke the field of 3D image segmentation with different challenges where the main clinical research part begins.

Segmentation is a technique used in medical imaging to recognize and distinguish between various bodily features, such as organs, tissues, and lesions. This is crucial because it enables medical personnel to measure and track the development of ailments as well as diagnose and treat diseases more precisely. The improvement of medical diagnosis and treatments could be greatly enhanced by the development of a deep learning-based 3D image segmentation technique for imaging data. It is possible to develop a system that can precisely segment various body structures with a high level of accuracy by training a deep learning algorithm on a huge collection of 3D medical photos. This might be especially helpful when manual segmentation is challenging or time-consuming.

The paper [6] introduces a brand-new model called PointNet, a unified framework for applications like object segmentation and categorization. A collection of points known as point clouds avoids imperfections. Only the three coordinates of each point are used to represent it (x, y, z). Point clouds are directly inputted into the PointNet architecture, which then produces either class labels or point segment-level labels. The classification network accepts n input points, transforms the input and features, and then pools all of the points' features together. A set of classification scores for k classes is the end outcome. A classification net extension is the segmentation network. It creates per-point scores by fusing global and local features. In terms of computing performance, PointNet outperforms techniques like MVCNN and Subvolume (3D CNN). Additionally, PointNet uses far less space than MVCNN. It provides a single framework for numerous applications, such as component segmentation, scene semantic parsing, and object classification.

In the work [18], a different viewpoint on 3D picture segmentation is presented. Here, a brand-new deep learning model called KerNet is recommended for identifying the eye condition Keratoconus. An end-to-end deep learning model that

uses the raw data from the Pentacam HR system can recognise the illness. The Pentacam HR system produces five slices of raw data (CUR-F, CUR-B, ELE-F, ELE-B, and PAC). To receive the five slices independently, five branches were added to KerNet. Multi-level fusion—low-level and high-level fusion—was created to account for any potential associations between the five slices. To improve the low-level fusion, in particular, the attention mechanism leveraging spatial modules was used. Only straightforward channel-wise concatenation is used for high-level fusion. Our own dataset-based comparative assessment studies have shown that suggested KerNet can outperform techniques for the clinical keratoconus.

In the paper [7], different methods submitted by different teams in the challenge of 3D shape part-level segmentation and reconstruction from single-view images are discussed. However, for our research, we will only be confined to the part-level segmentation of the 3D shapes. In Submanifold Sparse ConvNets, The point clouds are voxelized into a sparse 3D grid, with the points scaled to their L2-norm. valid size-3 sparse convolutions (VSCs) were stacked. The collection of active sites does not change since the filter outputs are only computed where the input locations are active. To partition data sparsely, this is perfect. Parallel routes in the ResNet fashion using strung convolutions were added, VSCs, and deconvolutions to broaden the receptive field. Pd-networks is another model used to classify point clouds. A bottom-up traverse of the Pd-tree is made by the recognition process beginning with certain basic representations given to leaves (which correspond to individual points). The vectorial feature representation for the parent node is computed for each calculation in the bottom-up step using the vectorial representations of the children. With the top-down pass, which computes the new feature representation for children from the representations of the parent nodes, the bottom-up pass is reversed. The bottom-up and top-down passes are connected by skip connections, resulting in a typical U-net design. In another model, a PointNet structure was redesigned as a dense connection block like DenseNet and the number of kernels was reduced. All layers were connected in the block and local & global features were extracted. Adversarial Loss for Shape Part Level, The PointAdLoss Data from a point cloud is segmented, and the pointwise class probability is calculated. It employs a discriminator to determine whether the probability is the ground truth label vector or the prediction vector after simultaneously receiving the data and the probability. The segmentation process in the K-D Tree network uses an order-invariant representation of the input point cloud. The point clouds' representation is created by building a K-D tree for them. A fully convolutional network with skip connections receives the output. Each category is trained on a different network. Therefore, segmentation results are trained from beginning to end using the available training data.

Superpixel segmentation is discussed in the work[23]. It divides spatial images into a number of semantic sub-regions with similar defining properties. In order to generate superpixels, graph-based or gradient-based approaches for hyperspectral images are considered. Superpixel-Guided Classification is one of several classification techniques that generates a classification map based on the probability that each pixel belongs to each class. The classification map is then further optimised or regularised in accordance with the segmentation map. In direct classification utilising superpixels, features are computed directly on the superpixels employed for classification. In superpixel-based deep learning, these networks are used to extract HSI features. It is possible to extract spectral, spatial, and spatial-spectral features. The superpixels are processed by means of ex-deep belief networks (DBN), convolutional neural networks (CNN), recurrent neural networks (RNN), generative adversarial networks (GAN), and so forth.

The paper [21] examines image segmentation, including fully convolutional models, which contain only convolutional layers and can produce segmentation maps of the same scale as the input image (ex-ParseNet). If we look for options we can get encoder-decoder-based models and it is based on two parts. One is a convolutional layer-based encoder and another one is a deconvolutional network. If we deep dive we can see that a deconvolutional network always receives a feature vector as input in return it gives probabilistic maps. There are also Criss-cross attention networks in addition there are attention-based models that repeatedly improve segment boundaries from lower and it outperform average and maximum pooling by employing the attention mechanism to evaluate the significance of features at multiple locations and scales. In addition, there are generative models and adversarial training, which employ a zero-sum game framework and include a generator and a discriminator. The discriminator attempts to distinguish synthetic data from real data, whereas the generator attempts to generate synthetic data that resembles a training dataset.

Different point cloud semantic segmentation strategies are addressed in the paper [17]. The traditional supervised machine learning techniques come first. Neighbourhood selection, feature extraction, feature selection, and semantic segmentation are the four stages of the process. In addition, deep learning approaches can extract using more than two hidden layers. The other early method utilized in deep-learning-based PCSS is the voxel-based method, which combines voxels with 3D CNNs. Both the unordered and the unstructured issues with the raw point cloud are resolved by voxelization. Similar to how pixels in 2D neural networks are treated, voxelized data can also be processed further using 3D convolutions.

In paper [20] the author applied Deep-Learning-Based Medical Image Segmentation. It can be described as a set theory. Here, 2D CNN and 3D CNN have been applied. FNC (Fully Convolutional Neural Networks) it is the most advanced deep learning in medical image segmentation. There are lots of networks that have been used here(SegNet, U-Net,2D and 3D U-Net, V Net). In different human organs such as the brain, eye, and chest it has been used mainly in CT, MRI, PET and X-ray. Also, there are lots of limitations of this segmentation because it is different from the natural images Limitations of existing medical image data sets, etc. Document [13] Utilizing deep learning algorithms, hierarchical and useful information has been extracted from 3D forms. Three popular uses of 3D deep learning are shape segmentation, recognition, and classification. The efficacy of form descriptors in all of these ways is yet still constrained by representation. Volumetric, points-cloud, and mesh-based methodologies have all been used in this work. In the A-CNN architecture, there are two networks: a classification network and a segmentation network. It is a point cloud A-CNN framework.

[14] In the paper One of the most significant computer vision research technologies is semantic segmentation. It is mostly reliant on two-dimensional images, and as a result, the segmentation performance is subpar due to the constraints of two-dimensional data in terms of occlusion and other factors. Two-dimensional data has inferior segmentation performance due to restrictions in occlusion and other areas. Indirect segmentation techniques include point CNN, RSNet, SO-net, and multi-view-based techniques like MVCNN, SHAPENET, and VOXnet, among others. They have preserved the execution duration, precision, spatial complexity, etc.

They employed 3D LiDAR datasets from robotics, mobile mapping, and autonomous driving in this research study [19]. It solves methodological and data set development issues as well as concerns with data hunger. The static dataset, sequential dataset, and synthetic dataset are used to make it work. Additionally, there are semanticPoss, semanticKITTI, and semantic3D. They have demonstrated that this dataset only accurately depicts a tiny number of real-world scenarios. Large domain gaps result in severe performance reduction, even while the data annotation is accurate. Three representative 3D LiDAR datasets are analyzed statistically once the primary datasets are done. What constitutes the model's crucial elements? As a result of the segmentation, a structured survey has been created. There may always be a need for data. It is crucial to rely on a method that can annotate data simply.

In this paper [3], the authors have created a refined architecture, "FC Network." The architecture has been modified to operate with lesser images which gives more accurate segmentation. They have very little training data, so data augmentation has been utilized with available images U-net's network architecture essentially consists of two components. The first device is an encoder, while the second device is a decoder. Here, each layer is convolutional in its entirety. It consists of repeatedly applying two 3x3 convolutions. Border loss in pixels causes each convolution cropping which is later on ready for the training datasets. This dataset

has a momentum of 99% which cause optimization into the deep neural network. The vast amount of CNN layer create a weight by initializing appbility of U-Net segmentation. This U-Net segmentation gives the biomedical segmented images.

In this article [27], they have proposed a network called NUMSnet which is basically a nested multi-class segmentation which can identify and label different structures or regions of interest within the image. They have used 3D image stacks of CT or MRI images. The NUMSnet required training images and once it has been trained, the three main analytical questions regarding multiclass segmentation in 3D medical stacks will appear. The research also conducted into an examination of loss function curves, with a particular emphasis on the Unet++ model's deep-supervision feature. The results show that the Unit model and its modifications are effective for multi-class segmentation, even when trained with only 10% of the volumetric scan's annotated data. Therefore, The authors conducted four sets of experiments, including a modification of the training dataset sequence to observe variations in segmentation performance. But there is also one key limitation when it scans written text on them. Moreover, multi-class segmentation with one key by using U-Net between the ROI size will impact significantly in the training period. For example, images of Lung spot regions are larger because of Dice's coefficient.

In this article [5], they have worked on atrous convolution image segmentation. This CNN, detailed convolution technique is used for segmentation which controls the view of the field of resolution. They have also improvised the 'Deeplab3' system significantly and because of that, they have seen a lot of improvement. They resolved the extraction of the resolution features. At first, there is an image pyramid and after that, there is an encoder-decoder and lastly the deeper w. Artrous convolutional wih. Here, they have applied the method of various convolutions for dense feature extraction and with this, they go deeper and deeper. There have been three 3×3 convolutions in blocks. Also, there is multi-grid methods have been applied. ResNet -50 and ResNet-101 have also been used to go deeper with various convolutions. Finally, in conclusion, we can say that their proposed model "DeepLabv3" employs various convolutions for capturing long-range context with lesser difficulties. DeepLabV3 used training time with 32 duplicates which reduced maximum hours of training time.

In paper [24], a novel DeepLab V3+ model is discussed, and it is shown to be capable of resolving the issues with the DeepLab V3+ model. The research makes use of a technique known as ResNet2 and has established a number of different loss functions in order to limit the model's overall efficiency. The procedure of DeepLab V3+ has been enhanced in this regard in two ways: the gathering of multiple scales contextual data and the use of the basic characteristics. After comparing the findings of the study with those of the original DeepLab V3+, we came to the conclusion that the PA, mPA, and mIoU all increased by 0.1%, 2.233%, and 2% respectively. However, this strategy has several limitations, such as the fact that training a model requires more processing power, slows down the execution in real-time, and makes it more difficult to guarantee the performance.

According to [28], The UNet 3+ and CBAM model combination is explored, and it is shown that this configuration offers more sophisticated capabilities. In the context of the study, a methodology known as Redesigned Multi-Scale Skip Connections and CBAM in the Decoder was implemented. This research uses three different types of datasets, each with its own unique characteristics, in order to verify its effectiveness. The data of the research were compared with UNet and Unet+, and after doing so, we got to the conclusion that the model that was presented had a superior segmentation performance. This method improves the precision and accuracy of image segmentation while also enhancing the capability of feature extraction to get a more in-depth comprehension of the picture. However, this technique suffers from a number of drawbacks, such as the fact that deformable convolution is applied without first assessing whether or not it can improve the feature extraction.

The difficulties of segmentation, detection, and classification in medical images are addressed in this study [11] by applying U-Net, a deep learning-based technique. The models that have been offered here basically are tested on three benchmarks of datasets. The segmentation that has been used here is the skin cancer and lung lesions among affected people. In this article they have used U-net architecture and the architecture uses recurrent convolutional neural networks. Also, there are some suggestions for some models which have done some experiments and after doing the experiments they have found three datasets with the same amount of parameters which basically proved that those suggested net models were found to outperform the Segnet, U-net and Res net models for segmentations. Our qualitative and quantitative analyses led us to the same conclusion training that those suggested models resulted in better test results. However, there are a few problems with this strategy. As an illustration, suppose you use deformable convolution without checking if it enhances feature extraction. In this paper [25]. we can see this in biomedical picture segmentation; it resolves the segmentation via Unet where a fully convolutional network has been used. This U-Net architecture has the consistency of the encoding phase and subsequent decoding phase. To get more accuracy in contextual information the TransUnet combines and transforms it as an encoder and works accordingly. Basically, it is a U-shaped design-based model architecture but it helps to calculate accurate global context design beside low-level CNN architecture. In this article their main goal or we can say their main work is to give an efficient concept regarding Unet based on different neural machine learning networks. Lastly, we can say by this article that it also has some limitations when it looks for accuracy in order to get accurate results.

This framework uses Tanimoto, an altered version of the Dice loss algorithm [15], which is an innovative multitasking deep learning architecture. This study examines the performance of this approach to that of previous research findings and explores the incorporation of context information to improve the algorithm's performance. The proposed trained handling multiple tasks model is the most effective variation because it forecasts not only the mask that is used to segment but also the outer limits of multiple classes, the distance transform (which indicates

the topological interaction of the components), and the factual restoration of the image being used. The conditioned multitasking model is the most effective model version we have created in terms of efficacy. Tanimoto loss is a recently developed and applied technique that improves training convergence and works well even with very unbalanced data sets. It is possible to conclude that the integration of ResUNet is trained to perform several tasks and the recommended loss function is an effective tool for executing semantic segmentation tasks after comparing their performance to that of the UZ 1,89.3 RIT L7,92.6 DST 5,92.5 CAS Y3,92.2 CASIA2,93.3 DPN MFF,92.4 HSN+OI+WBP model. When compared to previous state-of-the-art approaches, the performance of a unique depth-supervised strategy simultaneously requires fewer parameters, as stated in [26].In this method, MLP modules are used in conjunction with residual blocks to segment images of intervertebral discs. Its 91.74% DSC and 84.74% Jaccard are superior to that of U-Net, IVD-net, and U-net3+. When compared to the IVD-Net, a few drawbacks, including effects of migration fields, artifact motion, and MRI noises are the result of technical issues including those encountered during image capture and the limitations of imaging equipment. Some tissues of the intervertebral disc are excessively tiny and inconsequential, with too much similarity, poor distinction, weak edge contrast, and fuzzy borders. The paper [12], describes an innovative architecture devised by ResUnet++ for medical image segmentation. In order to generate the brand-new Kvasir-SEG dataset, an experienced gastroenterologist was recruited to aid in the annotation of the polyp class in the original Kvasir dataset. This architecture could function more efficiently if residual units, compression and excitation units, ASPP, and attention units are incorporated. The Kvasir-SEG dataset obtains a dice coefficient of 81.33 percent and a mIoU of 79.27 percent. In addition, it achieves a dice coefficient of 79.55 percent and a mIoU of 79.62 percent for the CVC-612 dataset, all of which demonstrate that ResUnet++ outperforms U-Net and ResUNet.This model has a defect in that it requires more parameters, which increases its execution time

Chapter 3

Methodology

3D image data needs to be preprocessed first to be loaded into a model. Data splitted into training and validation portions and later it sorted 3D images with segmented files and mask files. The sorted files are then read with the image loader function and stored in the list data structure.

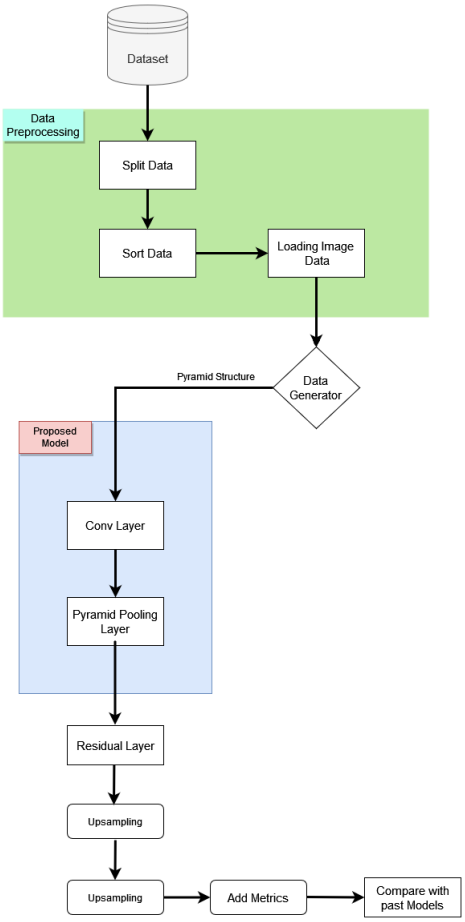


Figure 3.1: Flow chart of Proposed Merged 3D Segmentation Model

Preprocessed data is then bypassed into our model. First, the input will go through a dilated spatial pyramid pooling layer, the residual blocks, upsampling layers and output layer and generate the result. We have added several metrics namely dice coefficient, IoU and pixelwise accuracy. Lastly, the results will be passed through a comparison process.

3.1 Dataset

We have utilized the BRATS (Brain Tumour Segmentation)[2] [4] [8] dataset for training and validating our models. This dataset, specifically the BRATS 2020 version, is an exhaustive compilation of multimodal MRI scans for the segmentation and analysis of tumors of the brain, specifically gliomas. It is intended to evaluate cutting-edge methodologies in the field of segmentation. This NIFTI formatted

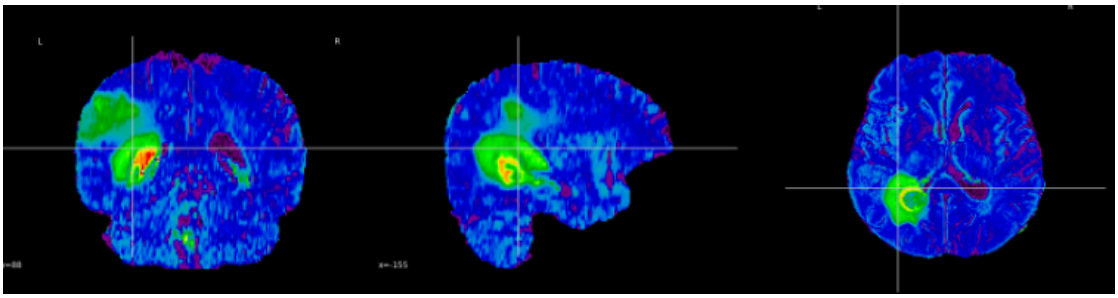


Figure 3.2: Brain tumour Flair (x, y, z) axis images.

file contains 3 axes of different side views of the particular brain image. One single 3D image of this dataset contains 240 units for the x and y-axis and 155 units for measurement for the z-axis respectively. This x, y and z axis of different angles of sides of view represents the 3D brain MRI image.

3.2 Data Description

The dataset is comprised of pre-operative MRI scans obtained from multiple institutions, contributing a variety of imaging information. The scans are offered in the NIfTI file format (.nii.gz) and include four distinct modalities:

- Native T1-weighted (T1): This modality provides baseline brain anatomical information.
- Post-contrast T1-weighted (T1ce): This modality is contrast enhancement for tumour regions and aids in identifying tumour borders.
- T2-weighted (T2): This modality emphasizes edema and provides additional information on tumour characteristics.

- (FLAIR): This represents the fluid signals that help to enhance the visibility of edema and tumor infiltration.

The scans within the dataset were acquired using a variety of clinical protocols and scanners from 19 distinct institutions, assuring a diverse and representative collection of imaging data.
The annotations in the dataset were painstakingly conducted by one to four raters using a standard annotation protocol. Neuroradiologists with extensive experience evaluated and accepted the annotations. These are the segmentation labels:

- GD-enhancing tumour (ET) - Label 4: This denotes the region of the tumour that displays enhancement following gadolinium contrast injection.
- Peritumoural edema (ED) - Label 2: This refers to the region surrounding the tumour that displays edema, which is induced by the tumour.
- Necrotic and non-enhanced tumor core (NET): This represents a certain portion of the central tumor.

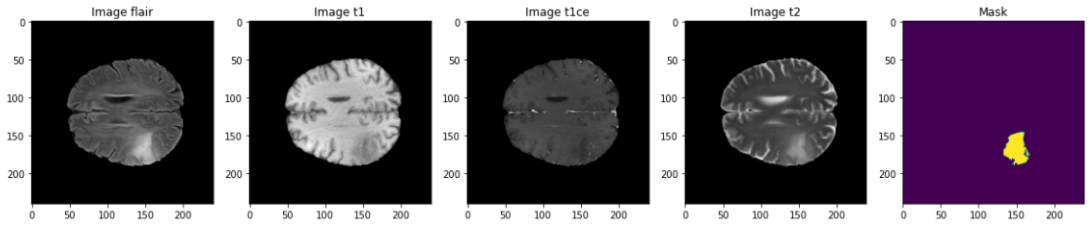


Figure 3.3: Categories in Dataset Files

The dataset's data have undertaken preprocessing steps to assure consistency and comparability. These phases include co-registering the scans to a standard anatomical template, interpolating to a uniform resolution of 1 mm3, and skull-stripping to eliminate non-brain tissues.
In addition to imaging data and manual segmentation annotations, the BRATS 2020 dataset contains clinical data pertaining to patient outcomes. This includes information regarding overall survival, progression status evaluation, and uncertainty estimation for the predicted tumour subregions. The (BRATS 2020) dataset gives a resource that gives medical image analysis. Moreover, the development of the algorithm for brain image segmentation gives the prediction and gives the differences between tumor and normal brain tissues.

3.2.1 Data Preprocessing

The total number of 3D images is 2,345 where the training 3D image set is 371 and the validation image set is 127. So in order to train and evaluate models, we divided dataset into three categories: train, test and validation portion of dataset.

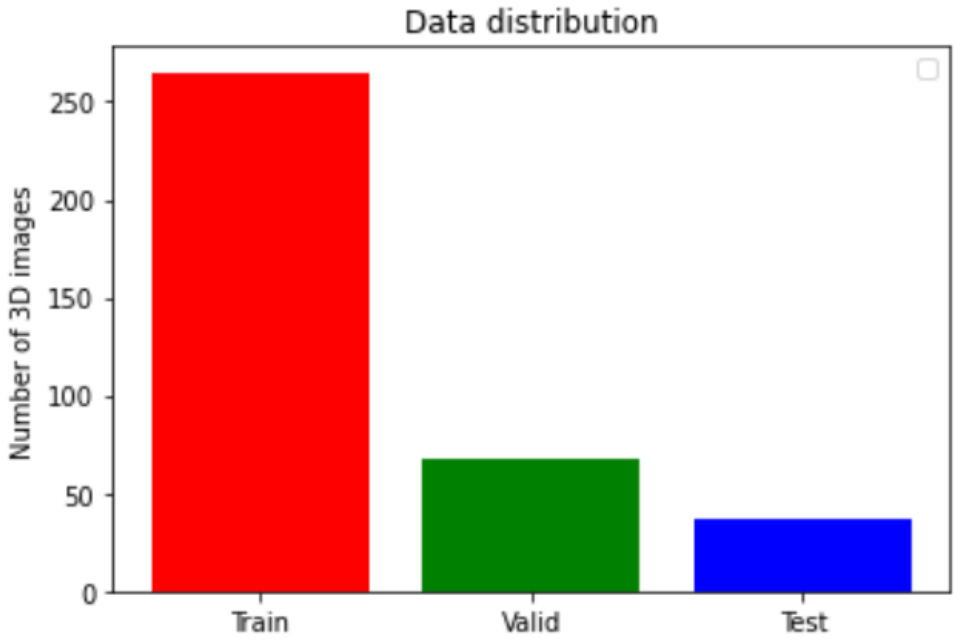


Figure 3.4: Data Distribution Graph.

We've designated 20% of our data as validation, 10% as test, and the remainder as training. The following chart depicts this distribution of images into training, test, and validation sets. Firstly, we have stored the training and mask dataset path into a variable. After storing the path splitting the dataset into "x-train", "x-test", "y-train", "y-test", "x-val" and "y-val". Then we implemented a test size for validation is 0.3 and a test size is 0.15 for the training part respectively. After that, images are loaded into a variable with two different functions "load image" and "load mask" which is then ready for training, validation and testing process.

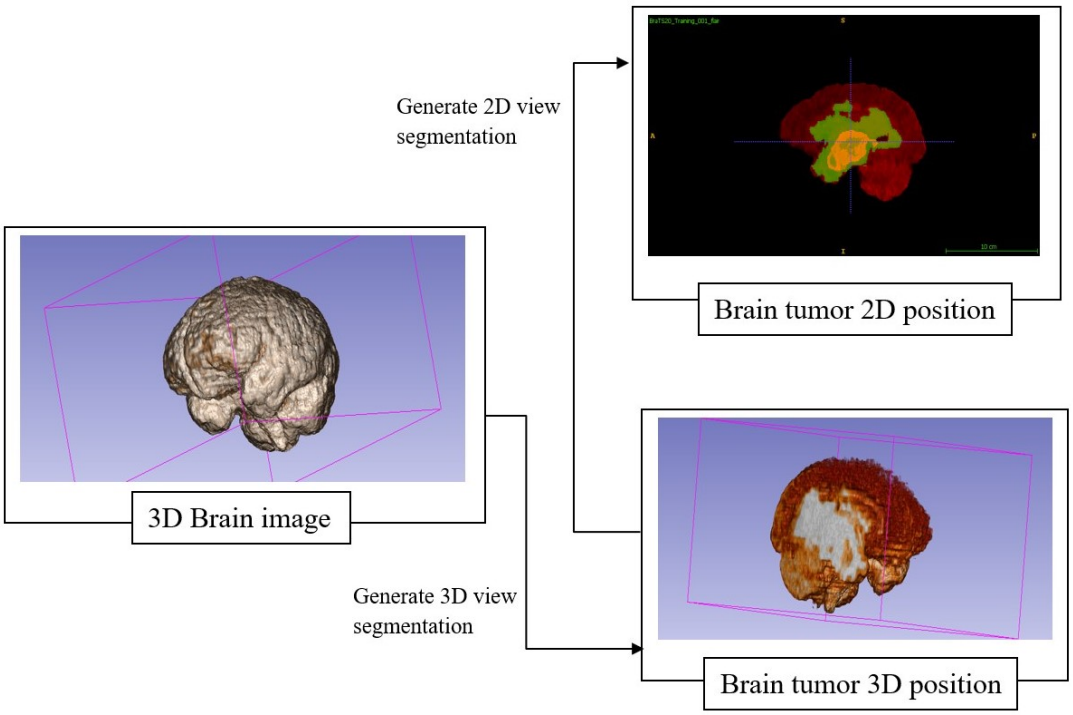


Figure 3.5: Original nifti file gets converted in 2D axis images for segmenting Brain tumor.

To ensure data integrity, in figure (3.5) raw NIfTI pictures are first screened for abnormalities. After that, spatial normalization techniques are used to put all pictures into a shared anatomical space, allowing for meaningful group-level comparisons. Following that, intensity normalization and bias correction are used to reduce any confounding effects by minimizing fluctuations in signal intensity. Fusion techniques of Pyramid Pooling and ResUNet are used to combine data from many imaging modalities, resulting in a more thorough knowledge of brain anatomy and function. Therefore, the 2D images are then ready to be segmented with the fusion technique model which will be implemented in the Dice coefficient for observing the accuracy of segmentation.

3.3 Model specification

In this research, we have used two models for training and testing the data. They are DeepLabV3+ and U-Net.

3.3.1 DeepLabV3+

General Introduction

DeepLabv3+ is an extension of the DeepLabv3 model, integrating advanced features such as atrous convolution and the encoder-decoder structure. DeepLabv3+'s encoder-decoder architecture enables the extraction of fine-grained features, semantic comprehension, and precise localization.

3.3.2 Architecture of the Model

The core of the DeepLabv3+ architecture is comprised of the following essential elements:

Encoder Network

In DeepLabv3+, the encoder network serves a crucial role. The ResNet-101 network is used as the backbone for feature extraction in this model. ResNet-101 is a variant of the ResNet architecture with 101 layers and residual connections that mitigate the problem of vanishing gradients. The ResNet-101 model is used to extract rich and discriminative visual features from image inputs.
Important encoder network layers include:

- Convolutional Layers: The Convolutional layers extract local patterns and structures from the input image by performing feature extraction operations.
- Residual Blocks: ResNet-101 is made up of residual blocks, each of which is integrated with multiple CNN layer connections. This enables to solve the gradient vanishing problem.

Atrous Spatial Pyramid Pooling (ASPP)

ASPP describe as a key component in DeepLabv3+ that which holds multi-scale context information through the use of Atrous CNN with multiple rates. Atrous convolutions permit which expands the network receptive area without downsampling the map's feature. By employing multiple atrous rates in parallel, ASPP captures contextual information at various scales, allowing the model to manage objects of varying proportions.
Important ASPP module layers include the following:

- Atrous Convolutional Layers: These layers utilize atrous convolutions at varying rates to capture context data at multiple dimensions. The rates influence the magnitude of the receptive field by determining the spatial sampling rates of the filters.

Decoder Network

A decoder network is utilized by DeepLabv3+ which helps in to improve in the area of the spatial resolution of segmentaion. The decoder network uses bilinear upsampling and 1x1 convolution to generate feature maps with the same resolution as the input image. To improve localization precision, the decoder network incorporates skip connections from previous phases of the encoder network.
Important decoder network layers include:

- Upsampling Layers: The upsampling layers increase resolution if the spatial ffefature map which mapped the original input image. They assist in the recovery of fine-grained details that were lost during the encoding procedure.
- Convolutional Layers: The layers of convolutional of the decoder network process the upsampled feature maps in order to extract high-level semantic information.

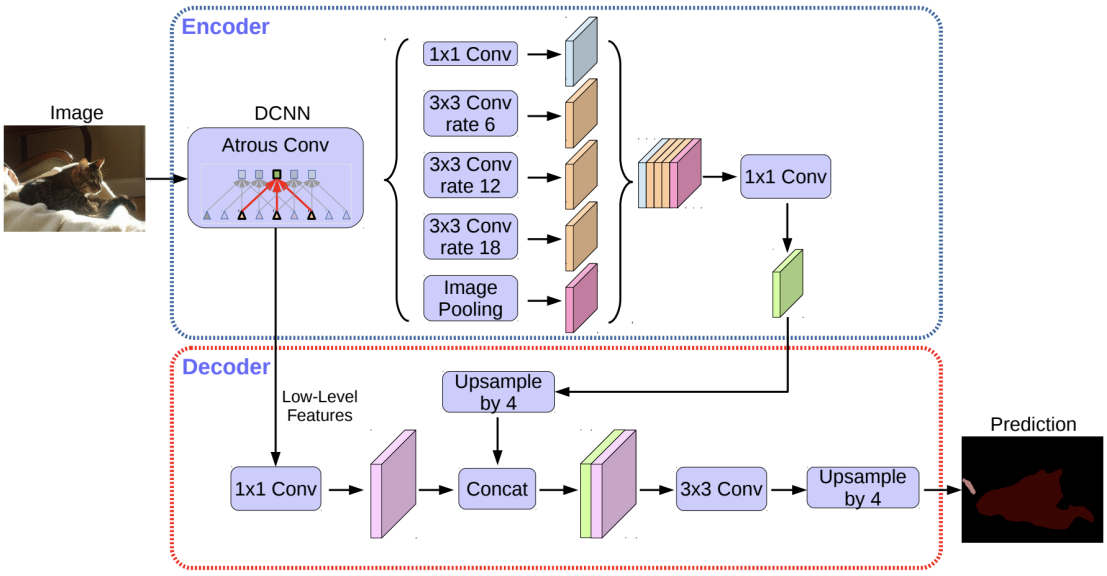


Figure 3.6: The DeepLabV3+ Architecture

Training and Inference

DeepLabv3+ is trained using the integration of pixel-wise entropy loss with loss functions to guarantee both precise object localization and fine-grained boundary prediction. Mathematically, cross-entropy can be defined as below:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c}) \tag{3.1}$$

Equation 3.1 represents the loss function.
where:

- N represents total number of pixels in the image,
- C represents number of classes,
- $y_{i,c}$ holds the ground truth label for pixel i and class c which represent as a one-hot vector,
- $p_{i,c}$ is the predicted probability for pixel i and class c.

During the training process, the pixel-wise cross-entropy loss is minimized using optimization algorithms such as stochastic gradient descent (SGD) or Adam. Minimizing this loss encourages generating segmentation maps which will match the labels of the ground truth, thereby enhancing the segmentation performance overall.
During training, the model is optimized with stochastic gradient descent (SGD) or more sophisticated optimization techniques such as Adam. Data augmentation techniques, such as random scaling, cropping, and rotating, are frequently used to enhance generalization by augmenting the training dataset.
DeepLabv3+ takes an input image during inference and generates a dense semantic segmentation map. The model processes the image using the encoding network, the ASPP module, and the decoder network. The resulting feature maps are upsampled to the original image resolution, and a softmax operation is performed to determine the class probabilities for each pixel.

3.3.3 U-Net

General Introduction

U-Net is used widely in CNN architectural field where it helps to segment biomedical images. It was introduced in 2015 and is composed of a contracting path (encoder) and an expanding path (decoder) linked by skip connections. U-Net effectively captures local and global context information while preserving particulars. It is commonly used for medical image analysis tasks such as organ and tumor segmentation. The network's unique architecture and extensive data augmentation techniques contribute to its capacity to manage limited annotated medical datasets.

3.3.4 Architecture of the Model

Encoder Path

U-Net's encoder path comprises of multiple convolutional and pooling layers. The number of convolutional layers can vary based on the implementation and task specifications. In the original U-Net architecture, the spatial dimensions of the input image are typically reduced by four or five phases of downsampling.
Important encoder path layers consist of:

- Convolutional Layers: These layers extract features from the input image using convolutional operations. They are indispensable for capturing hierarchical information.
- Pooling Layers: The pooling layers, which are typically implemented as max pooling or average pooling will reduce the resolution of the spatial features by expanding their receptive field.

Decoder Path

U-Net's decoder path attempts to recover the spatial resolution lost during encoding and generate pixel-by-pixel predictions. It consists of upsampling layers and skip connections that aid in maintaining fine-grained details.
Important decoder path layers include the following:

- Upsampling Layers: The upsampling layers, which are typically implemented as transposed convolutions or bilinear interpolation, enhance the spatial resolution of the feature maps, enabling the model to reconstruct the original input size.
- Concatenation Layers (Skip Connections): These layers combine the upsampled feature maps from the decoder path with the encoder path's feature maps. Skip connections enable U-Net to maintain both low-level and high-level properties, enabling accurate localization and enabling the model to use both local and global context data.

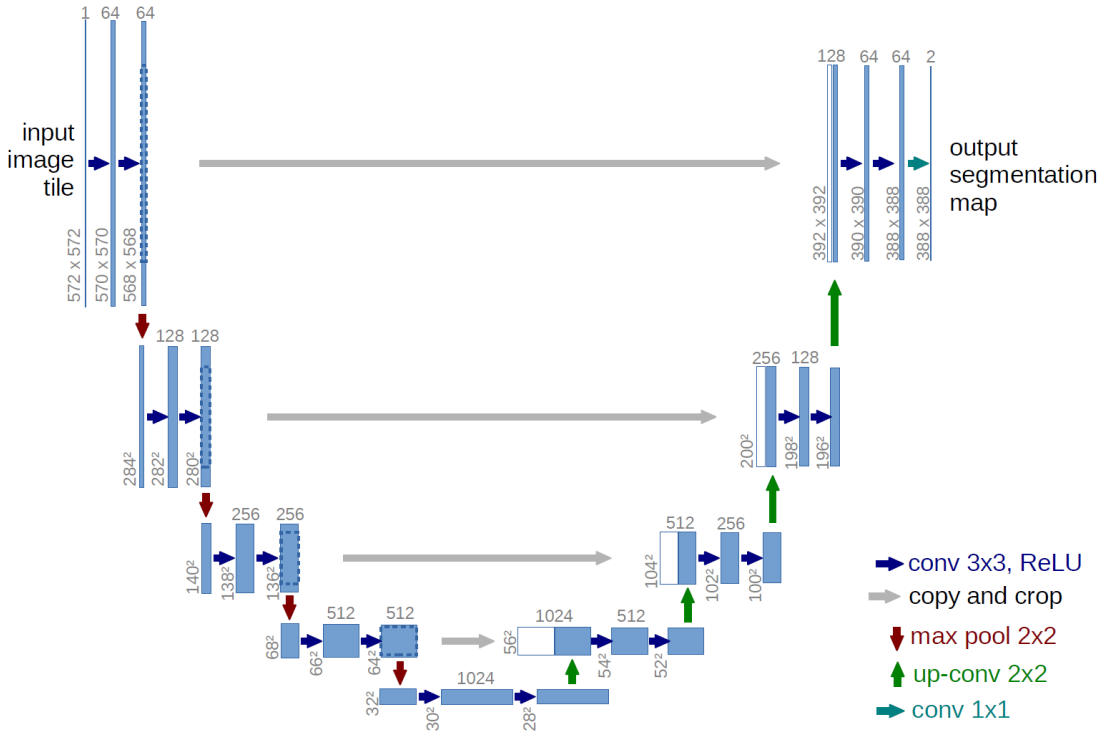


Figure 3.7: The U-Net Architecture

Training and Inference

U-Net is often trained using a pixel-by-pixel cross-entropy loss function that contrasts the predicted segmentation map with the actual segmentation map. To reduce loss and adjust model parameters during training, optimization techniques such as stochastic gradient descent (SGD) and Adam are utilized. Combining the cross-entropy loss function with pixel-wise softmax over the final feature map, the energy function is calculated. A definition of the softmax is

$$p_k(x) = \frac{\exp(a_k(x))}{\sum_{k'=1}^{K} \exp(a_{k'}(x))} \tag{3.2}$$

where $a_k(x)$ represents the activation of feature channel k at the corresponding pixel location $x \in \Omega$ with $\Omega \subset Z^2$. $p_k(x)$ is the approximation of the maximum-function, and K is the total number of classes. This translates to $p_k(x) \approx 1$ for the k that has the highest activation $a_k(x)$, and $p_k(x) \approx 0$ for every other k. The deviation of $p'(x)$ from 1 at each place is then penalized by the cross entropy using

$$E = \sum_{x \in \Omega} w(x) \log(p'(x)) \tag{3.3}$$

where $w(x)$ represents a weight assigned to pixel position x.

During inference, U-Net uses an input picture to produce a dense segmentation map. Following the decoder route's processing of the feature maps and combining them with the skip connections, the encoder path is used to extract features from the picture. The result is a segmentation map, where each pixel is tagged with the appropriate class in the end product.

3.3.5 ResUNet

General Introduction

ResUnet is a deep convolutional neural network (CNN) architecture that is based on the U-Net architecture [10]. ResUnet architecture has encoder and decoder network part which helps to connect the two bridges. The encoder network consists of a series of convolutional blocks, each of which is followed by a residual block. The ResUnet architecture represents the effectiveness of semantic segmentation. It has been used for a variety of applications, including road extraction, polyp segmentation, and brain tumor segmentation.
The residual block helps to address the problem of vanishing gradients, which can occur in deep CNNs. The decoder network is similar to the encoder network, but it is reversed. The bridge is a single convolutional block with a stride of 2. This helps to reduce the size of the output image while preserving spatial information.

3.3.6 Architecture of the Model

Encoder Path

- Convolutional Layers: These layers are responsible for extracting hierarchical features from the input image, similar to U-Net's encoder path.
- Pooling Layers: Pooling layers (e.g., max pooling or average pooling) are used to downsample the feature maps, increasing their receptive field while reducing spatial dimensions.
- Residual Units: Instead of using plain convolutional layers, ResUNet employs residual units in the encoder path. Two 3x3 convolutional blocks contained in each residual sector including an identity mapping. The residual units facilitate the training process and enable the network to go deeper without degradation issues.

Bridge

Transition between Encoder and Decoder: The bridge acts as a connection between the encoder and decoder paths, similar to U-Net.

Decoder Path

- Upsampling Layers: The decoder path utilizes upsampling layers, such as transposed convolutions or bilinear interpolation, to restore the spatial resolution of the feature maps.

- Concatenation Layers (Skip Connections): Similar to U-Net, ResUNet uses the connection to combine the upsampling of features map which extract from the decoder. This preserves both low-level and high-level information, enabling accurate localization and leveraging both local and global context data.
- 1x1 Convolution and Activation: After the last level of the decoding path, a 1x1 CNN layer with a suitable activation function (e.g., sigmoid) is applied to obtain the final pixel-wise segmentation output.

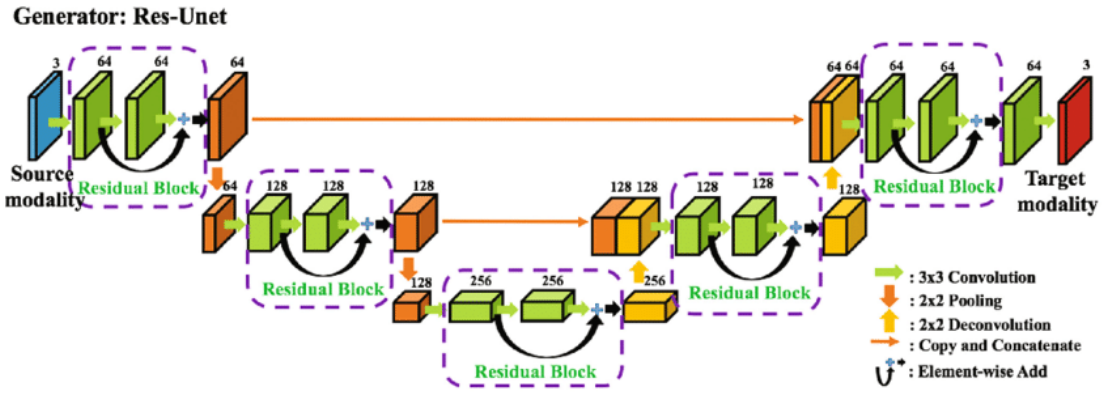


Figure 3.8: The ResUNet Architecture

Training and Inference

ResUNet can be trained using a pixel-by-pixel cross-entropy loss function, just like U-Net. However, ResUNet also includes residual connections, which can help to improve the training of the model. The residual connections allow the model to learn more complex features, which can lead to better performance on the segmentation task. The loss function is calculated using the softmax function to convert the output of the model to a probability distribution over the different classes. The cross-entropy loss is then used to measure the similarity between the predicted segmentation mask and the ground truth segmentation mask. The model is typically trained using the Adam optimizer.
The model extracts features from the input image using the encoder path. The feature maps are then processed by the decoder path which then helps the connections to skip are used to combine the features part of the encoder and decoder path. The output of the decoder path is a segmentation mask, which is used to identify the different objects in the image.

3.3.7 Proposed Hybrid Model

General Introduction

Combining the best of all three previous models, we have developed a novel hybrid model, mainly based on ResUnet. We have increased layers in the encoder path

and added a dilated spatial pyramid pooling layer before the residual units start.

Encoder Path

- Dilated Spatial Pyramid Pooling(DSPP): Unlike the Atrous Spatial Pyramid Pooling of Deeplabv3+, we have used DSPP in our encoder network. It has average pooling and convolutional layers. With this, the feature map is downsampled with average pooling before passing it to the residual layer. It provides a rich representation of the input image at multiple scales, enabling the network to have a broader context understanding. It helps in capturing both local and global context information, which is crucial for accurate semantic segmentation.
- Residual Layers: Similar to ResUNet, we have used residual blocks that take in the output of DSPP. This addresses the vanishing gradient problem of deep neural networks and makes the model learn better.
- Convolutional layers: These are used to extract features from the input, especially the hierarchical information.

Bridge

This is the transition between encoder and decoder networks.

Decoder Path

- Upsampling Layers: These increase the spatial dimensions (width and height) of an image or feature map. They are used to recover the original resolution of an image or match the resolution of a map of the features with input of the image size. We have used bilinear upsampling in this model.
- Skip Connections: These are used to allow information to bypass certain layers in a neural network, aiding in gradient flow and promoting better feature reuse during training.
- Final Convolution and Activation: Similar to ResUNet, we have used a 1x1 convolutional layer with sigmoid activation to get the filan segmentation output.

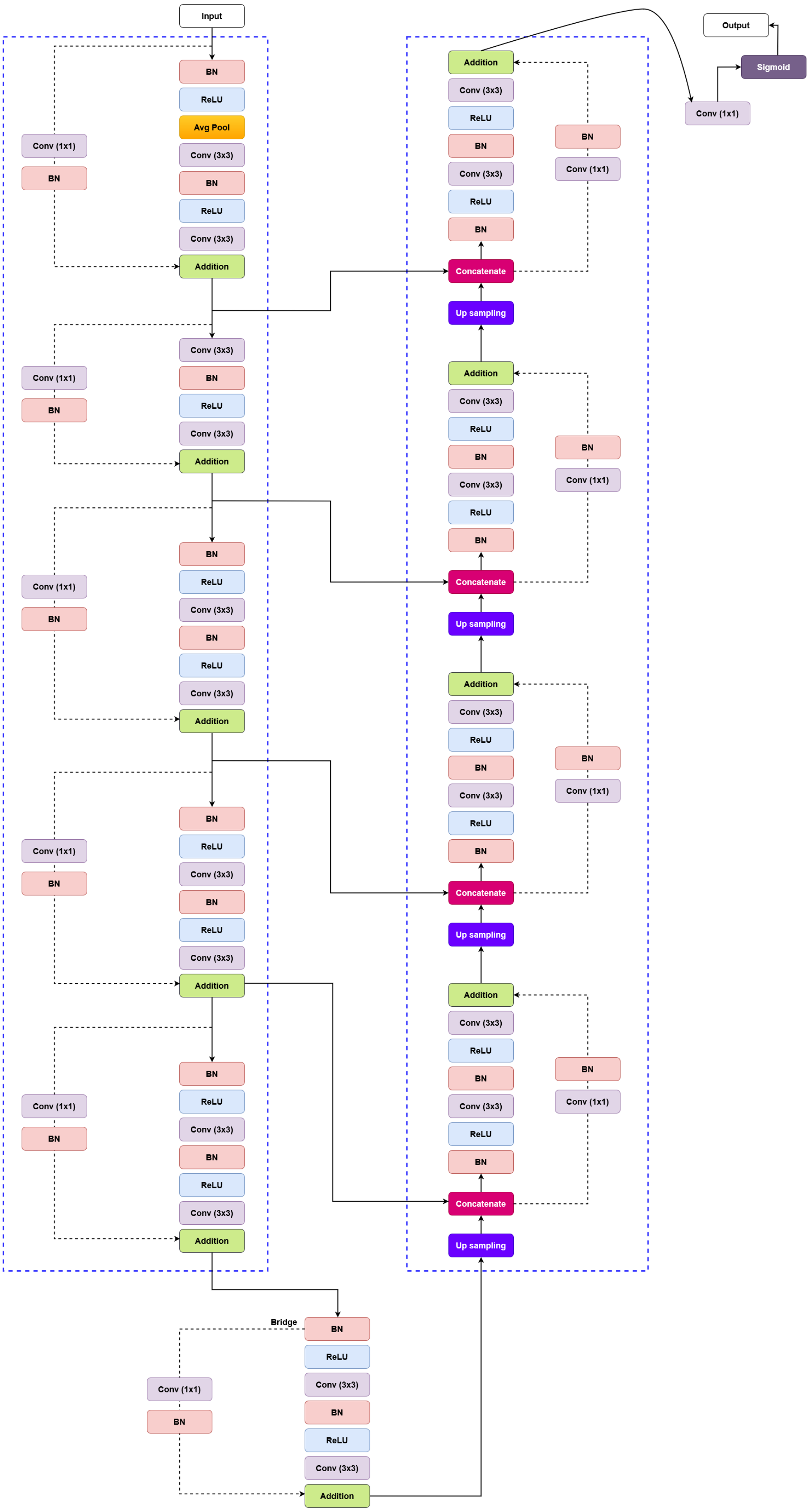


Figure 3.9: Architecture of our Proposed Model

Chapter 4

Result Analysis

After running DeepLabV3+, U-Net and Proposed 3D image segmentation models, we have determined the training and validation accuracy. Moreover, we have also discovered the tversky and dice coefficient accuracy with each of its loss values with a learning rate of 1e-6 unit. Python libraries like Pytorch, Keras and Tensorflow aid us in completing the training and validation testing result analysis.

4.1 Training and Testing on DeepLabV3+

In the training phase, the DeepLabV3+ processes with the iterative method which initialized trained weighted of the dataset and these training 3D MRI images are ready for segmentation. Moreover, loss functions like Tversky loss, dice coefficient loss, and training loss are acquired during the training of the model. After training, testing or validation will evaluate the performance of the trained model through 10% testing images.

4.1.1 Training and Testing Dice Similarity Coefficient

The result of the DeepLabV3+ training accuracy was accrued with about 88% after 10 epochs with 4 batches and 81.60% dice coefficient training accuracy.

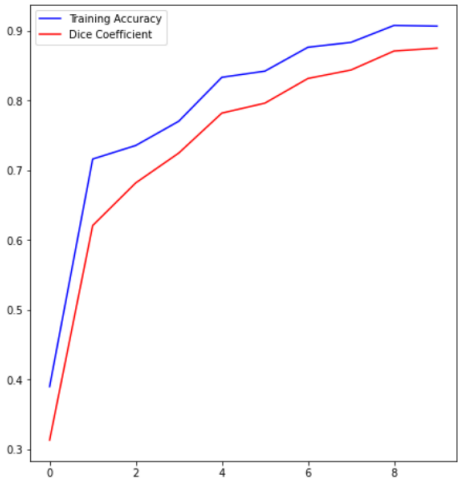


Figure 4.1: Dice Score and Training accuracy

After training, testing on DeepLabV3+ requires framework libraries like Pytorch and Keras which provide the efficient test result. About 85% testing results have been achieved through the dice coefficient process which shows the quality predicted segmentation overall performance.

4.2 Training and Testing on U-Net

U-Net dice coefficient training uses the U model CNN architecture which reduces the coefficient loss. This provides the prediction segmentation of the ground truth mask of the 3D images dataset.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot \nabla w_t \tag{4.1}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot \nabla (w_t)^2 \tag{4.2}$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{4.3}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{4.4}$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t + \varepsilon}} \cdot \hat{m}_t \tag{4.5}$$

This model uses a weights adjustment algorithm which is in our case we have used the Adam weight optimizer algorithm which helps to reduce loss in the training process. After training, validation testing has been performed with 10% of the data.

4.2.1 Training Dice Similarity Coefficient

In the U-Net model, the training accuracy was 87.40% and the dice coefficient accuracy is 84.40%.

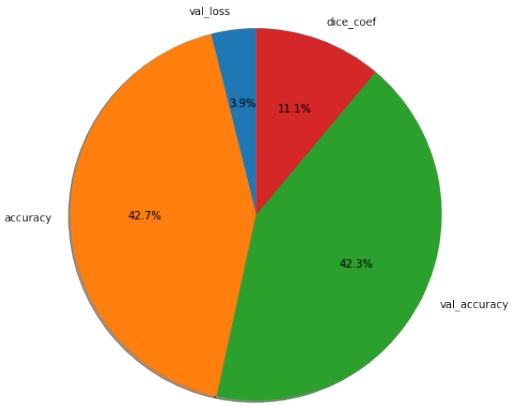


Figure 4.2: U-Net Result Piechart

The above pie chart represents the overall result of the U-Net model where the accuracy and the validation percentage show higher similarity. The loss validation

percentage is only 3.9% with 11% dice coefficient out of 100% respectively. With these results, it can measure the overlap of the two particular portions of the image area and give a prediction on the ground truth mask value.

$$DSC = \frac{|X \cap Y|}{|X| + |Y|} \tag{4.6}$$

Here, X and Y represent the area of the particular images and the (X union Y) represents the calculation of the correctly classified intersect area in the image. In this case, the DeepLabV3+ model gives better performance on classifying the segmented image with only 10 epochs.

4.2.2 Validation Testing

The validation result of the U-Net model is 87% which is lesser than the DeepLabV3+ model. This validation result proves the efficiency of the DeepLabV3+ model over the U-Net model.
In this result, the validation loss for the modified U-Net model is less than 2% and the validation accuracy for segmenting the 3D MRI brain images is greater than 95%. However, the validation result for the DeepLabV3+ is less than 90% where the images are less segmented compared to the modified U-Net model.

4.3 Training and Testing on Proposed Model

In this hybrid merged model, we have trained the same dataset and acquired a better result.

4.3.1 Validation Testing

We developed a cutting-edge proposed 3D segmentation model which gives the training accuracy result of 0.9859 for segmenting tumors on a 3D MRI dataset. Moreover, the validation accuracy result is 0.98 which shows the improved performance from apart individual DeepLabV3+ and ResUNet models.

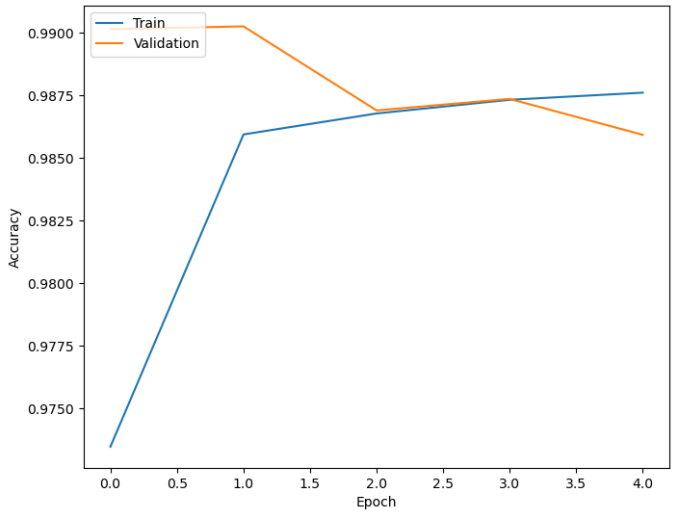


Figure 4.3: Merged 3D Segmentation Model Accuracy Result

By adding an extra layer of Pyramid Pooling from DeepLabV3+ into the ResUNet model, we set a new benchmark of better accuracy and validation result.

4.3.2 Dicecoefficient Result

The Dice coefficient result we found is 0.9691 which shows the outcome of the similarity of overlapping particular areas of the image. By synergizing ResUNet and DeepLabV3+ models, we also developed the validation for the Dice coefficient result which is 0.933 respectively.

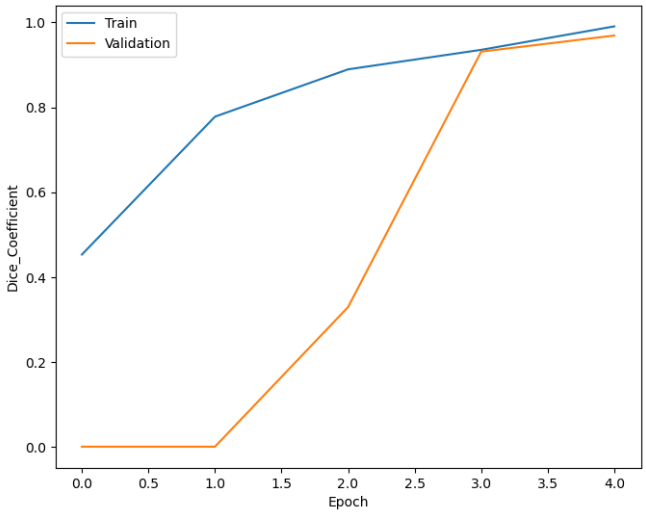


Figure 4.4: Merged 3D Segmentation Model Dice Coefficient Result

4.4 Comparison With Previous Models

After successful training and validation, our models could successfully segment the 3D Brain images as shown in the figures 4.5

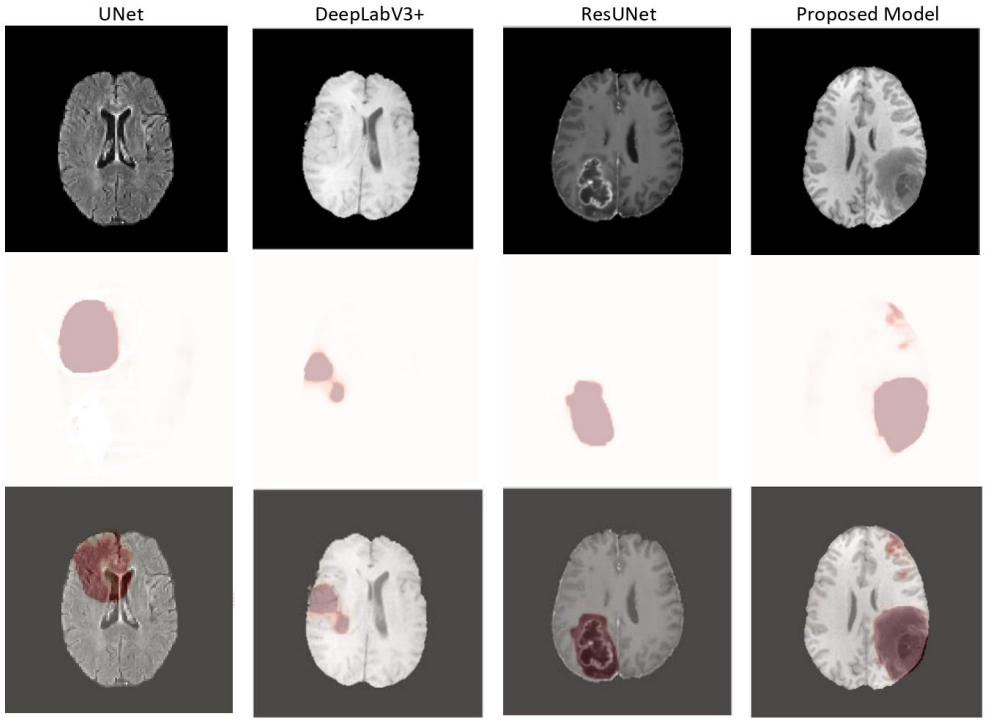


Figure 4.5: Segmentation Comparison(Original, Segmented and Finally Predicted Image

In the U-Net model image segmentation, the predicted brain tumor does not give accurate image results as input brain image. In the DeepLabV3+ model, the tumor prediction slightly matched the input image. In the ResUNet model, the tumor image has matched with the predicted image but there still has been a slight lack of overlapping seen in the segmented image area portion. In our proposed model, the tumor image has been identified and it shows improvement over the ResUNet segmentation image.

The following table 4.1 refers to the comparison between the models across different metrics. It demonstrates a Dice Coefficient of 0.844, Accuracy of 0.874 and IoU of 0.787 which indicates it will fail to perform of capturing highly detailed segmentation images.

DeepLabV3+ which is known as the pyramid pooling module shows slight improvement over the U-Net model. However, DeepLabV3+ achieves a Dice coefficient of 0.856, Accuracy of 0.884 and IoU of 0.806. ResUNet is a U-Net residual architecture and shows further improvement by achieving a Dice Coefficient of 0.862, Accuracy of 0.888 and IoU of 0.811. It performs better than U-Net and DeepLabV3+ 3D segmentation models.

Models	Dice Coefficient	Accuracy	IoU
UNet	0.844	0.874	0.787
DeepLabV3+	0.856	0.884	0.806
ResUNet	0.862	0.888	0.811
Proposed Model	**0.9691**	**0.9859**	**0.9642**

Table 4.1: Comparison among U-Net, DeepLabV3+, ResUNet and our proposed model

Our proposed modified 3D image segmentation model shows significant results. A high Dice Coefficient of 0.9691, Accuracy of 0.9859 and IoU of 0.9642 represent a new benchmark compared to U-Net, ResUNet and DeepLabV3+ 3D segmentation models. The successful integration of the pyramid pooling layer in the ResUNet model helps to leap forward in the field of accuracy percentage.
The bar graph below illustrates the comparison of four 3D segmentation models which were U-Net, DeepLabV3+, ResUNet and our proposed 3D image segmentation model. We have added three evaluation metrics which were Dice Coefficient, Accuracy and IoU will aid the segmentation performance. In the y-axis, there are four 3D segmentation models and the value of the results are in the x-axis of the bar graph.

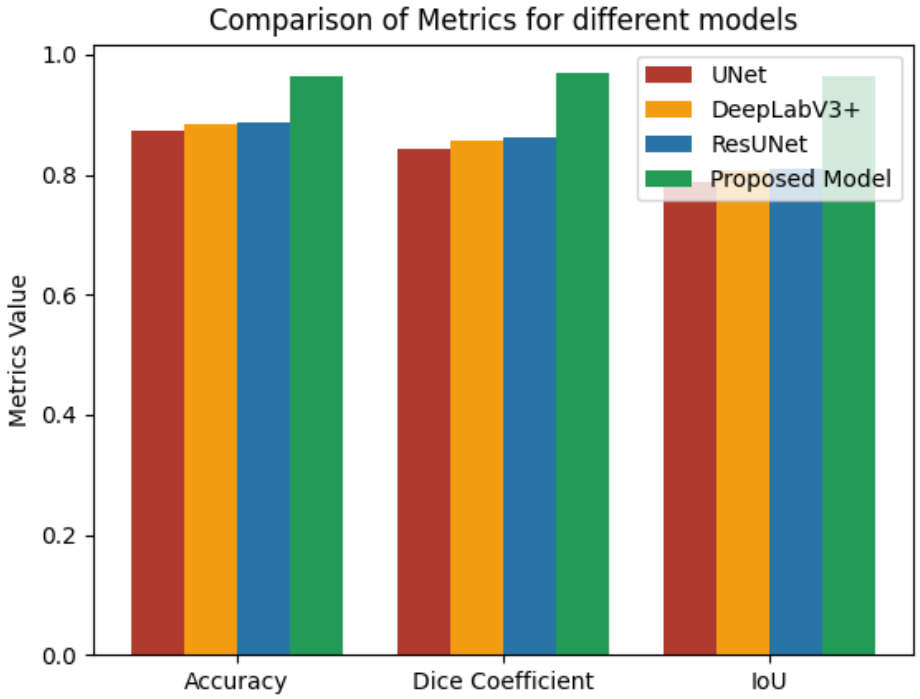


Figure 4.6: Comparison Graph

In every four 3D segmentation models, the accuracy is higher than the IoU and Dice scores. Starting with the U-Net model, the first red bar represents the Dice Score which is higher than the next IoU green bar. The next blue bar is the accuracy bar which is slightly less than the Accuracy bar of the DeepLabV3+ model. The Dice score bar of the ResUNet model pointed to a greater than 0.8 value which represents there is still a need for improvement on the image segmentation similarity area with Dice Coefficient. However, in our proposed modified 3D segmentation model, the Accuracy, Dice Scorce and IoU bar are slightly lesser than 1.0 and also greater than U-Net, ResU-Net and DeepLabV3+ models which represent the improved 3D image segmentation model.

Chapter 5

Conclusion

Medical 3D image segmentation is a crucial part that clarifies confusion between doctors and patients and more importantly, it makes revolutionary changes in the medical technology field. However, many medical sectors do not contain proper reliable new technology U-Net-based model which can segment the medical image. This segmentation is essential for tracking down diseases, fractures and similar internal problems. We used a combination of the DeepLabV3+ pyramid pooling layer within the ResUNet model to create a hybrid modified one that can perform better segmentation based on medical images. With a vast dataset, this hybrid model will function well, hence the accuracy percentage of segmenting images relies on the input dataset.

The use of our paradigm will be effective in a variety of contexts, despite the fact that it does have certain restrictions. When applied to a very limited dataset, this model will more than likely provide erroneous results. Additionally, the cost of this model is higher than the cost of other models since it needs additional processing to manage to skip connections. There is a chance that the bilinear approach will not be sufficient for the task at hand. In our future work, we would want to execute attention mechanisms and handline class imbalances. We hope to do this in the near future.

Bibliography

[1] X.-X. Yin, B. W.-H. Ng, Q. Yang, A. Pitman, K. Ramamohanarao, and D. Abbott, "Anatomical landmark localization in breast dynamic contrast-enhanced mr imaging," *Medical & biological engineering & computing*, vol. 50, pp. 91–101, 2012.

[2] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015. DOI: 10.1109/TMI.2014.2377694.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Springer, 2015, pp. 234–241.

[4] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, Sep. 2017. DOI: 10.1038/sdata.2017.117.

[5] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[6] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[7] L. Yi, L. Shao, M. Savva, H. Huang, Y. Zhou, Q. Wang, B. Graham, M. Engelcke, R. Klokov, V. Lempitsky, *et al.*, "Large-scale 3d shape reconstruction and segmentation from shapenet core55," *arXiv preprint arXiv:1710.06104*, 2017.

[8] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, M. Prastawa, E. Alberts, J. Lipková, J. B. Freymann, J. S. Kirby, M. Bilello, H. M. Fathallah-Shaykh, R. Wiest, J. Kirschke, B. Wiestler, R. R. Colen, A. Kotrotsou, P. LaMontagne, D. S. Marcus, M. Milchenko, A. Nazeri, M.-A. Weber, A. Mahajan, U. Baid, D. Kwon, M. Agarwal, M. Alam, A. Albiol, A. Albiol, A. Varghese, T. A. Tuan, T. Arbel, A. Avery, P. B., S. Banerjee, T. Batchelder, K. N. Batmanghelich, E. Battistella, M. Bendszus, E. Benson, J. Bernal, G. Biros, M. Cabezas, S. Chandra, Y.-J. Chang, and et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *CoRR*, vol. abs/1811.02629, 2018. arXiv: 1811.02629. [Online]. Available: http://arxiv.org/abs/1811.02629.

[9] W. Sun and R. Wang, "Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with dsm," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 3, pp. 474–478, 2018.

[10] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018. DOI: 10.1109/LGRS.2018.2802944.

[11] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual u-net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, pp. 014 006–014 006, 2019.

[12] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen, "Resunet++: An advanced architecture for medical image segmentation," in *2019 IEEE international symposium on multimedia (ISM)*, IEEE, 2019, pp. 225–2255.

[13] A. Komarichev, Z. Zhong, and J. Hua, "A-cnn: Annularly convolutional neural networks on point clouds," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 7421–7430.

[14] J. Zhang, X. Zhao, Z. Chen, and Z. Lu, "A review of deep learning-based semantic segmentation for point cloud," *IEEE Access*, vol. 7, pp. 179 118–179 133, 2019.

[15] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020.

[16] F. Renard, S. Guedria, N. D. Palma, and N. Vuillerme, "Variability and reproducibility in deep learning for medical image segmentation," *Scientific Reports*, vol. 10, no. 1, pp. 1–16, 2020.

[17] Y. Xie, J. Tian, and X. X. Zhu, "Linking points with labels in 3d: A review of point cloud semantic segmentation," *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 4, pp. 38–59, 2020.

[18] R. Feng, Z. Xu, X. Zheng, H. Hu, X. Jin, D. Z. Chen, K. Yao, and J. Wu, "Kernet: A novel deep learning approach for keratoconus and sub-clinical keratoconus detection based on raw data of the pentacam hr system," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3898–3910, 2021.

[19] B. Gao, Y. Pan, C. Li, S. Geng, and H. Zhao, "Are we hungry for 3d lidar data for semantic segmentation? a survey of datasets and methods," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[20] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability*, vol. 13, no. 3, p. 1224, 2021.

[21] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[22] S. Roychowdhury, "Few shot learning framework to reduce inter-observer variability in medical images," in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, 2021, pp. 4581–4588.

[23] S. Subudhi, R. N. Patro, P. K. Biswal, and F. Dell'Acqua, "A survey on superpixel segmentation as a preprocessing step in hyperspectral image analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5015–5035, 2021.

[24] Y. Wang, C. Wang, H. Wu, and P. Chen, "An improved deeplabv3+ semantic segmentation algorithm with multiple loss constraints," *Plos one*, vol. 17, no. 1, e0261582, 2022.

[25] X.-X. Yin, L. Sun, Y. Fu, R. Lu, and Y. Zhang, "U-net-based medical image segmentation," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[26] H. Liu, S. Lu, and F. Zhao, "Mlp-res-unet: Mlps and residual blocks-based u-shaped network intervertebral disc segmentation of multi-modal mr spine images.," *Current Medical Imaging*, 2023.

[27] S. Roychowdhury, "Numsnet: Nested-u multi-class segmentation network for 3d medical image stacks," *arXiv preprint arXiv:2304.02713*, 2023.

[28] Y. Xu, S. Hou, X. Wang, D. Li, and L. Lu, "A medical image segmentation method based on improved unet 3+ network," *Diagnostics*, vol. 13, no. 3, p. 576, 2023.