A Thesis on

Automated Glioma Segmentation on Post-Treatment MRI scans using Deep Neural Networks

By
Vikramjit Bora, 2623830

Under the supervision of

Dr Qingjie Meng

Assistant Professor,

School of Computer Science

For the degree of

M.Sc. in Artificial Intelligence and Machine Learning

Submitted to the

School of Computer Science

University of Birmingham

Abstract

This research addresses the complex challenge of segmenting gliomas in post-treatment MRI scans, a critical task in the ongoing management of brain tumor patients. Gliomas, particularly glioblastomas, are among the most aggressive brain tumors, making accurate diagnosis, treatment planning, and monitoring vital for improving patient outcomes. While MRI is the gold standard for glioma imaging, post-treatment scans present unique difficulties due to changes such as resection cavities and treatment-induced artifacts. The heterogeneous nature of gliomas, which infiltrate surrounding brain tissue and vary in appearance across MRI modalities, further complicates accurate glioma segmentation.

This research introduces a novel Residual 3D U-Net with Attention mechanism designed to improve the differentiation between residual tumor tissue and treatment-related changes. This model leverages deep learning techniques to enhance the accuracy and reliability of segmentation in complex post-treatment imaging conditions. A key innovation is the combination of residual connections and spatial attention mechanisms, which together improve the model's ability to capture and focus on relevant features, leading to superior segmentation performance.

The outcomes of this research not only contribute to advancing the field of medical image analysis but also hold significant implications for clinical practice. By providing more accurate and consistent segmentation results, the proposed model has the potential to enhance the quality of care for glioma patients, ultimately contributing to better treatment outcomes and improved survival rates.

Acknowledgement

I would like to express my sincere gratitude to my supervisor Dr. Qingjie Meng for her guidance and valuable insights during the project. I would also like to thank my project inspector, Dr. Miqing Li for his helpful feedback during the project demonstration. And last but not the least, I would like to thank my parents and family for their continued encouragement, support and patience. They are the main source of inspiration in my life.

I would also like to confirm that this work was solely undertaken by me and no help was provided from any other source than those permitted and referenced in this thesis.

Contents

List of Figures

Acronyms

2D	Two-dimensional
3D	Three-dimensional
Adam	Adaptive Moment Estimation
CNN	Convolutional neural network
CNS	Central Nervous System
CT	Computed tomography
DCNN	Deep Convolutional neural network
FCN	Fully convolutional network
GAN	Generative Adversarial Network
GBM	Glioblastoma Multiforme
GPU	Graphics processing unit
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
ReLU	Rectified Linear Unit
Tanh	Hyperbolic Tangent
VAE	Variational Auto Encoder

1 Introduction

1.1 Introduction

Gliomas are among the most prevalent and aggressive forms of brain tumors, originating from glial cells that support and protect neurons. These tumors vary greatly in their behaviour and response to treatment, with glioblastomas (GBMs) being the most aggressive subtype. Despite extensive research and numerous advancements in medical technology, the prognosis for glioblastoma patients remains grim, with a median survival time often measured in months rather than in years. This dire outlook underscores the critical need for accurate diagnosis, effective treatment planning, and careful post-treatment monitoring to improve patient outcomes.

Magnetic resonance imaging (MRI) is the gold standard (“Neuroradiology 101,” n.d.) for diagnosing and monitoring gliomas, which provides detailed images of the brain that reveal the size, location, and extent of the tumor. However, post-treatment MRI scans pose certain challenges for monitoring gliomas. After surgery, the brain undergoes various changes due to the resection of the tumor, radiation therapy, and chemotherapy. These changes can include the presence of resection cavities, treatment-induced inflammation, and other artifacts, which makes it difficult to distinguish between residual tumor tissue and treatment-related effects. The heterogeneous nature of gliomas, due to their diffuse infiltration into surrounding brain tissue, and variable appearance across different MRI modalities, further complicates this task (de Verdier et al., 2024).

Beginning in 2012, the Brain Tumor Segmentation (BraTS) Challenge has played a pivotal role in advancing the field of medical image analysis, particularly in segmentation of gliomas. The BraTS Challenge has provided a standardized platform for researchers to develop, test, and compare various algorithms for segmenting brain tumors in MRI scans. This has led to significant advancements in the accuracy and efficiency of glioma segmentation, with deep learning techniques, especially convolutional neural networks, emerging as powerful tools for this task. However, most prior challenges have focused on pre-treatment imaging, leaving a critical gap in the segmentation of gliomas in post-treatment MRI scans.

1.2 Motivation

The motivation for focusing on post-treatment glioma segmentation is rooted in both clinical needs and technological challenges. From a clinical perspective, accurate and reliable segmentation of post-treatment MRIs is critical for the ongoing management of glioma patients. The ability to distinguish between residual tumor tissue and post-surgical changes such as resection cavities or treatment-induced inflammation, such as post-radiation inflammation, is essential for determining the next steps in patient care. For instance, if a residual tumor is detected, additional treatment options such as further surgery, radiation therapy, or chemotherapy may be considered. On the other hand, a different strategy may be used if the detected changes are due to treatment therapy instead of tumour recurrence.

Manual segmentation of post-treatment MRIs is a time-consuming and labour-intensive process. Even among experienced radiologists, there can be significant variability in the interpretation of these images, leading to inconsistencies in treatment planning and patient outcomes. Automated segmentation algorithms have the potential to reduce this inconsistency and provide more consistent and accurate results. These models that can accurately identify residual tumor tissue and differentiate it from treatment-related changes can help clinicians make more informed decisions about patient management, leading to better treatment outcomes and improved quality of life for glioma patients (de Verdier et al., 2024).

The unique challenges presented by post-treatment MRIs also provide a compelling motivation for advancing the field of artificial intelligence (AI) and machine learning (ML). The variability in imaging, the subtlety of tumor boundaries post-treatment, and the need to integrate information from multiple MRI modalities all require innovative approaches. The post-treatment imaging represents a different scenario than pre-treatment imaging. As a result, most existing segmentation algorithms trained on pre-treatment images perform poorly on the post-treatment images.

1.3 Project Objectives

This research aims to address the above gap by concentrating on the segmentation of gliomas in the post-treatment setting using the 2024 BraTS dataset of post-treatment MRI scans. This is essential for assessing the effectiveness of surgical interventions, detecting residual or recurrent tumor tissue, and planning further treatment.

Specific objectives for the project include:

1. Enhancing the accuracy and reliability of segmentation algorithms tailored for post-treatment imaging conditions, which are often complicated by resection cavities and treatment-induced artifacts.

2. Developing robust deep neural networks to improve the segmentation between tumor recurrence and post-treatment effects. This includes designing an optimal model based on available hardware resources and evaluating the model on the available data.

3. Visualising predicted segmentation masks of the proposed model to qualitatively evaluate the model performance in addition to quantitative evaluation.

1.4 Contributions

This research work makes the following key contributions:

- Combines residual connections with spatial attention mechanism in a 3D U-Net architecture to perform glioma segmentation on post-treatment MRI scans. The proposed model improves segmentation accuracy by focusing on relevant features and learning deeper patterns.
- Handles class imbalance in the dataset (particularly for underrepresented tumor and resection cavity regions) by implementing a combination of focal loss and cross-entropy loss, leading to better model performance in difficult cases.
- Analyses and evaluates the performance of the proposed model in comparison to several U-Net based architectures, including 3D U-Net, 3D U-Net with Attention, Residual 3D U-Net.

1.5 Thesis Structure

This report is divided into seven chapters, each progressing from foundational concepts to the detailed methodology, results, and analysis. Chapter 1 introduces the research, outlining the motivation, objectives, and overall structure of the report. Chapter 2 lays the foundation by providing the essential medical and technical context. It covers medical and technical details about gliomas, the role of MRI, and the basics of CNNs and image segmentation. Chapter 3 delves into existing research on brain tumor segmentation, tracing the evolution from statistical methods to deep learning and highlighting the challenges of post-treatment imaging. Chapter 4 details the methodology, explaining the architecture of the proposed model along with implementation details. Chapter 5 presents the experimental setup, dataset, and evaluation metrics, followed by a thorough presentation of the results, both quantitative and qualitative. Chapter 6 critically analyses the findings and discusses its limitations and future research directions. Finally, Chapter 7 concludes the report by summarizing the key contributions and findings of the research.

2 Background

2.1 Medical background

2.1.1 Gliomas

“Glioma is a type of tumor that originates in the glial cells of the brain or spinal cord” (“What is a Glioma?,” 2024). Glial cells are an essential component of the central nervous system (CNS) that support and protect neurons by providing them with insulation, nutrients, and structural support. Gliomas are among the most common types of primary brain tumors and are classified based on their location within the brain, the type of glial cells they grow from, and the extent of their malignancy (“What is a Glioma?,” 2024).

Gliomas can be classified into the following categories based on the glial cells they originate from:

1. Astrocytoma
2. Glioblastoma (glioblastoma multiforme)
3. Ependymoma
4. Oligodendroglioma
5. Medulloblastoma etc.

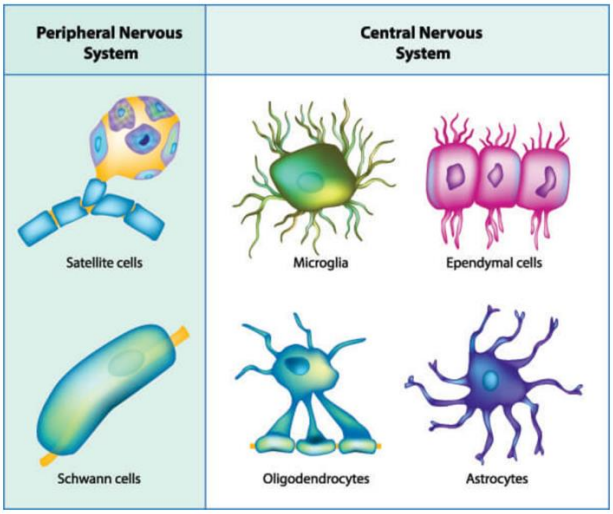

Figure 1: Types of glial cells (Knapp, 2020)

Gliomas can also be classified into four grades based on their aggressiveness using the World Health Organization (WHO) classification system (Louis et al., 2016):

1. Grade I (Pilocytic Astrocytoma): These are generally benign and slow-growing tumors. They are usually localized and have a good prognosis after surgical removal.
2. Grade II (Diffuse Astrocytoma): These are also low-grade tumors, but they are infiltrative in nature. As a result, they spread to surrounding healthy brain tissue making it difficult to remove them completely through surgery. These tumors tend to progress to higher-grade tumors eventually.
3. Grade III (Anaplastic Astrocytoma): These are malignant tumors that grow more quickly and have a worse prognosis than Grade II gliomas. They are more likely to recur after treatment and can progress to Grade IV.
4. Grade IV (Glioblastoma Multiforme or GBM): This is the most aggressive and common form of glioma, characterized by rapid growth and a poor prognosis. GBMs are highly invasive and tend to spread within the brain.

The aggressiveness of gliomas varies significantly, and their “treatment typically involves a combination of surgery, radiation therapy, and chemotherapy”(Isachesku et al., 2023). Despite advances in medical science, high-grade gliomas, particularly GBM, remain challenging to treat and are associated with a poor prognosis. As

previously mentioned, this thesis will deal with segmentation of post treatment MRIs of GBMs.

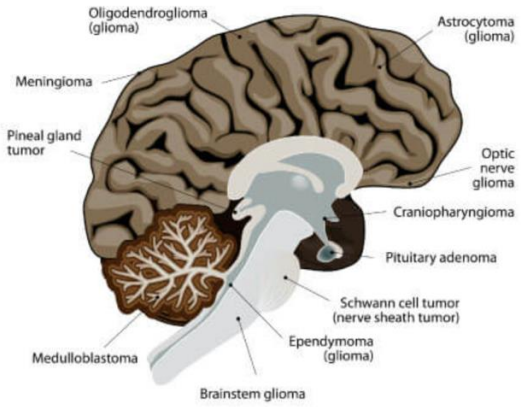

Figure 2: Common primary brain tumors (Hospitals, 2018)

2.1.2 Importance of MRI in Glioma Treatment Management

MRI (Magnetic Resonance Imaging) is a critical imaging tool in the management of gliomas, providing superior soft-tissue contrast compared to other modalities. It's used throughout the entire process, from initial diagnosis and surgical planning to treatment monitoring and post-treatment assessment. Its ability to differentiate between tumor recurrence and post-treatment effects makes it particularly valuable.

MRI plays several essential roles in the management of gliomas:

1. Diagnosis and Pre-treatment Evaluation: MRI is used to precisely locate the tumor, determine its size, and assess the surrounding healthy brain tissues.
2. Surgical Planning and Guidance: An MRI is essential for planning the surgical resection of gliomas. Surgeons rely on MRI scans to differentiate between tumor tissue and normal brain tissue, which is crucial for achieving maximum safe removal of the tumor while preserving healthy tissue and important neurological functions.
3. Post-treatment Monitoring: After surgery, MRI is used to assess the extent of tumor resection and to detect any residual tumor tissue. This is used to distinguish between residual tumors and post-surgical changes, such as blood

or edema. Patients with gliomas require regular follow-up MRI scans to monitor for tumor recurrence or progression. MRI can detect changes in the tumor's size, shape, and enhancement patterns, which may indicate recurrence.

2.1.3 Tumor Tissues in Gliomas

MRI is also used to identify and assess different types of tumor tissues for both pre-operative and post-operative treatment management. These tissues include (de Verdier et al., 2024):

- **Enhancing Tissue** (ET): This refers to areas of the tumor that appear enhanced on contrast-enhanced MRI scans, typically indicating regions with a disrupted blood-brain barrier and active tumor growth. Enhancing tissues are often targeted for surgical resection and are closely monitored for recurrence.
- **Non-Enhancing Tumor Core** (NETC): These are parts of the tumor that do not enhance with contrast on MRI. Despite not showing enhancement, these regions can still contain viable tumor cells and are significant in assessing the total tumor burden, especially in high-grade gliomas.
- **Surrounding Non-Enhancing FLAIR Hyperintensity** (SNFH): This refers to areas surrounding the tumor that appear hyperintense on FLAIR MRI sequences. These regions are usually made up of edema, infiltrative tumor cells, or a mix of the two. They are useful for figuring out how far the tumor has spread beyond the visible edges.
- **Resection Cavity** (RC): Post-surgical MRI scans will show the resection cavity, which is the space left after tumor removal. Evaluating this area is crucial for detecting residual tumor tissue and planning further treatment.

“The whole tumor (ET plus SNFH plus NETC) defines the whole extent of the tumor, including the tumor core, infiltrating tumor, peritumoral edema and treatment-related changes” (de Verdier et al., 2024).

Different MRI sequences are used to provide information about these tumor tissues, aiding in diagnosis, treatment planning, and post-treatment evaluation (Murphy and Gaillard, 2015):

- **T1-Weighted Imaging (T1)**

  T1-weighted imaging is a fundamental MRI sequence that provides high-resolution images of the brain's structure. In these scans, tissues rich in fat, such as white matter, appear bright, while water-containing tissues like edema, cysts, and cerebrospinal fluid (CSF) appear dark. T1-weighted images are particularly effective for assessing brain structure and determining the tumor’s location and size. However, T1-weighted imaging alone does not provide sufficient contrast between tumor and healthy brain tissue, making it less effective for fully identifying gliomas. Consequently, it is often combined with other sequences, such as T1-weighted imaging with gadolinium contrast (T1-Gd), to enhance tumor detection.

- **T1-Weighted Imaging with Gadolinium Contrast (T1-Gd)**

  Gadolinium-based contrast agents enhance the visibility of specific tissues on T1-weighted images by shortening the T1 relaxation time, making these tissues appear brighter (D. Shen et al., 2017). In glioma cases, T1-Gd imaging is particularly valuable for identifying the enhancing tumor (ET) regions, where the blood-brain barrier is compromised. This sequence is crucial for delineating the extent of the enhancing tumor, evaluating residual tumor tissue after surgery, and monitoring tumor changes over time. Moreover, the enhancing areas on T1-Gd images often correspond to zones of active tumor growth, making this sequence essential for treatment planning and follow-up. Additionally, it helps differentiate between tumor recurrence and treatment-related changes, such as radiation necrosis, which may appear similar on other MRI sequences.

- **T2-Weighted Imaging (T2)**

  T2-weighted imaging is another core MRI sequence, highly sensitive to water content. In these images, fluid-filled structures like CSF, edema, and cysts appear bright, while tissues with less water content, like white matter, appear darker. T2-weighted imaging is particularly useful for visualizing the extent of tumor infiltration into surrounding brain tissue. Additionally, T2-weighted imaging is valuable for identifying non-enhancing tumor regions.

- **Fluid-Attenuated Inversion Recovery (FLAIR)**
  FLAIR is a specialized MRI sequence that is highly sensitive to changes in water content in the brain, like T2-weighted imaging. However, FLAIR suppresses the signal from free water, like CSF, making it easier to detect abnormalities in brain parenchyma. This suppression improves the contrast between lesions and surrounding brain tissue, especially in areas where T2-weighted imaging may be less effective. In gliomas, FLAIR imaging is particularly useful for identifying non-enhancing tumor regions, peritumoral edema, and infiltrative tumor components. It is frequently used alongside T2-weighted imaging to provide a more comprehensive assessment of tumor infiltration and its effects on surrounding brain structures (Bakshi et al., 2001).

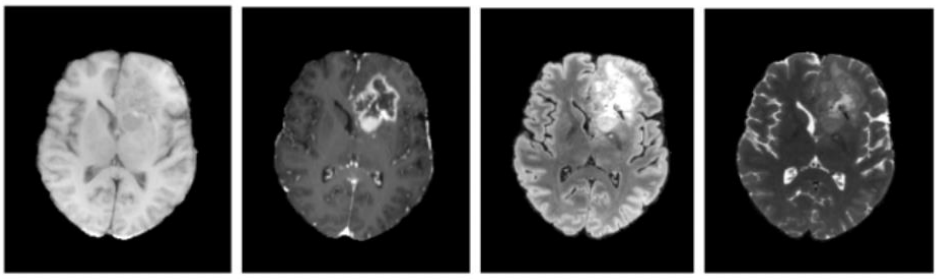

Figure 3: Comparison of T1-weighted, T1-Gd, FLAIR and T2-weighted MRI scans

These MRI sequences collectively provide a detailed picture of the tumor's characteristics and its interaction with surrounding brain tissues, guiding both surgical and non-surgical treatment strategies. They are crucial for ongoing monitoring to detect changes in the tumor post-treatment, such as recurrence or treatment-related effects.

2.2 Technical Background

2.2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are deep learning models specifically designed for working with images. They have gained significant popularity in the field of computer vision by enabling machines to perform image classification, object detection, as well as image segmentation. CNNs, particularly deep convolutional neural networks, have gained popularity in computer vision tasks due to their ability to learn the spatial hierarchies of features in images. These networks can automatically learn image features from low-level edges to high-level object segments, directly from raw images.

2.2.1.1 Architecture of CNNs

A CNN typically consists of several types of layers, each serving a specific purpose in the network's architecture:

1. **Input layer**: This layer consists of the pixel values of the image. For example, an image of size 32x32 with 3 colour channels (RGB) would have an input layer of dimensions 32x32x3.
2. **Convolutional layers**: These layers are at the core of CNNs. They apply filters or kernels with learnable parameters to the input data or to the output from the previous layer. Each filter slides over the input, performing a convolution operation. This is in essence computing the dot product or a weighted sum of the pixel values between the filter and the input patch. Output of this operation is a feature map (Hassaballah and Awad, 2020). Multiple filters can be applied within the same layer to produce multiple feature maps. These feature maps are basically a spatial representation of the input's features such as edges, corners, texture etc.

   Mathematically, the convolution operation can be defined as:

$$O(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i+m, j+n) \cdot K(m,n)$$

Where:

- $O(i,j)$ is the output feature map at position (i,j).
- $I(i+m,j+n)$ is the input image value at position $(i+m,j+n)$.
- $K(m,n)$ is the kernel value at position (m,n).
- $M \times N$ is the size of the kernel (e.g. 3x3)

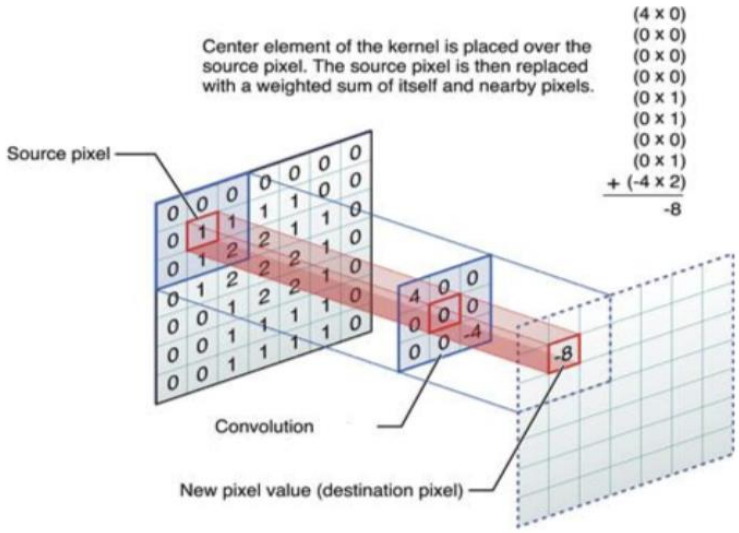

Figure 4: Convolution on 7x7 matrix by 3x3 kernel (Basavarajaiah, 2022)

3. **Activation functions**: An activation function is used to introduce non-linearity into the model. This non-linearity is important as it enables the model to learn complex patterns in the input data. Common activation functions used in CNNs are Rectified Linear Units (ReLU), Sigmoid and Hyperbolic Tangent (Tanh). (More about activation functions in subsection 2.2.1.2 below)
4. **Pooling layers**: Pooling layers are used to reduce the spatial dimensions (width and height) of the feature maps. This helps in decreasing the complexity and computational load of the network. It also helps in achieving spatial invariance, meaning that the exact location of features within the input image is less important. The most common type of pooling is Max-Pooling, where the maximum value within a kernel sliding over the feature map is selected.

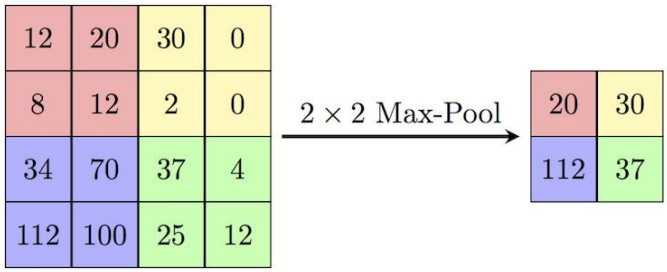

Figure 5: Max-Pooling with a 2x2 filter and stride 2 (“Papers with Code - Max Pooling Explained,” n.d.)

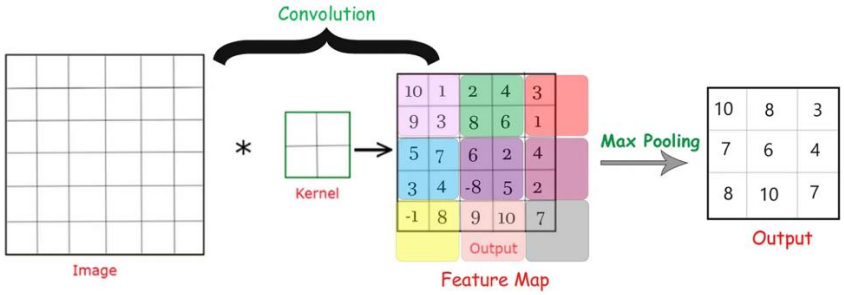

Figure 6: Convolution and Pooling layer applied to an input image (Dixit, 2021)

5. **Fully connected layers**: Similar to multilayer perceptrons (MLPs), these layers connect each neuron in the previous layer to every neuron in the subsequent layer of the network. They process the features learned by the numerous previous convolution layers and use them to make final predictions, such as classifying an image into categories.
6. **Output layer**: This layer produces the final output of the network, often using an activation function such as a softmax function for classification tasks to generate probabilities for each class.

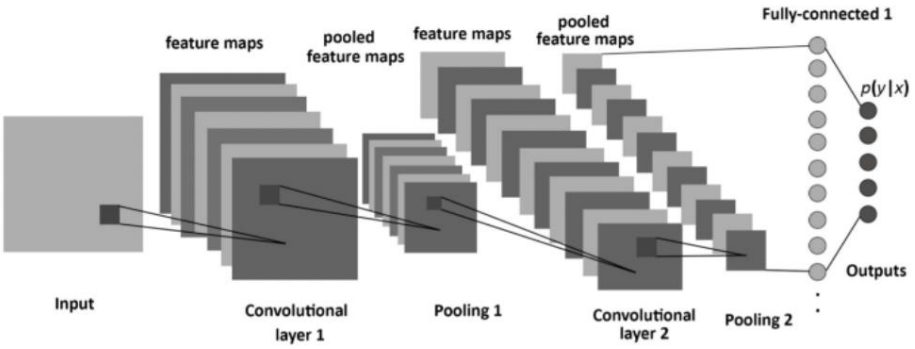

Figure 7: CNN Architecture consisting of convolutional layer, pooling layer and fully connected layer. (Albelwi and Mahmood, 2017)

2.2.1.2 Activation Functions

Activation functions play a critical role in all deep learning networks, including CNNs, by introducing non-linearity into the model. This non-linearity is responsible for the model being able to learn complex patterns in the data. Without the use of activation functions, a neural network model will behave like a linear model limiting its ability to learn complex and intricate patterns in the data. Here we discuss some of the most used activation functions in CNNs:

1. Sigmoid: The sigmoid function maps the input values to a range of 0 to 1, which is useful for binary classification tasks. As we can see from the Figure 8, it has a smooth gradient making it a popular choice for gradient based optimisation tasks. However, a limitation of the sigmoid function is the vanishing gradient problem, where gradients become close to zero at extreme values. Sigmoid activation function is defined as:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

2. Hyperbolic Tangent (Tanh): The tanh function is like the sigmoid function but restricts the input values between -1 and 1. Tanh is often preferred over sigmoid because the output of the tanh function is centred around zero. This leads to faster convergence during training compared to the sigmoid function. However,

like sigmoid, it also suffers from the vanishing gradient problem. Mathematically, it is defined as:

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

3. Rectified Linear Unit (ReLU): The ReLU function is defined as $\mathrm{ReLU}(x) = \max(0, x)$. It is one of the most widely used activation function in CNNs. Unlike the sigmoid and tanh, it addresses the vanishing gradient problem by setting all negative inputs to zero while not changing any positive input values. It is also very efficient as it requires thresholding only at zero. However, a potential drawback of ReLU is the 'dying ReLU' problem, which refers to the possibility of the network stopping to learn if the output of the neurons is consistently negative, resulting in a zero ReLU output.

   A variant of ReLU known as Leaky ReLU tries to mitigate this problem by returning a small, non-zero output for negative values instead of setting negative values to zero. This prevents neurons from dying during training.

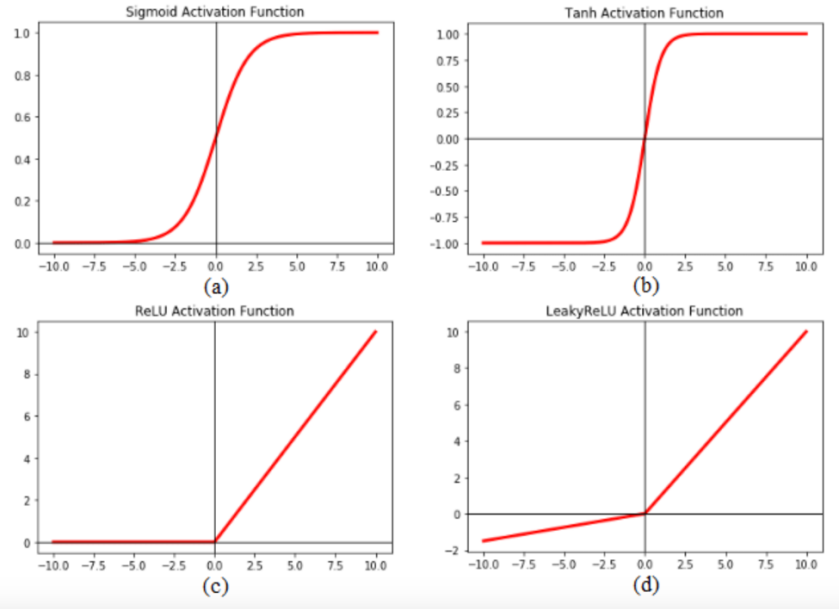

Figure 8: Plots of different activation functions: (a) Sigmoid (b) Tanh (c) ReLU (d) LeakyReLU (Kandel and Castelli, 2020)

4. Softmax: Softmax is typically used in the output layer of a CNN when performing multi-class classification. It converts the raw output logits from the

previous layer into probabilities for each class or label with one output neuron for each class.

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

2.2.1.3 Advantages of CNNs

CNNs are the preferred deep learning architecture to work with images due to the below advantages:

1. Spatial Hierarchy: By focusing on local regions of the input, CNNs can efficiently capture spatial hierarchies in the data. This allows them to learn and generalise features in the input image from local to global levels efficiently.
2. Parameter Sharing: Unlike MLPs, CNNs use shared weights across the receptive field. This drastically reduces the number of parameters that the model must learn making them more efficient and less prone to overfitting.
3. Translation Invariance: The convolutional and pooling operations provide translation invariance, meaning the network can recognize objects or patterns even if they appear in different locations in the image.

CNNs have transformed the field of computer vision, allowing machines to process visual data with significant accuracy. Convolution layers, pooling layers, and fully connected layers, along with various activation functions, allow CNNs to learn complex patterns in images.

2.2.2 Image Segmentation

Image segmentation has become increasingly important across various domains, particularly in medical imaging for analysing anatomical structures or identifying pathologies (Elnakib et al., 2011). Despite its complexity due to issues like noise and artifacts, image segmentation is an important preliminary step in image processing tasks (Namburu and Barpanda, 2020). It involves dividing an image into distinct segments, where each segment shares certain characteristics. These segments are used to identify objects and boundaries within the image.

Segmentation methods are generally divided into following types:

Semantic segmentation is a process where an image is divided into distinct segments, with each pixel in the image being classified into a specific label which represent the label of object it represents (Guo et al., 2020). In other words, it is a pixel level classification task. For example, in an image containing people, everyone in the image is considered to belong to the same class and the background is another class.

Instance segmentation involves identifying and segmenting each distinct object within an image as a separate entity. Unlike semantic segmentation, which classifies all objects of the same type (e.g., all people) as a single category, instance segmentation not only classifies the objects but also distinguishes between different instances of the same object category (Yi et al., 2019). For example, segmenting each person as a unique and separate object.

These concepts are illustrated in Figure 9

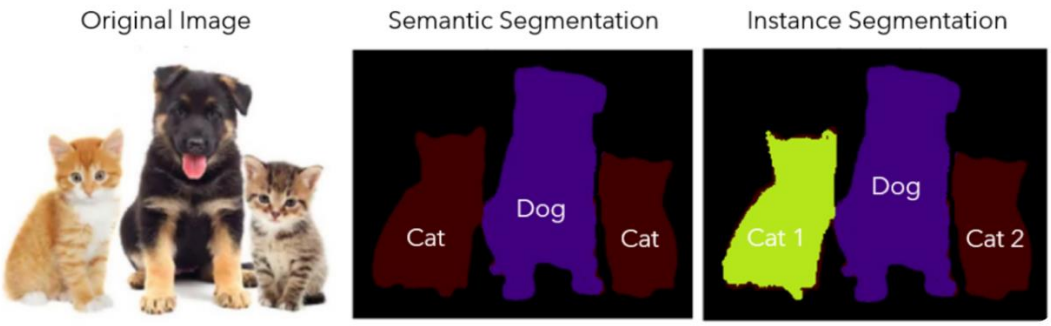

Figure 9: Semantic and Instance segmentation (“What Is Instance Segmentation?,” n.d.)

2.2.3 Medical Image Segmentation

In the context of medical imaging, there are multiple segmentation algorithms that are used. Below are some commonly used techniques for brain segmentation (Despotović et al., 2015):

1. Manual Segmentation: This is a widely used method, especially for creating ground truth data for training or assessing other segmentation techniques. It involves radiologists or clinicians manually segmenting and labelling 3D volumes of medical images slice by slice. However, it is time-consuming, labour-intensive and prone to inter-observer variability among radiologists and physicians.

2. Thresholding: This is a straightforward method based on the pixel intensity of an image. It is quick and computationally efficient but sensitive to noise since it doesn't account for spatial context. Thresholding involves setting a threshold value to differentiate between different regions. Variants of thresholding include single, local, multi, and adaptive thresholding.
3. Edge-Based Segmentation: This approach is based on detecting the edges of image objects, as edges provide significant information. It includes:
    - Search-Based Edge Detection: Computes the strength of edges by measuring local gradients.
    - Zero-Crossing Based Edge Detection: Identifies edges by computing zero crossings in a derivative expression of the image such as second order derivatives.
4. Region-Based Segmentation: Common region-based segmentation methods are:
    - Region Growing: This method starts with an initial region defined by a user and the algorithm expands this region by adding neighbouring pixels that meet pre-defined criteria for homogeneity. This process continues until the region stops growing.
    - Region Splitting and Merging: This involves dividing the image into regions with similar characteristics and then merging those regions that meet predefined criteria. This method is also known as split-merge segmentation.
5. Watershed Segmentation: This method treats the image like a topographic surface where pixel intensity is viewed as height. Image segmentation is done by finding the lines that run along the top of the "ridges" in this topography.
6. Clustering-Based Segmentation: In this method, pixels are grouped into clusters based on their features such as colour, texture etc. Common clustering algorithms used in segmentation include K-means, Fuzzy C-Means etc.
    - K-means Clustering: The image is divided into K clusters, and each pixel is assigned to the cluster whose centroid is closest to it. After all pixels are assigned, the centroids are recalculated, and the process is repeated until convergence, or a pre-defined number of iterations occur.

- Fuzzy C-Means: In this approach, unlike K-means, each pixel can belong to more than one cluster. The method involves an optimization function that influences the final clustering outcome.

7. Supervised Methods: These methods are based on supervised learning, in which image features, such as intensity values, edges, or textures, are used for classification. These features typically form multi-dimensional datasets. A key requirement for supervised methods is the availability of labelled data which serve as training data for the algorithm. This need for labelled data is a drawback compared to clustering techniques.
8. Deformable Models: These models perform segmentation by utilizing parametric curves (for 2D images) or surfaces (for 3D images) that deform under internal or external forces (Özdemir and Reski, 2021). Internal forces maintain the regularity of the curve or surface, while external forces guide the contour to fit the image features. To ensure that the model accurately delineates the region boundaries, an initial rough manual segmentation is performed.
9. Atlas-Based Methods: Like classification techniques, atlas-based methods are applied in the spatial domain. This approach is frequently used for brain segmentation and relies on a template or atlas, which is a composite of multiple manually segmented images. The template must be aligned with the target image through a process called image registration. Once aligned, the image is labelled accordingly.
10. Deep Learning based methods: Recent advancements in computational power, such as the development of more powerful graphics processing units (GPUs), have significantly accelerated progress in deep learning. Nowadays, deep learning is widely used to solve various computer vision problems including those in medical image segmentation and has achieved remarkable success in it. This has been largely due to the availability of large datasets which has been trained on CNNs using powerful GPUs. A primary task in medical imaging is semantic segmentation, which classical CNNs are not well-suited for because they lose spatial information when the convolutional features are passed into fully connected layers (Ronneberger et al., 2015). To address this, adaptations of CNNs have been developed, such as the Fully Convolutional Network (FCN) and U-Net based models, which are covered in more detail in section 3.3 below.

3 Literature Review

3.1 Statistical Machine Learning

Most brain tumor segmentation algorithms between the late 1990s and the early 2000s were based on traditional machine learning algorithms. These methods generally relied on hand-crafted features and domain knowledge. A clustering algorithm along with a rule-based expert system and multi spectral histogram analysis was used by (Clark et al., 1998) to segment tumor regions. (Prastawa, 2004) used the concept of outlier detection to segment brain tumors. Brain tumors are considered outliers in the MR image that are significantly different from healthy brain tissue and the algorithm performs segmentation by effectively identifying these outliers. (Wels et al., 2008) used a combination of discriminative machine learning models with graph cut algorithms to segment tumor regions. The discriminative model is used to identify tumor tissue using supervised machine learning techniques. Graph cuts are then used to model the segmentation task as an optimisation problem. This improves the segmentation results by taking into account spatial relationships and multi-spectral voxel intensities. (Menze et al., 2010) used a generative probabilistic model for brain tumor segmentation. It adds a probabilistic atlas of healthy tissue priors to a latent atlas of brain tumor tissues which is then used to detect tumor boundaries.

Although these early works were pioneering, they have certain shortcomings. Firstly, most of these are developed under strict assumptions. Secondly, these algorithms are dependent on manually designed feature engineering which is dependent on domain knowledge. Due to this, these approaches have poor generalisation. Finally, certain challenges such as inherent data imbalance in medical imaging data are not generally addressed by these early research works.

3.2 Deep Learning based approaches

(Krizhevsky et al., 2012) revolutionized the field of computer vision by demonstrating the exceptional power of deep convolutional neural networks. This sparked

widespread interest among researchers to apply Deep CNNs (DCNN) to various computer vision problems including medical image analysis, leading to its rapid adoption and various breakthroughs in this field.

(Havaei et al., 2017; Pereira et al., 2016; Zikic et al., 2014) were some of the early works that gradually increased the network depth by stacking multiple convolutional layers on top of each other to perform accurate brain tumor segmentation.

Early research work involved experimentation with DCNNs by stacking convolutional blocks with different kernel sizes along with non-linear activation functions and pooling layers in the network. Works such as (Dvořák and Menze, 2016), (Rao et al., 2015), (Zhao et al., 2018) built DCNNs using convolutional blocks with a larger kernel size greater than 5. Meanwhile, research works such as (Pereira et al., 2016), (Zikic et al., 2014) used smaller sized kernel (normally 3) as the basic convolution block for DCNNs. Still other works such as (Havaei et al., 2017) used different sized kernels in the network where larger sized kernels were used to detect global features with its larger receptive field while smaller sized kernels were used to detect more localized features with its smaller receptive field (Liu et al., 2023).

Since using two 3x3 kernels creates the same receptive field as a single 5x5 kernel, but with fewer parameters to train, later work started to focus on building deep networks by using 3x3 kernels as basic building blocks for DCNNs. This idea was extended to 3D convolutions in order to use the volumetric information of MRI scans by using 3x3x3 kernels as done in (Casamitjana et al., 2017), (Jesson and Arbel, 2018).

3.3 Encoder Decoder architectures

In the DCNNs discussed above used for brain tumor segmentation (semantic segmentation), the input to the model is an image patch, and the output is the predicted label of the centre pixel of that patch which has the highest probability of belonging to that class (Fernando and Tsokos, 2023). It is often quite difficult to accurately classify the correct label using the patch level information. This is because in these networks, the last fully connected layer is responsible for mapping the features to the correct label. Such an architecture with a fully connected layer at the end is not able to fully

capture all the feature information while a more complicated fully connected layer will increase the number of parameters drastically and hence will be computationally very expensive.

(Long et al., 2015) tackled this problem by introducing **fully convolutional neural network** (FCN) which is tailored for the task of semantic segmentation. It facilitates an end-to-end training of deep neural networks for pixel level prediction tasks enabling an effective and efficient way of performing semantic segmentation of the input image. As we can see from Figure 10, it achieves this by replacing the fully connected layers in DCNN by deconvolution layers. These deconvolution layers are used for upsampling the output feature maps to match the input image resolution thus allowing for pixel wise predictions.

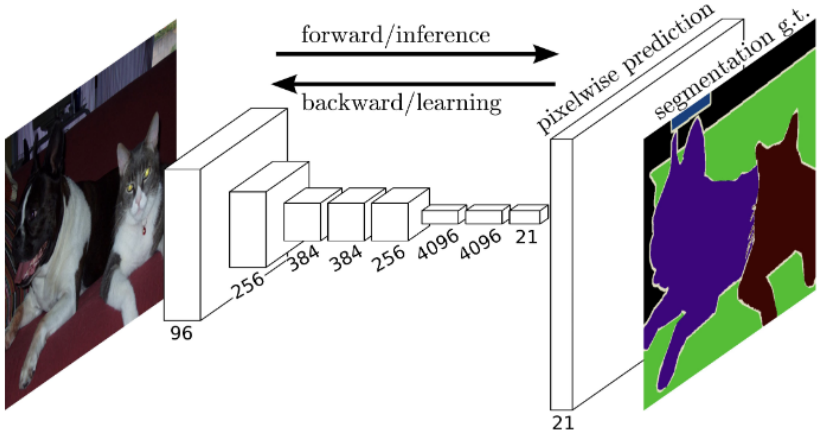

Figure 10: FCN architecture with deconvolution layers (Long et al., 2015)

(H. Shen et al., 2017) proposed a modified version of FCN with two upsampling branches. One of the branches is used to identify the tumor boundary as a classification decision. The other branch then detects the subregions within the tumor region to segment different tumor regions. The output of both these branches are concatenated and fed to the final softmax layer to output the final segmentation results.

One of the significant enhancements to the FCN is the **U-Net architecture** proposed by (Ronneberger et al., 2015). The U-Net architecture consists of a symmetric encoder (contracting) and decoder (expanding) path, with skip connections between corresponding layers. The encoder is responsible for capturing the high-level context of the image. The decoder then uses transposed convolutions for performing precise localisation. The network can learn from both high-level and low-level features because of the skip connections between corresponding encoder and decoder layers, which also increase the model's segmentation accuracy. Post-processing steps are not necessary because the model is trained end-to-end and can predict the segmentation mask for the input image directly. It predicts a label for each pixel in the input image and approaches the segmentation task as a pixel-by-pixel classification task. Figure 11 illustrates this architecture. This has inspired subsequent research in image segmentation with many models building upon its principles. The modifications to this architecture that has been used for this project are explained in section 4.1

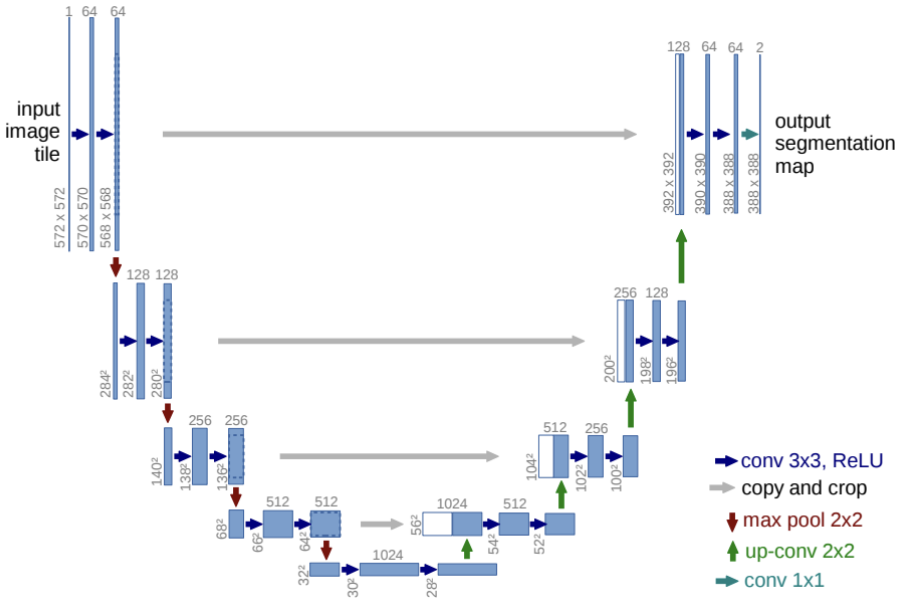

Figure 11: U-Net architecture for 32x32 pixels at its lowest resolution level. Blue boxes indicate a feature map with multiple channels. White boxes are feature maps copied from the encoder. The number of channels and the dimensions are displayed on top and bottom left corner of each box respectively. Arrows represent different operations within the network (Ronneberger et al., 2015)

(Oktay et al., 2018) modified the U-Net architecture by adding an attention mechanism between the encoder feature maps and the respective decoder feature maps. This attention mechanism helps the network focus on the most relevant parts of the image. In medical image segmentation, where the region of interest is small and surrounded by similar tissues, this selective focus is crucial. The attention gates help the network learn which regions of the encoder feature map are the most important by taking into account the feature map's context from the previous decoder block. In order to do this, the encoder feature map is multiplied by the weights determined by the attention gates.

In contrast to U-Net, which employs direct skip connections between the feature maps of the encoder and the decoder, (Zhou et al., 2018) used a nested and dense skip connection structure by using convolutional blocks between encoder and decoder feature maps. These stacked skip connections help in reducing the semantic gap between the encoder and decoder part of the network. The redesigned skip connections, consisting of convolutional blocks, are crucial in capturing fine details in the image resulting in better segmentation results.

(Milletari et al., 2016) modified the U-Net architecture by using 3D convolutions tailored specifically for 3D medical images and was one of the first to handle volumetric medical images. It is able to maintain spatial information of the input 3D volumes by processing the images in 3D rather than in a 2D slice by slice manner. This model is trained in an end-to-end manner, which means it takes in 3D volumes as input and predicts 3D segmentation labels without the need for extensive preprocessing or post processing.

(Myronenko, 2019) combined the use of autoencoders as a regularization step with an encoder decoder architecture. This method extracts features from the input 3D MRI scans using an encoder. Subsequently, a variational autoencoder (VAE) is used to reconstruct the input image based on the extracted features. Additionally, a decoder is used to generate the segmentation maps for the input image. By using variational autoencoder regularisation along with the U-Net architecture, overfitting can be reduced and the model's ability to generalise to new unseen data can be improved.

(Henry et al., 2021) used a self-ensembled 3D U-Net architecture where multiple models are trained, and their outputs are averaged to get the final segmentation map. This improved overall segmentation quality by leveraging the strengths of each of the

individual model. Additionally, the network uses deep supervision by adding auxiliary loss functions at each of the decoder layers except the final layer. This helps the network learn better by accelerating the convergence of training and leads to improved segmentation results (Wang et al., 2015).

Most of the previous research in brain tumor segmentation has been done using pre-treatment MRI scans. In fact, a survey found that 98.3% of published research in glioma segmentation used pre-treatment MRI scans using data from The Cancer Imaging Archives (TCIA) or Brain Tumor Segmentation (BraTS) challenges prior to 2024 (de Verdier et al., 2024).

More recently, researchers have started exploring brain tumor segmentation on post-treatment MRI scans. For instance, (Ghaffari et al., 2022) introduced a novel deep learning approach for segmenting brain tumors in post-operative MRI scans, a task that is complicated by the anatomical changes following surgery. The authors proposed using transfer learning from pre-operative images, where tumor boundaries are more clearly defined, to improve segmentation accuracy in post-operative images. By training the model on pre-operative data and fine-tuning it for post-operative images, the approach significantly enhanced the model's ability to generalise well.

4 Methodology

4.1 Model Architecture

The U-Net architecture forms the backbone of the proposed model. The U-Net structure consists of two main components: the encoder (downsampling path) and the decoder (upsampling path), connected through a bottleneck layer.

Residual Blocks

At the core of this architecture is the residual block, a building block inspired by the ResNet architecture (He et al., 2016). Residual learning was introduced to solve the problem of vanishing gradients by allowing gradients to flow smoothly during backpropagation and preventing them from being too close to zero. The residual block consists of two 3D convolutional layers, each followed by batch normalization and a ReLU activation function. It also includes a skip connection layer, which directly adds the input of the block to its output. This skip connection helps gradients flow more effectively through the network during training. These connections also contribute to the network's robustness by allowing it to bypass unnecessary layers, thereby focusing on essential features. This characteristic is particularly important in brain tumor segmentation, where the model needs to capture both fine details and high-level structures.

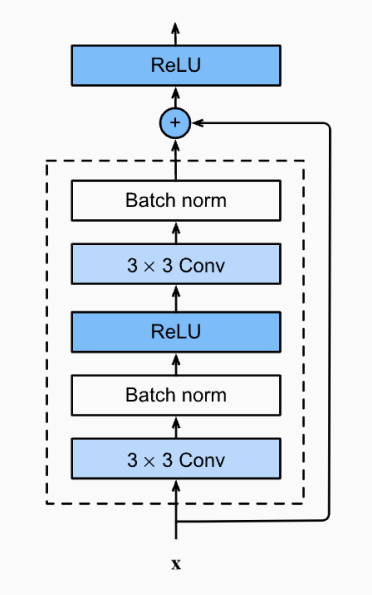

Figure 12: Residual block (“8.6. Residual Networks (ResNet) and ResNeXt — Dive into Deep Learning 1.0.3 documentation,” n.d.)

Encoder

The encoder path extract features from the input while gradually reducing its spatial resolution. It consists of four residual blocks described above, each followed by a max-pooling operation. The max-pooling layers are used to reduce the dimensions of the feature maps, which allows the network to focus on more abstract and complex features. As the network’s depth increases, it captures progressively global context information, which is essential for recognizing the overall structure of the target tumor or resection cavity regions in the MRI scan.

Each residual block in the encoder captures essential spatial features at different levels of abstraction. Early blocks capture fine details such as edges and textures, while deeper blocks focus on more complex patterns and structures. The use of residual connections ensures that the model can capture both shallow and deep features, improving the overall segmentation accuracy.

Bottleneck

The bottleneck layer serves as a bridge between the encoder and decoder paths. This layer captures the most abstract and high-level features from the input image. To prevent overfitting, a dropout layer is applied after the bottleneck, which randomly zeroes out some of the activations during training. This dropout layer performs the role of regularisation in the model and helps the network generalise better to unseen data.

The bottleneck layer's ability to capture high-level, abstract features is crucial for understanding the overall context of the image. This understanding is especially important in brain tumor segmentation, where the model must distinguish between different tissue types and identify regions of interest such as tumors or lesions.

Decoder

The decoder path is responsible for reconstructing the spatial resolution of the input image while refining the segmentation masks. Each upsampling operation in the decoder is implemented using a transpose convolution layer, which increases the spatial dimensions of the feature maps. The upsampled features are then concatenated with the corresponding features from the encoder path. This ensures that the decoder can learn from both high-level context and fine details. In addition to the bottleneck layer, dropout regularisation layers are applied after each decoder block to further improve the model's ability to generalise well to unseen data.

Attention Block

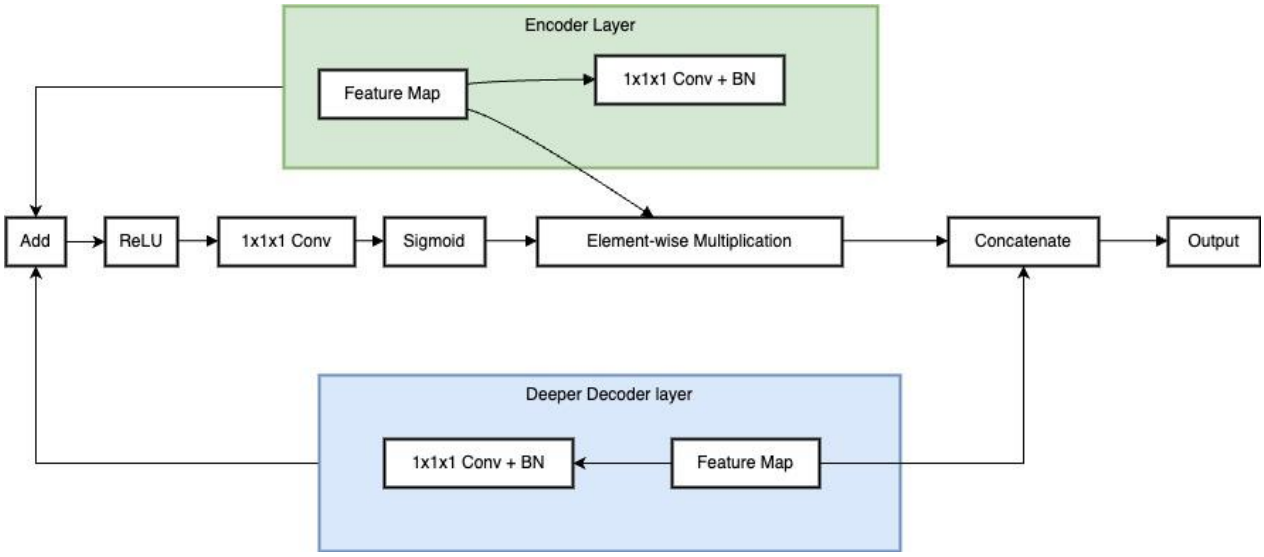

Figure 13: Attention block: Feature maps from encoder and decoder layers undergo 1x1x1 convolutions and batch normalization. After summing and applying ReLU activation, a 1x1x1 convolution reduces the map to a single-channel attention map. A sigmoid function normalizes it, and the attention mask is multiplied with encoder features before concatenation with decoder feature maps.

Although residual blocks help the network learn progressively more abstract features in an efficient manner in the encoder path while increasing the network's depth, repeated application of residual connections in the network leads to loss of spatial information in the feature maps (Liu et al., 2023). To recover this loss of spatial information, a spatial attention block is used between the upsampling feature maps of the decoder and the corresponding feature maps from the encoder. This allows the network to focus on the most relevant regions in the feature maps while suppressing irrelevant background information.

The attention mechanism works by combining the global features from the upsampled decoder output and the local features from the encoder. The feature maps from the deeper decoder layer and the encoder layer are each passed through separate 1x1x1 convolutional layers followed by batch normalization. These layers reduce the number of channels and transform the feature maps into a common dimensional space. After this transformation, the features are added element-wise. This sum is then passed through a ReLU activation function, which introduces non-linearity into the model. This step allows the network to learn complex combinations of the input features and highlight the areas that are important for the specific task at hand. The combined feature map is then passed through another 1x1x1 convolutional layer that reduces it to a single-channel attention map. A sigmoid activation function is then used to normalize this map to values between 0 and 1, creating a spatial attention mask that highlights important regions. This spatial attention map is then applied to the original encoder feature maps by performing an element-wise multiplication before they are concatenated with the decoder features. Figure 13 illustrates the working of the attention block.

Final Output

The final layer of the model is a 1x1x1 convolution, which reduces the number of channels to match the desired output which is the number of segmentation classes. This layer generates the final segmentation map, which contains the predicted labels for each voxel in the input volume. The use of a 1x1x1 convolution ensures that the model can produce a segmentation map with the same spatial resolution as the input image.

The architecture of the proposed model is shown in Figure 14

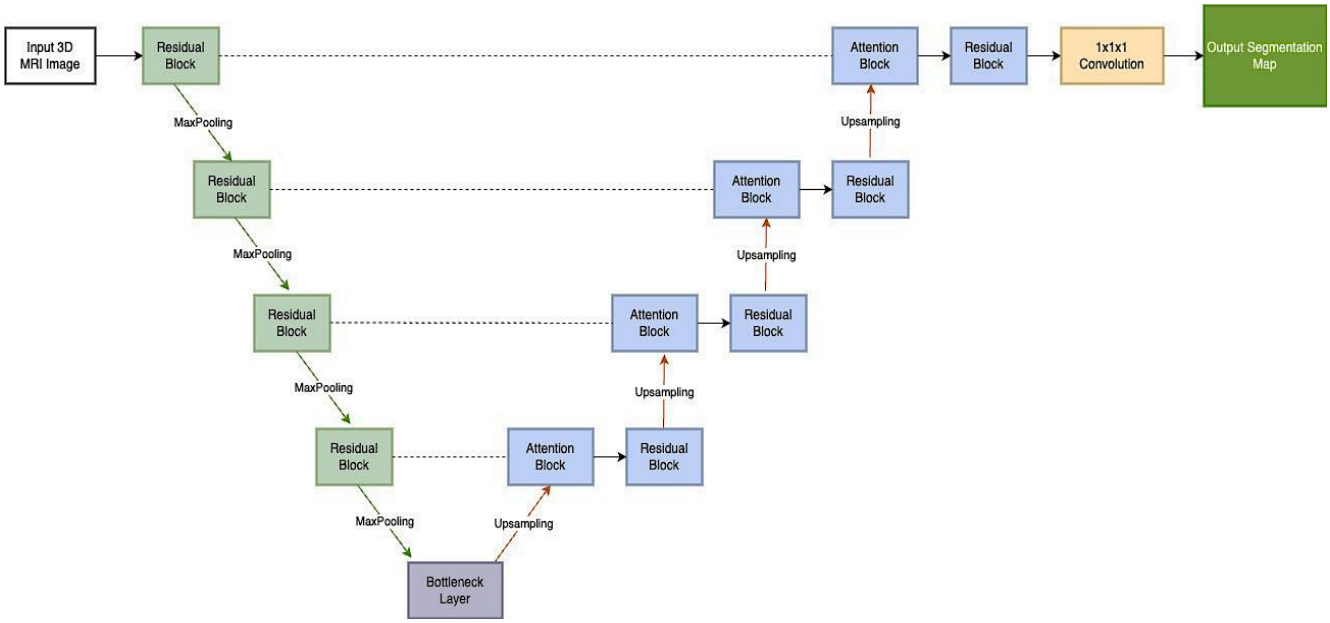

Figure 14: Residual 3D U-Net with Attention architecture

4.2 Loss Function

This project uses a customised loss function by combining cross-entropy loss and focal loss to mitigate the effects of the inherently imbalanced data.

Cross-entropy loss is a common loss function used for segmentation tasks. It is defined as:

$$CrossEntropyLoss(CE) = -\sum_{c=1}^{C} y_c \log(\hat{y}_c)$$

where y_c is the true label, $\hat{y}_c$ is the predicted probability for class c and C is the number of classes.

However, cross-entropy loss treats all pixels equally, which can result in the model focusing too much on the majority class. Focal loss was introduced to counter this issue (Lin et al., 2017). It modifies cross-entropy loss by adding a modulating factor that up-weights the loss for hard-to-classify examples and down-weights the loss for easy-to-classify examples. This means that Focal Loss inherently focuses more on the Whole Tumor and Resection Cavity classes, which are underrepresented and harder for the model to segment accurately. Focal loss is defined as

$$\text{Focal Loss, FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$

where, p_t is the predicted probability for the correct class, α_t is a weighting factor for class balance, and γ is the focusing parameter, which adjusts the rate at which easy examples are down-weighted.

When the model is trained using only focal loss, it can cause it to focus too much on the hard examples. This leads to overfitting on the tumor and resection cavity classes and underfitting on the majority class. Including the cross-entropy loss provides a regularizing effect by preventing the model from being overly biased toward any specific class. The combined loss used in this project can be defined as

$$CombinedLoss = (1-\lambda) \times \left(-\sum_{c=1}^{C} y_c \log(\widehat{y_c})\right) + \lambda \times (-\alpha_t(1-\widehat{y_t})^{\gamma}\log(\widehat{y_t}))$$

Where λ is a hyperparameter used to control the effect of each of the individual loss components in the combined loss equation.

4.3 Implementation detail

The initial training was conducted using the Focal loss function (described above), Adam optimizer and a learning rate of 0.0001, with a batch size of 1. However, these parameters led to instability with larger datasets. Increasing the learning rate caused the algorithm to diverge, producing NaN values for the loss function. Conversely, lowering the learning rate led to early stopping and halted the training process prematurely. To address this, a new set of parameters was implemented. First, a learning rate scheduler ReduceLROnPlateau has been used, which reduced the learning rate when the validation loss stopped improving. Additionally, the loss function has been customised by combining focal loss with cross-entropy loss as explained in section 4.2 above to ensure network stability. Various loss functions, including Tversky loss (Salehi et al., 2017) and Dice loss (Sudre et al., 2017), were also tested, but the combined loss function showed the most stable behaviour. The optimizer, batch size and other parameters remained unchanged.

5 Experiments and Results

The experimental setup in this study involved the steps of loading the data along with data preprocessing, a comparative study of different U-Net based architectures and quantitative evaluation of their segmentation performance along with a qualitative evaluation of the proposed Residual 3D U-Net with Attention model.

The dataset used in this project is from the 2024 Brain Tumor Segmentation (BraTS) Challenge focusing on the segmentation of post-treatment glioma using MRI data (de Verdier et al., 2024). Due to resource constraints, a subset of the data has been used out of the total 205 GB of the data made available as part of the challenge.

The dataset is composed of multiparametric MRI (mpMRI) acquisition protocols, which include pre-contrast T1-weighted (T1), contrast-enhanced T1-weighted (T1-Gd), T2-weighted (T2), and T2-weighted fluid-attenuated inversion recovery (FLAIR) sequences for post-treatment imaging along with the ground truth segmentation masks for each patient. The pixel intensities have been normalised by subtracting the mean and dividing it by the standard deviation.

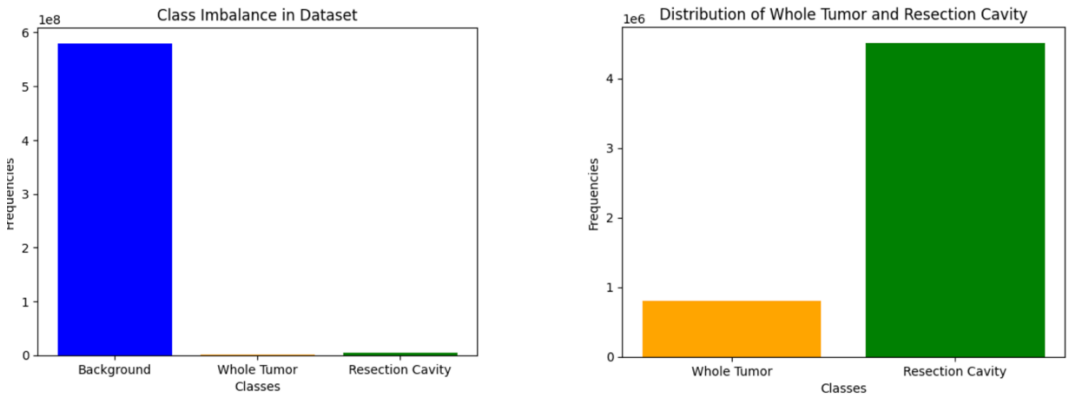

Figure 15: Left: Class imbalance in dataset. Right: distribution of Whole Tumor and Resection Cavity in the dataset

As we can see from Figure 15, the dataset has a significant class imbalance. In the first chart, the background labels dominate the dataset with an overwhelming

frequency of 99% while the Whole Tumor and Resection Cavity labels have a much lower frequency highlighting their relatively small representation in the data. As expected, this imbalance suggests that the data consists of mainly non-tumor regions, which could potentially bias a segmentation model towards predicting the background class more often than the actual tumor or resection cavities.

The second chart offers a more detailed look at the distribution between the Whole Tumor and Resection Cavity classes specifically. Here, the Resection Cavity has a higher frequency compared to the Whole Tumor. This difference further underscores an imbalance between these two classes, with the Resection Cavity being more prevalent in the dataset.

To address this imbalance, a customized loss function approach has been implemented that combines Focal Loss and Cross-Entropy Loss as described in section 4.2 above. This aims to enhance the model's ability to learn from underrepresented classes (Whole Tumor and Resection Cavity) and improve overall segmentation results.

5.1 Evaluation metrics

For this project, Lesion-wise Dice Coefficient and Lesion-wise Sensitivity have been used for evaluating the results of the segmentation models. The Lesion-wise Dice Coefficient gives insight into the overlap accuracy for each individual lesion, while the Lesion-wise Sensitivity focuses on the model's ability to detect all lesions.

Lesion-wise Dice Coefficient

The Lesion-wise Dice Coefficient is a variation of the standard Dice Coefficient tailored specifically to evaluate the segmentation performance on a per-lesion basis. In medical imaging, particularly when dealing with diseases like cancer, it is crucial not only to segment the entire affected area but also to accurately identify each individual lesion.

The Lesion-wise Dice Coefficient is calculated similarly to the standard Dice Coefficient but is adapted to assess the similarity between the predicted lesions and the ground truth lesions on an individual basis. It is defined as:

$$Lesion - Wise\ Dice\ Coefficient = \frac{2 \times |L_{\text{pred}} \cap L_{\text{true}}|}{|L_{\text{pred}}| + |L_{\text{true}}|}$$

where, L_{pred} is the predicted lesion and L_{true} is the ground truth lesion in the image.

The range for Lesion-wise Dice Coefficient is from 0 to 1, where 0 means there is no overlap between predicted lesion and the actual lesion and 1 means perfect segmentation with predicted lesions being the same as the actual lesions.

This metric is particularly useful in scenarios where it is critical to assess how well the model detects each individual lesion rather than just the overall volume.

Lesion-wise Sensitivity

Lesion-wise Sensitivity measures the ability of the model to correctly identify all true lesions. It focuses on the detection rate of lesions rather than the pixel-wise overlap.

Lesion-wise Sensitivity is calculated as:

$$Lesion - wise\ Sensitivity = \frac{\text{TP}_{\text{lesions}}}{\text{TP}_{\text{lesions}} + \text{FN}_{\text{lesions}}}$$

where, $\text{TP}_{\text{lesions}}$ are samples correctly classified as lesions and $\text{FN}_{\text{lesions}}$ are lesions incorrectly classified as negative (not a lesion).

Lesion-wise Sensitivity ranges from 0 to 1, where 1 means that all actual lesions have been correctly detected and 0 indicates that none of the actual lesions have been detected.

This metric is particularly important in clinical practice, as missing a lesion can lead to significant consequences in patient management. It ensures that the model can identify every lesion present, which is crucial for diseases such as glioma where early and comprehensive detection of all abnormal tissue areas is vital.

5.2 Comparative Study

A comparative study has been done to evaluate the performance of different U-Net based architectures for post-treatment MRI segmentation for gliomas, specifically focusing on the whole tumor and resection cavity. Lesion wise dice coefficient and

sensitivity has been used (as explained in section 5.1 above) as the primary evaluation metrics.

5.2.1 Models Evaluated

For this study, four different models have been experimented with:

1. **3D U-Net**: The 3D U-Net model has been considered the baseline model. This is a modification to the standard U-Net architecture by using 3D convolutions instead of 2D convolutions to process volumetric data such as brain MRI scans. Inspired by (Milletari et al., 2016), 3D convolutions used in this model are able to process the volumetric data as a whole in comparison to standard 2D convolutions in U-Net models which process data in a 2D slice by slice manner.

2. **3D U-Net with Attention**: Based on (Oktay et al., 2018), attention mechanisms were introduced to the 3D U-Net model discussed above. As discussed previously, the idea behind adding attention is to help the network focus on the most relevant parts of the image, thereby potentially improving the segmentation quality by focusing on different features according to their importance.

3. **Residual 3D U-Net**: Next, the impact of adding residual connections (He et al., 2016) to the 3D U-Net architecture was explored. As described earlier, residual connections are a type of skip connection that help address the vanishing gradient problem and improve model training.

4. **Residual 3D U-Net with Attention**: Finally, both residual connections and attention mechanisms were combined in the proposed method. The hypothesis is that leveraging both these enhancements could improve segmentation performance further by combining the advantages of residual connections as well as attention mechanisms.

5.3 Results

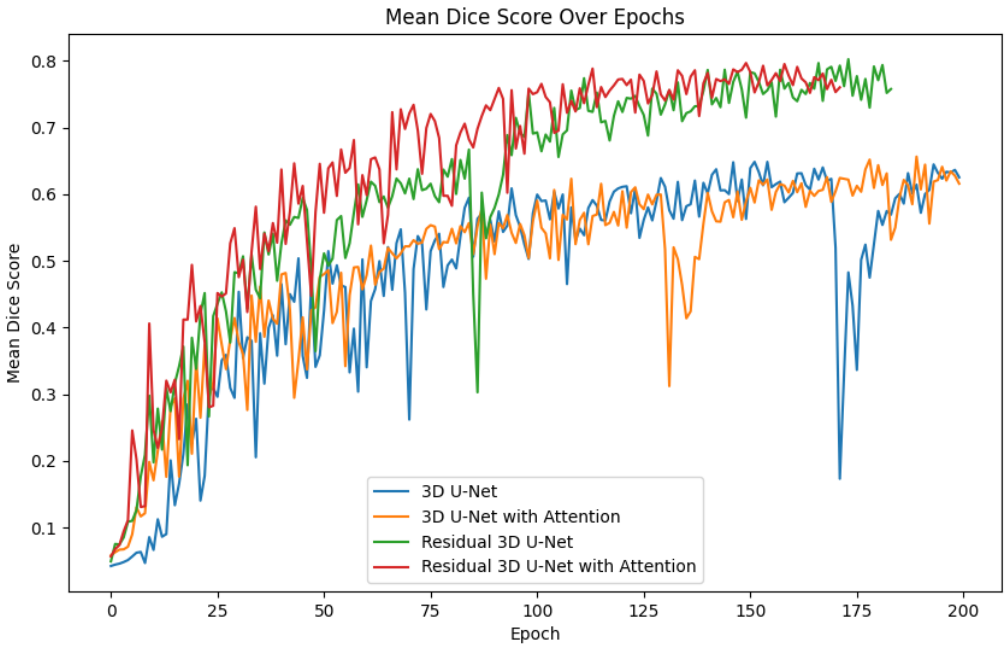

Figure 16: Mean Dice scores over epochs during training

Figure 16 shows the mean Dice score for Whole Tumor and Resection Cavity on the validation data across epochs for the four models: 3D U-Net, 3D U-Net with Attention, Residual 3D U-Net, and Residual 3D U-Net with Attention. As training progresses, the Dice scores steadily improve, indicating that the models are learning to better predict segmentation masks. There is some fluctuation in the Dice scores during the early epochs, which is expected as the models adjust their parameters and learn to better predict the segmentation masks.

The proposed Residual 3D U-Net with Attention model outperforms the others, consistently achieving the highest scores and demonstrating more stable learning. This model, along with the Residual 3D U-Net, eventually stabilizes around a mean Dice score of close to 0.8, suggesting these models are more effective for the segmentation task. In contrast, the 3D U-Net models without residual connections exhibit more significant fluctuations and lower final Dice scores, indicating potential limitations in their learning capacity. Overall, the graph highlights that incorporating

residual connections and attention mechanisms enhances the model's segmentation performance, leading to more accurate and reliable predictions.

5.3.1 Quantitative Results

In the comparative study, certain differences have been observed in segmentation performance across the four models. These differences are summarized by the lesion wise Dice coefficients as well as the lesion wise sensitivity for the whole tumor, resection cavity as shown in Table 1

Table 1: Quantitative results for comparative study

Models	Whole Tumor		Resection Cavity	
	Dice score	Sensitivity	Dice score	Sensitivity
3D U-Net	0.75	0.99	0.56	0.56
3D U-Net with Attention	0.73	1.00	0.56	0.58
Residual 3D U-Net	0.84	0.98	0.54	0.54
Residual 3D U-Net with Attention	0.87	0.96	0.63	0.61

- **3D U-Net**: The baseline 3D U-Net model achieved a Dice coefficient of 0.75 for whole tumor segmentation, suggesting a strong ability to identify and segment the whole tumor region. The model also displayed a very high sensitivity for the whole tumor, indicating its effectiveness in identifying most of the tumor regions. However, its performance for the resection cavity is much lower, with a Dice coefficient of 0.56 and sensitivity also at 0.56. This decline is due to the challenges of segmenting the resection cavity, which may vary in shape and contain different materials like fluid and air. Overall, this model performs reasonably well for tumor segmentation but struggles with the resection cavity.
- **3D U-Net with Attention**: When attention mechanisms were added to the 3D U-Net, the Dice coefficient for whole tumor segmentation slightly decreased to 0.73. Despite this, the model maintained perfect sensitivity (1.00) for whole tumor segmentation, meaning it still captures all of the whole tumor tissues. The

performance for the resection cavity remains the same as the baseline model, with a Dice coefficient of 0.56 and a slight increase in sensitivity to 0.58. This suggests that the addition of attention mechanisms does not significantly enhance the segmentation of more challenging regions, like the resection cavity, but ensures better recall for tumor regions.

- **Residual 3D U-Net**: Introducing residual connections into the 3D U-Net architecture led to a notable improvement in performance. The Dice coefficient for whole tumor segmentation increased to 0.84, demonstrating the model's better ability to learn from complex, deeper features in the data. Sensitivity for the whole tumor remains high at 0.98. However, for the resection cavity, the Dice coefficient dropped slightly to 0.54, and sensitivity too decreased to 0.54. Although the model excels in whole tumor segmentation, its ability to segment the resection cavity is less robust, which could be due to the cavity's complexity and varying content.
- **Residual 3D U-Net with Attention**: The combination of residual connections with attention mechanisms resulted in the highest performance across the models. For whole tumor segmentation, this hybrid approach achieved a Dice coefficient of 0.87 with a sensitivity of 0.96, indicating superior performance in identifying and segmenting the tumor tissues correctly. For resection cavity segmentation, this model also saw the highest improvement, achieving a Dice coefficient of 0.63 and a sensitivity of 0.61. This improvement highlights the complementary nature of residual connections and attention mechanisms: the residual connections help the model learn from deeper layers, while attention mechanisms allow the model to focus on the most relevant regions for segmentation.

Overall, the results suggest that adding residual connections significantly improve segmentation performance, particularly for whole tumor segmentation by enabling the model to learn deeper features. The attention mechanism alone did not drastically enhance performance, but when combined with residual connections, it led to the highest overall performance by improving the model's ability to segment both the whole tumor and the resection cavity. Although the resection cavity remains challenging to segment across all models, the inclusion of attention mechanisms along with residual connections helps refine the segmentation process.

5.3.2 Qualitative results

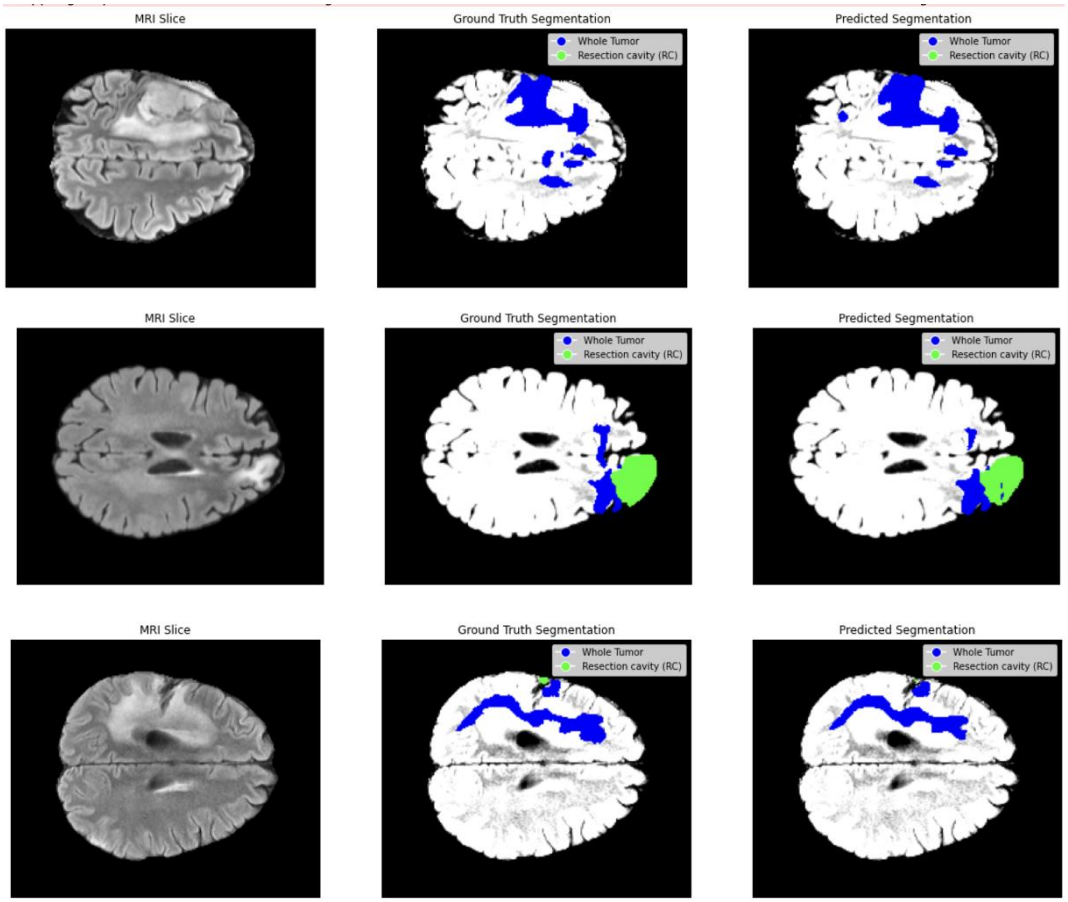

Figure 17: Comparison of predicted segmentation and ground truth segmented labels

Figure 17 provides a visual representation of the segmentation performance of the proposed model. Each row corresponds to a different MRI slice, which shows the model's ability to segment the whole tumor (WT) and resection cavity (RC) compared to the ground truth labels. This evaluation helps us understand not only where the model succeeds but also where it faces challenges, which is crucial for refining its performance in clinical applications.

In the first scan, the model demonstrates a high degree of accuracy in segmenting the whole tumor. The model accurately captures the overall shape and size of the tumor, including the more subtle variations in intensity within the MRI slice. This level of

precision indicates that the model is effectively learning to differentiate between normal and abnormal brain tissues. The close match between the model's output and the ground truth suggests that the attention mechanism is effectively helping the model focus on relevant regions, enhancing its ability to detect and segment the tumor as well as post-treatment changes in the MRI scan.

The second scan is more complex where the resection cavity has a larger, irregular shape. Despite these complexities, the model still performs relatively well, capturing the general shape and position of both the whole tumor and the resection cavity. However, there are some discrepancies in the segmentation boundaries. For instance, the green region representing the resection cavity appears slightly more expansive in the predicted segmentation compared to the ground truth label. This could be a result of the model’s sensitivity to variations in intensity and texture within the MRI scan, resulting in the model segmenting the surrounding tissues as resection cavity. While this sensitivity is useful for detecting diffuse tumor regions, it also poses a challenge in identifying precise boundaries, especially in cases with heterogeneous tissues such as those present in gliomas.

The third scan illustrates a case where the segmentation task becomes significantly more challenging due to the irregular shape of both the tumor and resection cavity. In this instance, the model’s segmentation deviates more significantly from the ground truth. The segmented whole tumor extends beyond the actual boundary as defined by the ground truth labels, especially around the resection cavity. Additionally, there are regions where the whole tumor’s segmentation does not align with the ground truth, particularly in regions where the tumor has thin, elongated extensions. This discrepancy indicates that while the model performs well in segmenting more well-defined regions, it struggles with finer structures that require a higher level of detail and precision.

6 Discussion and Future Work

6.1 Limitations

Overall, this project performs well in most cases, accurately identifying both whole tumors and resection cavities. However, there are some limitations and challenges associated with this research. One of the main challenges was evaluating and comparing the performance of different segmentation models against the proposed model. The longer model training times, due to limited hardware capabilities, made it difficult to train other state-of-the-art models within the available timeframe. Additionally, because only a subset of the dataset was used for training, it was not possible to fully compare the proposed model with other advanced methods, which may have affected the model's generalisation ability.

Moreover, this problem is relatively unexplored, and there are limited existing solutions for post-treatment brain tumor segmentation, mainly for resection cavity segmentation. Most current brain tumor segmentation models are trained on pre-treatment MRI scans and cannot be directly applied to post-treatment scenarios. These existing models need to be retrained from scratch to make a fair comparison, which was not feasible given the resource constraints.

Lastly, the use of advanced techniques such as attention mechanisms and residual blocks added to the model's complexity. Due to the limited hardware, training had to be done with fewer epochs with longer training times, which also likely resulted in less thorough hyperparameter optimization.

6.2 Future Work

Future work can address several key areas to enhance the effectiveness of segmentation models, particularly in overcoming the limitations imposed by limited computational resources in the current project.

First, expanding the training data to use the entire BraTS dataset is a logical next step. Only a subset of the dataset was used in this project, which may have restricted the model’s generalisation capabilities. Training on the full dataset can lead to better performance.

Second, future work could explore the use of generative models like GANs and VAEs to generate synthetic data, particularly for underrepresented classes. These generative models can create diverse, realistic images that could enhance the model's performance across different images. Again, the implementation of these advanced techniques is resource-intensive, which posed a challenge in the current project.

Cross-dataset generalisation is another crucial area for exploration. Future work could aim to develop models that are not only effective on post-treatment imaging but can also perform well on pre-treatment scans. Techniques like transfer learning and domain adaptation could be utilized to bridge this gap, ensuring that models are versatile and can be applied in different clinical scenarios.

Using data from multiple different modalities represents another promising avenue of research. While this study focused on MRI scans, combining data from other imaging modalities like PET or CT scans could significantly improve the accuracy of segmentation models. These modalities provide different types of tissue contrast and biological markers, which can complement MRI data, leading to more robust models.

Finally, ensemble methods offer a way to combine the strengths of different models trained in this study. By integrating models like the 3D U-Net, Residual 3D U-Net, and their attention-based variants, an ensemble could deliver superior performance by leveraging the unique advantages of each model, leading to more accurate and reliable segmentation results.

Addressing these areas in future work will not only overcome the limitations encountered in this study but also significantly advance the field of glioma segmentation, making the models more robust, generalizable, and clinically relevant.

7 Conclusion

In conclusion, this research has demonstrated that the proposed model significantly enhances the segmentation of gliomas in post-treatment MRI scans. The results indicate that including both residual connections and attention mechanisms together allows the model to learn more complex features as well as focus on relevant features, leading to better segmentation performance compared to other models tested.

A key contribution of this project is its focus on the area of post-treatment glioma segmentation, addressing a critical gap in current research. While most prior studies have concentrated on pre-treatment scans, this work provides a novel approach tailored specifically to the challenges of post-treatment imaging. This advancement is important for improving clinical decision-making and treatment planning in glioma patients, improving overall patient care.

8 Bibliography

8.6. Residual Networks (ResNet) and ResNeXt — Dive into Deep Learning 1.0.3 documentation [WWW Document], n.d. URL https://d2l.ai/chapter_convolutional-modern/resnet.html (accessed 9.6.24).

Albelwi, S., Mahmood, A., 2017. A Framework for Designing the Architectures of Deep Convolutional Neural Networks. Entropy 19, 242. https://doi.org/10.3390/e19060242

Bakshi, R., Ariyaratana, S., Benedict, R.H.B., Jacobs, L., 2001. Fluid-Attenuated Inversion Recovery Magnetic Resonance Imaging Detects Cortical and Juxtacortical Multiple Sclerosis Lesions. Arch Neurol 58, 742. https://doi.org/10.1001/archneur.58.5.742

Basavarajaiah, M., 2022. 6 basic things to know about Convolution. Medium. URL https://medium.com/@bdhuma/6-basic-things-to-know-about-convolution-daef5e1bc411 (accessed 9.1.24).

Casamitjana, A., Puch, S., Aduriz, A., Vilaplana, V., 2017. 3D Convolutional Neural Networks for Brain Tumor Segmentation: A Comparison of Multi-resolution Architectures. https://doi.org/10.48550/ARXIV.1705.08236

Clark, M.C., Hall, L.O., Goldgof, D.B., Velthuizen, R., Murtagh, F.R., Silbiger, M.S., 1998. Automatic tumor segmentation using knowledge-based techniques. IEEE Trans. Med. Imaging 17, 187–201. https://doi.org/10.1109/42.700731

de Verdier, M.C., Saluja, R., Gagnon, L., LaBella, D., Baid, U., Tahon, N.H., Foltyn-Dumitru, M., Zhang, J., Alafif, M., Baig, S., Chang, K., D'Anna, G., Deptula, L., Gupta, D., Haider, M.A., Hussain, A., Iv, M., Kontzialis, M., Manning, P., Moodi, F., Nunes, T., Simon, A., Sollmann, N., Vu, D., Adewole, M., Albrecht, J., Anazodo, U., Chai, R., Chung, V., Faghani, S., Farahani, K., Kazerooni, A.F., Iglesias, E., Kofler, F., Li, H., Linguraru, M.G., Menze, B., Moawad, A.W., Velichko, Y., Wiestler, B., Altes, T., Basavasagar, P., Bendszus, M., Brugnara, G., Cho, J., Dhemesh, Y., Fields, B.K.K., Garrett, F., Gass, J., Hadjiiski, L., Hattangadi-Gluth, J., Hess, C., Houk, J.L., Isufi, E., Layfield, L.J., Mastorakos, G., Mongan, J., Nedelec, P., Nguyen, U., Oliva, S., Pease, M.W., Rastogi, A., Sinclair, J., Smith, R.X., Sugrue, L.P., Thacker, J., Vidic, I., Villanueva-Meyer, J., White, N.S., Aboian, M., Conte, G.M., Dale, A., Sabuncu, M.R., Seibert, T.M., Weinberg, B., Abayazeed, A., Huang, R., Turk, S., Rauschecker, A.M., Farid, N., Vollmuth, P., Nada, A., Bakas, S., Calabrese, E., Rudie, J.D., 2024. The 2024 Brain Tumor Segmentation (BraTS) Challenge: Glioma Segmentation on Post-treatment MRI. https://doi.org/10.48550/ARXIV.2405.18368

Despotović, I., Goossens, B., Philips, W., 2015. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. Computational and Mathematical Methods in Medicine 2015, 1–23. https://doi.org/10.1155/2015/450341

Dixit, P., 2021. Max Pooling, Why use it and its advantages. Geek Culture. URL https://medium.com/geekculture/max-pooling-why-use-it-and-its-advantages-5807a0190459 (accessed 9.1.24).

Dvořák, P., Menze, B., 2016. Local Structure Prediction with Convolutional Neural Networks for Multimodal Brain Tumor Segmentation, in: Menze, B., Langs, G., Montillo, A., Kelm, M., Müller, H., Zhang, S., Cai, W., Metaxas, D. (Eds.),

Medical Computer Vision: Algorithms for Big Data, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 59–71. https://doi.org/10.1007/978-3-319-42016-5_6
Elnakib, A., Gimel'farb, G., Suri, J.S., El-Baz, A., 2011. Medical Image Segmentation: A Brief Survey, in: El-Baz, A.S., Acharya U, R., Laine, A.F., Suri, J.S. (Eds.), Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies. Springer New York, New York, NY, pp. 1–39. https://doi.org/10.1007/978-1-4419-8204-9_1
Fernando, K.R.M., Tsokos, C.P., 2023. Deep and statistical learning in biomedical imaging: State of the art in 3D MRI brain tumor segmentation. Information Fusion 92, 450–465. https://doi.org/10.1016/j.inffus.2022.12.013
Ghaffari, M., Samarasinghe, G., Jameson, M., Aly, F., Holloway, L., Chlap, P., Koh, E.-S., Sowmya, A., Oliver, R., 2022. Automated post-operative brain tumour segmentation: A deep learning model based on transfer learning from pre-operative images. Magnetic Resonance Imaging 86, 28–36. https://doi.org/10.1016/j.mri.2021.10.012
Guo, D., Pei, Y., Zheng, K., Yu, H., Lu, Y., Wang, S., 2020. Degraded Image Semantic Segmentation With Dense-Gram Networks. IEEE Trans. on Image Process. 29, 782–795. https://doi.org/10.1109/TIP.2019.2936111
Hassaballah, M., Awad, A.I. (Eds.), 2020. Deep Learning in Computer Vision: Principles and Applications. CRC Press, Boca Raton. https://doi.org/10.1201/9781351003827
Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2017. Brain tumor segmentation with Deep Neural Networks. Medical Image Analysis 35, 18–31. https://doi.org/10.1016/j.media.2016.05.004
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
Henry, T., Carré, A., Lerousseau, M., Estienne, T., Robert, C., Paragios, N., Deutsch, E., 2021. Brain Tumor Segmentation with Self-ensembled, Deeply-Supervised 3D U-Net Neural Networks: A BraTS 2020 Challenge Solution, in: Crimi, A., Bakas, S. (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 327–339. https://doi.org/10.1007/978-3-030-72084-1_30
Hospitals, Y., 2018. What is the most aggressive form of glioma? Is there a complete cure? [WWW Document]. Yashoda Hospitals. URL https://www.yashodahospitals.com/blog/glioma-causes-symptoms-diagnosis-treatment/ (accessed 9.1.24).
Isachesku, E., Braicu, C., Pirlog, R., Kocijancic, A., Busuioc, C., Pruteanu, L.-L., Pandey, D.P., Berindan-Neagoe, I., 2023. The Role of Non-Coding RNAs in Epigenetic Dysregulation in Glioblastoma Development. IJMS 24, 16320. https://doi.org/10.3390/ijms242216320
Jesson, A., Arbel, T., 2018. Brain Tumor Segmentation Using a 3D FCN with Multi-scale Loss, in: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries,

Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 392–402. https://doi.org/10.1007/978-3-319-75238-9_34
Kandel, I., Castelli, M., 2020. Transfer Learning with Convolutional Neural Networks for Diabetic Retinopathy Image Classification. A Review. Applied Sciences 10, 2021. https://doi.org/10.3390/app10062021
Knapp, S., 2020. Glial Cells - The Definitive Guide. Biology Dictionary. URL https://biologydictionary.net/glial-cells/ (accessed 9.1.24).
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.
Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal Loss for Dense Object Detection. https://doi.org/10.48550/ARXIV.1708.02002
Liu, Z., Tong, L., Chen, L., Jiang, Z., Zhou, F., Zhang, Q., Zhang, X., Jin, Y., Zhou, H., 2023. Deep learning based brain tumor segmentation: a survey. Complex Intell. Syst. 9, 1001–1026. https://doi.org/10.1007/s40747-022-00815-5
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, pp. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965
Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W., 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathologica 131, 803–820. https://doi.org/10.1007/s00401-016-1545-1
Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., Golland, P., 2010. A Generative Model for Brain Tumor Segmentation in Multi-Modal Images, in: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 151–159. https://doi.org/10.1007/978-3-642-15745-5_19
Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV). Presented at the 2016 Fourth International Conference on 3D Vision (3DV), IEEE, Stanford, CA, USA, pp. 565–571. https://doi.org/10.1109/3DV.2016.79
Murphy, A., Gaillard, F., 2015. MRI sequences (overview), in: Radiopaedia.Org. Radiopaedia.org. https://doi.org/10.53347/rID-37346
Myronenko, A., 2019. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization, in: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, Cham, pp. 311–320. https://doi.org/10.1007/978-3-030-11726-9_28
Namburu, A., Barpanda, S.S. (Eds.), 2020. Recent Advances in Computer Based Systems, Processes and Applications, 1st ed. CRC Press. https://doi.org/10.1201/9781003043980
Neuroradiology 101: What An MRI of the Brain Can Reveal [WWW Document], n.d. URL https://www.myssmi.com/blog/neuroradiology-101-what-an-mri-of-the-brain-can-reveal (accessed 8.18.24).

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D., 2018. Attention U-Net: Learning Where to Look for the Pancreas. https://doi.org/10.48550/arXiv.1804.03999

Özdemir, B., Reski, R., 2021. Automated and semi-automated enhancement, segmentation and tracing of cytoskeletal networks in microscopic images: A review. Computational and Structural Biotechnology Journal 19, 2106–2120. https://doi.org/10.1016/j.csbj.2021.04.019

Papers with Code - Max Pooling Explained [WWW Document], n.d. URL https://paperswithcode.com/method/max-pooling (accessed 9.1.24).

Pereira, S., Pinto, A., Alves, V., Silva, C.A., 2016. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Trans. Med. Imaging 35, 1240–1251. https://doi.org/10.1109/TMI.2016.2538465

Prastawa, M., 2004. A brain tumor segmentation framework based on outlier detection*1. Medical Image Analysis 8, 275–283. https://doi.org/10.1016/j.media.2004.06.007

Rao, V., Sharifi, M., Jaiswal, A., 2015. Brain tumor segmentation with deep learning. MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS) 56–59.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.48550/arXiv.1505.04597

Salehi, S.S.M., Erdogmus, D., Gholipour, A., 2017. Tversky loss function for image segmentation using 3D fully convolutional deep networks. https://doi.org/10.48550/ARXIV.1706.05721

Shen, D., Wu, G., Suk, H.-I., 2017. Deep Learning in Medical Image Analysis. Annu. Rev. Biomed. Eng. 19, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Shen, H., Wang, R., Zhang, J., McKenna, S.J., 2017. Boundary-Aware Fully Convolutional Network for Brain Tumor Segmentation, in: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 433–441. https://doi.org/10.1007/978-3-319-66185-8_49

Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J., 2017. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. https://doi.org/10.48550/ARXIV.1707.03237

Wang, L., Lee, C.-Y., Tu, Z., Lazebnik, S., 2015. Training Deeper Convolutional Networks with Deep Supervision. https://doi.org/10.48550/ARXIV.1505.02496

Wels, M., Carneiro, G., Aplas, A., Huber, M., Hornegger, J., Comaniciu, D., 2008. A Discriminative Model-Constrained Graph Cuts Approach to Fully Automated Pediatric Brain Tumor Segmentation in 3-D MRI, in: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 67–75. https://doi.org/10.1007/978-3-540-85988-8_9

What is a Glioma? - Spiegato, 2024. URL https://spiegato.com/en/what-is-a-glioma (accessed 8.30.24).

What Is Instance Segmentation? [2024 Guide & Tutorial] [WWW Document], n.d. URL https://www.v7labs.com/blog/instance-segmentation-guide, https://www.v7labs.com/blog/instance-segmentation-guide (accessed 9.2.24).

Yi, J., Wu, P., Jiang, M., Huang, Q., Hoeppner, D.J., Metaxas, D.N., 2019. Attentive neural cell instance segmentation. Medical Image Analysis 55, 228–240. https://doi.org/10.1016/j.media.2019.05.004

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., Fan, Y., 2018. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Medical Image Analysis 43, 98–111. https://doi.org/10.1016/j.media.2017.10.002

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. https://doi.org/10.48550/arXiv.1807.10165

Zikic, D., Ioannou, Y., Brown, M., Criminisi, A., 2014. Segmentation of Brain Tumor Tissues with Convolutional Neural Networks.

GitLab repository

The code for this project is stored in a Jupyter notebook titled **segmentation.ipynb**, which is located in the root directory of the GitLab repository at https://git.cs.bham.ac.uk/projects-2023-24/vxb330.

This notebook includes all aspects of data loading, preprocessing, model development, training, evaluation, and analysis. To run the project, start by executing the setup cells to import necessary libraries and dependencies, and the helper functions. Then, proceed to execute the cells in sequence to reproduce the results. Keep in mind that the experiments require substantial GPU resources – the notebook has been executed using a 22.5 GB L4 GPU using Google Colab Pro services. Make sure to run the cells in the correct order, as later experiments rely on earlier code.