ALGERIAN DEMOCRATIC AND POPULAR REPUBLIC
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
KASDI MERBAH UNIVERSITY OUARGLA
FACULTY OF NEW TECHNOLOGIES OF INFORMATION AND COMMUNICATION
DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

ACADEMIC MASTER thesis
Domain: Computer science and Information technology
Field: Computer science
Speciality: Artificial Intelligence and Data Science
by Ilhem LAOUAMER & khaoula DRID

THEME

U-Net based deep architecture for brain tumor segmentation

Publicly discussed on: 18/06/2023
Before the Jury:

Pr.	**Mohammed el amine Abderrahim**	**President**	**UKM Ouargla**
Dr.	**Oussama Aiadi**	**Supervisor**	**UKM Ouargla**
Dr.	**Bachir Said**	**Examiner**	**UKM Ouargla**

Academic year: 2022/2023

Acknowledgment

First we would thank the one above all of us, the omnipresent GOD, for answering our prayers and for giving us the strength to plod thank you so much GOD.

We express our sincere gratitude to our supervisor, Dr. Oussama Aiadi, for the continuous support, sage advice, insightful criticisms, and encouragement. He always motivates us to solve our research problems. His guidance helped us in all the time of research and writing of our thesis. And special thanks for his confidence in us.

Likewise, we extend our respectful thanks to the jury members for accepting to review our work we are honored to have you and we thank you before for your remarks.

We sincerely thank all our teachers who taught us through our course. Also for giving us the opportunity to show our capacities, thanks to all.

There are no words to express our gratitude and thanks to our Beloved parents, for supporting us in our daily lives, for guiding us always, and for their prosperity and love.

A special thanks to our family and friends for always standing by us. Their love has been the major spiritual support in our lives.

As a final word, we would like to thank each and every individual who has been a source of support and encouragement and helped us to achieve our goal and complete our dissertation work successfully.

Dedication

Briefly to my MOM

Yours
Khaoula

Praise be to Allah. Alhamdulillah for everything after a study process that has brought with it many difficulties, hardship, and fatigue,

I am honored to dedicate this graduation to my parents, the two individuals who have provided me with the necessary tools and values to reach where I stand today. My parents have supported me unconditionally and taught me the importance of working hard to achieve my aspirations. I especially dedicate this to my dear mother, who taught me how to hold the pen and write words without regret. May God bless her with a long and prosperous life.

I extend my heartfelt dedication to my teachers, who have gone above and beyond to educate and support us. I am grateful for your unwavering commitment. Furthermore, I would like to express my deep appreciation to my supervisor, Dr. Oussama Aiad, for the guidance and mentorship throughout this journey. your expertise, encouragement, and valuable insights have played a pivotal role in shaping the direction of this research. Thank you for your unwavering commitment and support.

Lastly, I dedicate this dissertation to my numerous friends who have stood by me throughout the process. This work is dedicated to all of you

Yours
Ilhem

ABSTRACT

Deep learning has achieved very high and significant results in many interesting fields, the medical imaging field is one of these active areas and it is advancing each day. In this work, we are mainly interested in brain tumor segmentation. Glioma is one of the most common brain tumors, it is divided according to its grade. MRI images are relevant and commonly used on a wide scale by scientists in the diagnostic of LGG images this is what makes it recommended for efficient results. Accurate brain tumor diagnosis and the ability to identify size, location, and shape are very important to save patients' lives.

In view of the impressive performance of U-Net architecture in brain tumor segmentation, as revealed by several literature studies, we propose in this work a U-Net-based architecture for brain tumor segmentation. In particular, we considered an ensemble learning scheme in which three pre-trained networks are incorporated to achieve the final decision. These networks are MobileNet, DeepLabV3+, ResNet, and DenseNet we use them as an encoder part with the U-Net architecture then ensemble learning is applied in many ways to get the best result. However, our methodology could be well generalized as well as could be investigated by so many other architectures and methods. Generally, our obtained results were promising for IOU, Dice-coeff, and accuracy we achieved 0.86, 0.92, and 0.99 respectively thus our followed method improves the importance of applying deep learning in the brain tumor segmentation domain.

Keywords: *Image segmentation, Brain tumor, Pre-trained models, U-Net, CNN.*

ملخص

حقق التعلم العميق نتائج عالية وهامة للغاية في العديد من المجالات المثيرة للاهتمام ،كمجال التصويرالطبي الذي هو في تقدم كل يوم. في هذا العمل ، نحن مهتمون بتجزئة ورم الدماغ، اهمها الورم الدبقي بحيث يعتبر أكثرها شيوعًاوتعتبر صور الرنين المغناطيسي الاكثر استخداما من قبل العلماء بحيث التشخيص الدقيق لأورام الدماغ مهم جدًا لإنقاذ حياة المرضى. في ضوء الأداء المذهل لهندسة U-Net نقترح في هذا العمل بنية تعتمد عليها . على وجه الخصوص درسنا مخطط التعلم الجماعي الذي فيه يتم دمج اربعة شبكات مدربة مسبقًا و هي MobileNet ، DeepLabV3+ ، ResNet و DenseNet بحيث نستخدمها كجزء مشفر مع بنية U-Net ثم يتم تطبيق التعلم الجماعي بعدة طرق للحصول على أفضل نتيجة. بشكل عام ،كانت النتائج التي حصلنا عليها واعدة بالنسبة إلى IOU ومعامل النرد والدقة التي حققناها 0.86 و 0.92 و 0.99 على التوالي ومنه ، منهجيتنا يمكن تعميمها بشكل جيد وكذلك يمكن التطوير و التحسين فيها من قبل العديد من البنى والطرق الأخرى.

الكلمات المفتاحية: تجزئة الصورة ، ورم الدماغ ، نماذج مدربة مسبقا ، U-Net ، CNN.

RÉSUMÉ

L'apprentissage en profondeur a obtenu des résultats trés élevés dans de nombreux domaines intéressants comme le domaine de l'imagerie médicale. Dans ce travail, nous intéressons á la segmentation des tumeurs cérébrales exactementle Glioma. Les images IRM sont couramment utilisées á grande échelle par les scientifiques dans le diagnostic des images LGG, c'est ce qui les rend recommandées pour des résultats efficaces. Un diagnostic précis des tumeurs et la capacité d'identifier la taille, l'emplacement et la forme sont trés importants pour sauver la vie des patients. Compte tenu des performances impressionnantes de l'architecture U-Net dans la segmentation des tumeurs cérébrales, comme révélé par plusieurs études de la littérature, nous proposons dans ce travail une architecture basée sur U-Net. En particulier, nous avons considéré un schéma d'apprentissage d'ensemble dans lequel trois des réseaux préformés sont incorporés pour parvenir á la décision finale. Ces réseaux sont MobileNet, DeepLabV3+, ResNet et DenseNet nous les utilisons comme partie codeur avec l'architecture U-Net ensuite, l'apprentissage d'ensemble est appliqué de plusieurs maniéres pour obtenir le meilleur résultat. En général, nos résultats obtenus étaient prometteurs pour IOU, Dice-coeff et la précision que nous avons obtenue 0,86, 0,92 et 0,99 respectivement, ainsi notre méthode suivie améliore l'importance d'appliquer apprentissage profond dans le domaine de la segmentation des tumeurs cérébrales.

Keywords: *Segmentation d'image, Tumeur cérébrale, modéles Pré-entinés, U-Net, CNN.*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION

1.1 Introduction

1.1.1 Background and Motivation

Brain tumors are a significant health concern, contributing to a high global death rate. They are characterized by the growth of abnormal cells within the human brain. Among the various types of brain tumors, gliomas are the most common malignant ones. Gliomas can be further classified into high-grade gliomas (HGG) and low-grade gliomas (LGG)[137]. The World Health Organization (WHO) classifies brain tumors by cell origin and behavior, from least to most aggressive type[72]. The HGGs are particularly threatening, often limiting the life expectancy of patients to a maximum of two years. On the other hand, LGGs provide a more favorable prognosis, potentially allowing sufferers to live for many years [75].

1.1.2 Research Problem

Segmentation of brain tumors from neuroimaging modalities is a crucial step in enhancing disease diagnosis [37]. However, accurate segmentation of brain tumors poses significant challenges due to their diverse characteristics, including variable locations, varying shapes and sizes, and poor contrast [7] this results in overlap with the intensity levels of healthy brain tissues [38]. Therefore, it is not easy to distinguish healthy tissue from the tumor. A common approach to address this challenge is to integrate information obtained from various modalities of Magnetic Resonance Imaging (MRI) such as T1-weighted MRI (T1), T1-weighted MRI with contrast (T1c), T2-weighted MRI (T2), Fluid-Attenuated Inversion Recovery (FLAIR) MRI [37]. MRI is typically preferred for structural brain analysis as it offers images with high soft tissue contrast and spatial resolution, without posing

any known health risks [3]. Additionally, MRI offers an in-depth scan that can ease to spot brain tumors and other infections [7].

Brain tumor segmentation methods can be broadly categorized into three groups based on the level of human interaction they require: manual, semi-automatic, and fully automatic segmentation [69], Manual segmentation of brain tumors requires the laborious task of manually delineating the tumor boundaries and relevant structures, often accomplished through drawing or painting [38] making it a time-consuming and subject to rater variability [37]. Therefore, during the past two decades, there have been hundreds of different algorithms developed for the reliable automatic and semi-automatic segmentation of brain tumors[69]. However, automated segmentation of brain tumors from multimodal MR images has become essential for the evaluation and control of the progression of the disease. This is due to the persistent challenges in accurately segmenting tumors which are primarily caused by the malignant and heterogeneous characteristics of gliomas. Furthermore, the presence of ambiguous and fuzzy boundaries between cancerous tissue and normal brain tissue adds to the complexity of the segmentation task [127].

1.1.3 Significance and Scope

In this context, Significant efforts have been dedicated to the development of classical machine learning for the segmentation of normal (e.g. white matter and gray matter) and abnormal brain tissues (e.g., brain tumors) in MRI [3]. Deep learning algorithms excel at tasks such as semantic segmentation due to the ability to learn and extract relevant features automatically, eliminating the need for hand-crafted feature engineering. By training on large amounts of data, deep learning algorithms demonstrate impressive results and generalizability [3] while traditional machine learning algorithms often strug-

gle with generalization and it needs to creation relies on carefully engineered imaging features and domain-specific expertise. To address these limitations, the medical imaging research community has turned to a promising machine learning technique with deep learning [65]. Various types of deep learning approaches have been developed for different purposes, including object detection and segmentation in images, speech recognition, and genotype/phenotype detection. In the context of image segmentation and disease classification, convolutional neural networks (CNNs) are widely applied and recognized for their effectiveness. In particular, U-Net has emerged as one of the top choices for segmentation, especially in the context of brain tumor segmentation. Its widespread usage in this domain serves as evidence of its effectiveness and success.

Comparing the performance of different brain tumor segmentation methods and improving the development and evaluation of algorithms has proven challenging due to variations in input data, types of brain tumors, and clinical stages [37]. As a result, A multimodal Brain Tumor Image Segmentation Benchmark (BraTs) challenge in conjunction with the International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI) focused to assess the state-of-the-art in automated brain tumor segmentation and compare the different proposed methods [77]. This initiative aimed to establish a standardized dataset that facilitates fair comparisons among different segmentation methods in the field of brain tumor segmentation.

1.1.4 Objectives and Aims

The aim of this thesis is to develop and implement a deep learning scheme for automatic segmentation of brain tumors. The performance of different network architectures will be evaluated through experiments. To achieve this, we will unravel and implement the

fundamental building blocks of the U-Net structure, with the objective of enhancing accuracy and efficiency in brain tumor segmentation within medical imaging. Additionally, we aim to reduce the time complexity associated with the segmentation process.

Our contribution to the field of medical image segmentation, particularly in brain tumor segmentation, lies in the exploration and extension of the successful U-Net approach. Through our experiments, we have made significant advancements by employing ensemble models. These models have yielded remarkable results, with an Intersection over Union (IoU) of 0.8692, a Dice coefficient of 0.92027, and an accuracy of 0.99901 on the test data. These findings emphasize the effectiveness of our approach in accurately delineating brain tumors, thus contributing to improved diagnostic capabilities and treatment planning in clinical settings.

1.2 Image Segmentation

Image segmentation[80] is a sub-domain of computer vision and the processing of digital images that seek to classify and group related regions or segments of an image[13]. Also, image segmentation is an active technique and it has received great attention from researchers in several areas. Segmentation is defined as the split of an image (or view) into areas that correspond to various surfaces. As a result, the segmentation's borders will be depth and very specified boundaries[81]. Humans and many other organisms adopt a mechanism known as visual perception[114][64][94] to fast determine which areas of an image require detailed processing and which may be disregarded[30]. This enables us to deal with large amounts of visual information while also efficiently utilizing the capabilities of our visual system. Trying to each computer to see and also understand what they

are seeing has proven extremely difficult [33]. For computer vision, researchers have to deal with exactly the same problems, so learning from the behavior of humans provides a promising way to improve existing algorithms for image segmentation[56]. Image segmentation tasks[48][138][76][52] are categorized into three types based on the amount and type of information they provide.

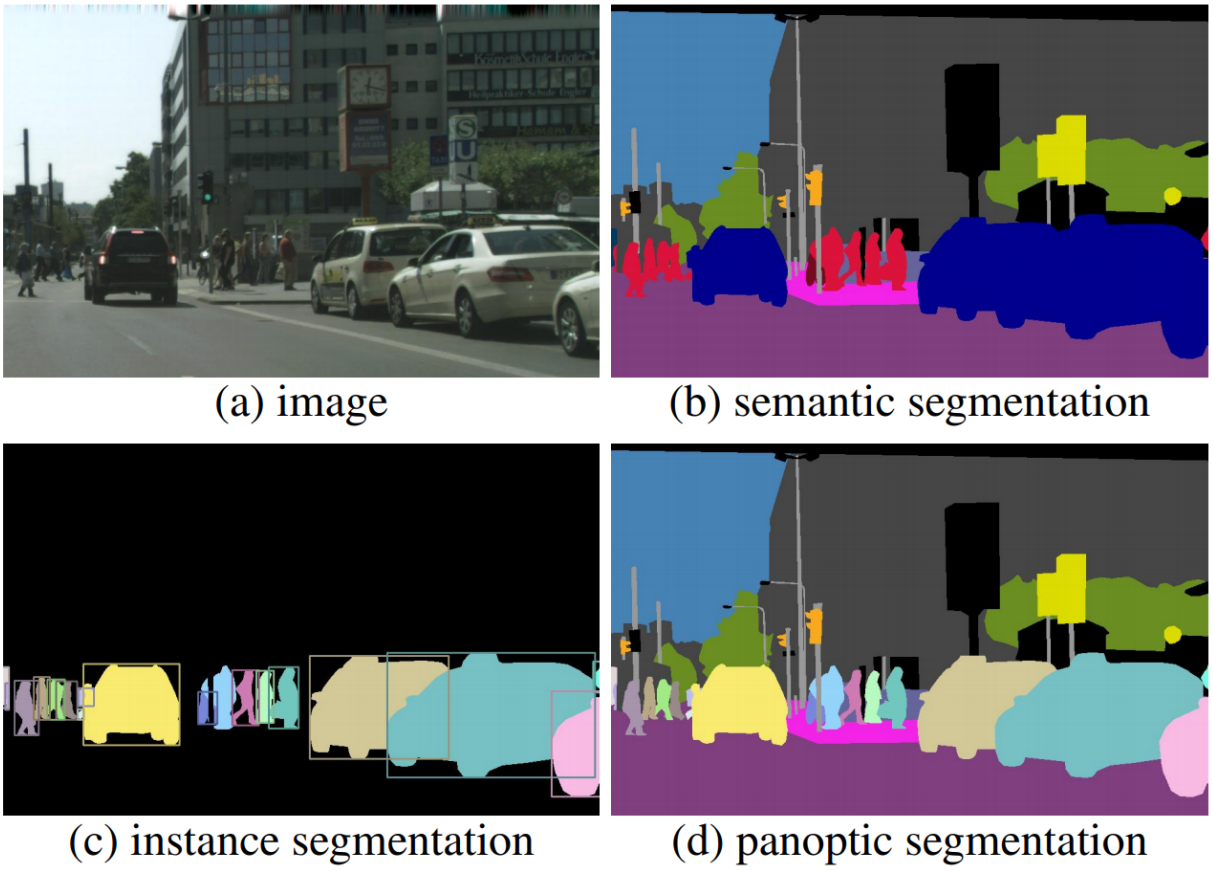

Figure 1.1: The difference between semantic, instance, and panoptic segmentation[59]

1. **Semantic Segmentation:**

   Semantic segmentation is the process of dividing a given image into numerous visually significant or interesting regions for further image analysis and visual understanding[84]. Then grouping together image elements that correspond to the same object class see Figure 1.1.

2. **Instance Segmentation:**

   Instance segmentation has emerged as a rather significant, sophisticated, and difficult field in computer vision research. It localizes different classes of object instances present in distinct images, with the goal of predicting the object class label and the pixel-specific object instance mask. Instance segmentation is primarily designed to help robots, autonomous driving, surveillance, and other similar applications[40]. Several instance segmentation frameworks have been proposed with the discovery of deep learning, mainly convolutional neural networks such as[11][68], Mask R-CNN[43]

3. **Panoptic Segmentation:**

   Panoptic segmentation(PS) combines the traditionally separate tasks of semantic segmentation (assigning a class label to each pixel) and instance segmentation (detecting and segmenting each object instance)[59]. Panoptic segmentation is currently being researched to enable get a more detailed understanding of visual situations[32].

1.3 Applications of image segmentation

Image segmentation has received great attention from researchers in several areas and many interesting applications are using image segmentation[13]. Some of the most important applications are the following:

1.3.1 Medical imaging:

Medical imaging includes a variety of technologies used to observe the human body in order to diagnose, monitor, or treat medical conditions[34]. In medical imaging, image

segmentation is commonly used to recognize and separate certain anatomical features or diseases see Figure1.2. And here we have some medical images :

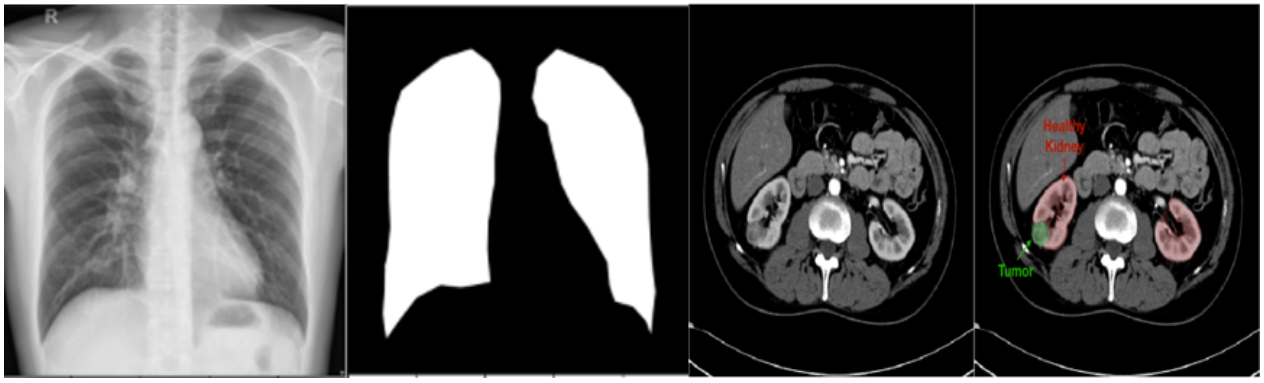

Figure 1.2: An example of medical imaging segmentation

- X-Ray segmentation.
- CT scan organ segmentation.
- Dental instance segmentation.
- Digital pathology cell segmentation.
- Surgical video annotation.
- **Brain Tumor Segmentation:** In the segmentation of brain tumors, labels are assigned to different tissues that share similar characteristics. The process involves dividing the brain into tumor and non-tumor parts. Within the tumor tissues that exhibit the same characteristics can be further divided into subcategories, with each subcategory assigned its own label as illustrated in Figure 1.3.

  These labels are assigned to each MRI slice and are used for training and evaluating the network. It is crucial that these labels are accurately assigned in order to obtain a reliable segmentation network. To ensure the quality of the labeling, multiple

experts are often involved in the data labeling process. Their expertise helps minimize errors and inconsistencies, resulting in more accurate and reliable segmentation models[63].

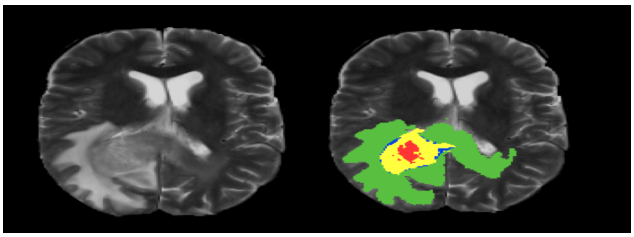

Figure 1.3: An example of brain segmentation with Four Distinct Regions Labeled

1.3.2 Robotics

In robotics, image segmentation is used to recognize and detect objects in a scene. It may be used to segment and monitor the human body[131] image segmentation enhances machine perception and locomotion by highlighting objects in their direction of motion, allowing them to change directions effectively and comprehend their surroundings.[13]

1.3.3 Self Driving Cars

Autonomous vehicles feature a number of sensor systems aboard to identify obstructions, lanes, and available parking spaces, among other things[89][79]. Image segmentation, which employs camera pictures to identify each pixel, is a commonly used approach in this discipline. The expected visuals can be utilized to plan the behavior of the vehicle and avoid collisions[17]. For more understanding see Figure 1.4

Figure 1.4: An example self driving car[101]

1.3.4 Agriculture

The correct segmentation of crops and weeds in agronomic images has traditionally been the focus of precision agriculture. Several approaches have been developed, but clean and crisp segmentation of crops and weeds remains a difficult issue for images with a significant weed presence[118]. This aids in the detection of weeds at an early stage, their eradication, the exact use of herbicides, and their management. While removing soils is a simple procedure, detecting and separating weeds from crops has always been difficult[73].

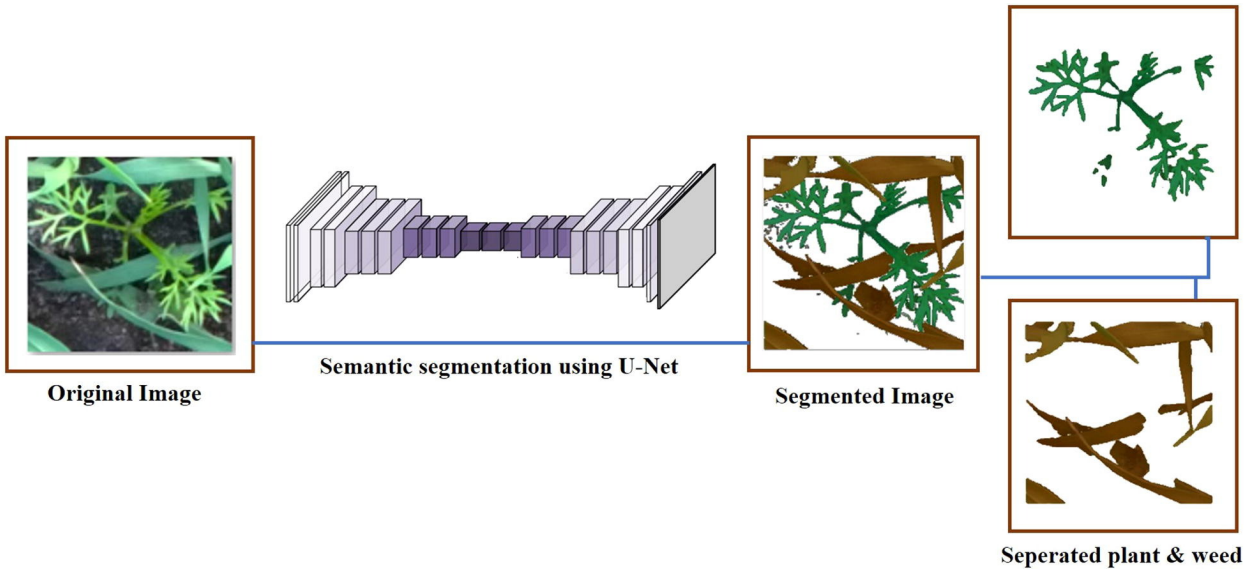

Figure 1.5: An example image segmentation in Agriculture[118]

1.3.5 Smart Cities

One of the major and challenging study fields is the concept of Smart Cities[111]. Smart Cities often have CCTV cameras for real-time monitoring of pedestrians, traffic, and crime. With the use of image segmentation, this monitoring may be simply automated[13]. Also, crimes can be reported more quickly with AI-based surveillance, traffic accidents can be responded to with quick ambulances, and speeding vehicles can be simply detected and penalized. That's why the use of image segmentation monitoring flexibility and stability enhances people's lifestyles. Some of the most useful applications in smart cities are:

- Pedestrian detection
- Traffic analytics
- License plate detection
- Video Surveillance

1.4 Brain tumor segmentation based on MR images

Medical image segmentation is important in clinical diagnosis, it is a challenging problem because medical images frequently have poor variations, multiple types of noise, and missing or diffusive boundaries [24]. The brain's structure can be scanned by Magnetic Resonance Imaging (MRI) scan[20] [14][9]or Computed Tomography (CT) scan[4]. For diagnosis, an MRI scan is more satisfying than a CT scan because it does not use radiation. it has no effect on the human body it is based on the magnetic field and radio waves [93]. MRI scans provide high-quality images that are particularly useful for brain tumor segmentation. Brain tumor segmentation from MR images has the potential to significantly impact diagnostics, growth rate prediction, and treatment planning.

A widely approved imaging protocol for brain tumor MRI acquisition includes the utilization of these four modalities [5], namely T2, T1, T1c, and FLAIR see Figure 1.6, which are distinct forms of MRI. T2 provides detailed images of brain tissue, T1 highlights the anatomical structures, T1c enhances the contrast between tumor and normal tissue after the administration of contrast agent, and FLAIR suppresses the signal from cerebrospinal fluid, enabling better visualization of abnormalities. These modalities have gained significant recognition and are particularly prevalent in the field of brain tumor segmentation.

1.5 Structure of thesis

The rest of this thesis is structured as follows:

In chapter 2, we will take a look at the state of the art in the image segmentation domain and will review the most important algorithms for image segmentation based on U-Net.

In chapter 3, we will present our methodology, we will start by explaining the base

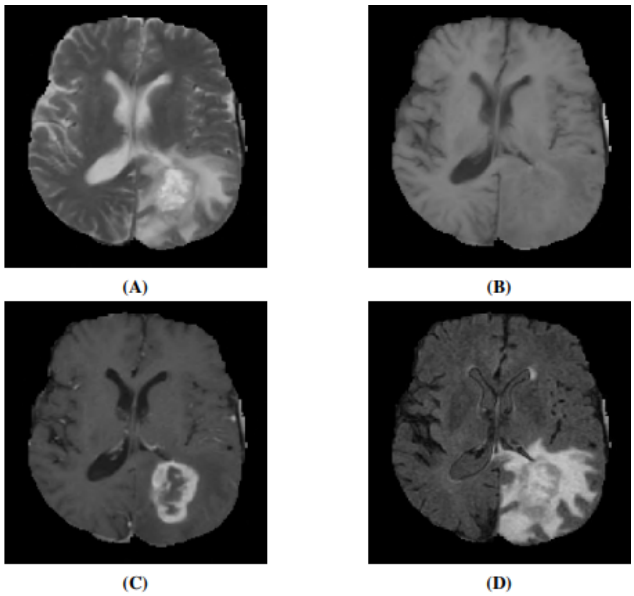

Figure 1.6: MRI Images (A) T2 (B) T1 (C) T1c (D) FLAIR [12]

method in detail and mentioning its limitations, then we explain our methodology and how each step works deeply.

In chapter 4, we will describe the experiment steps and environment also the used evaluation metrics, then we discussed the obtained results.

In chapter 5, the General conclusion is where we draw the conclusion of the thesis, as it illustrates the main outcome of it and what more can be achieved in future works and research.

CHAPTER 2

AN OVERVIEW ON THE STATE OF THE ART

To get more knowledge about the primary basics of medical image segmentation in this chapter we are going to discuss the brain tumor segmentation techniques and steps. In the first section, we will present brain medical images basically the MRI images. Next in the second section, we will explain the most common brain tumor segmentation methods and techniques. After that in the third section, we will explain and compare various U-Net architectures and variants. Finally, the conclusion of this chapter.

2.1 Methods of Brain Tumor Segmentation

The principal purpose of image segmentation is to divide an image into similar areas based on a predefined standard. When it comes to brain tumor segmentation, the tumor is divided into segments by separating the various tumor tissues, such as edema, necrosis, and active tumor from normal brain tissue such as cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM)[85].

For brain tumor segmentation, typical segmentation methods are applied to get the objective measure to define the homogeneity of each tissue. These methods can be divided into several categories as shown in Figure 2.1

2.1.1 Hybrid techniques

The hybrid technique is one of the most known concepts, it is based on combining and collecting two or more models in a way that their strengths are combined one to each another to get better result performance [8]. In the segmentation domain, it is widely used for getting better segmentation and for more productive and reliable results. As a special case, the hybrid technique is used for MRI brain tumor segmentation and it achieves very

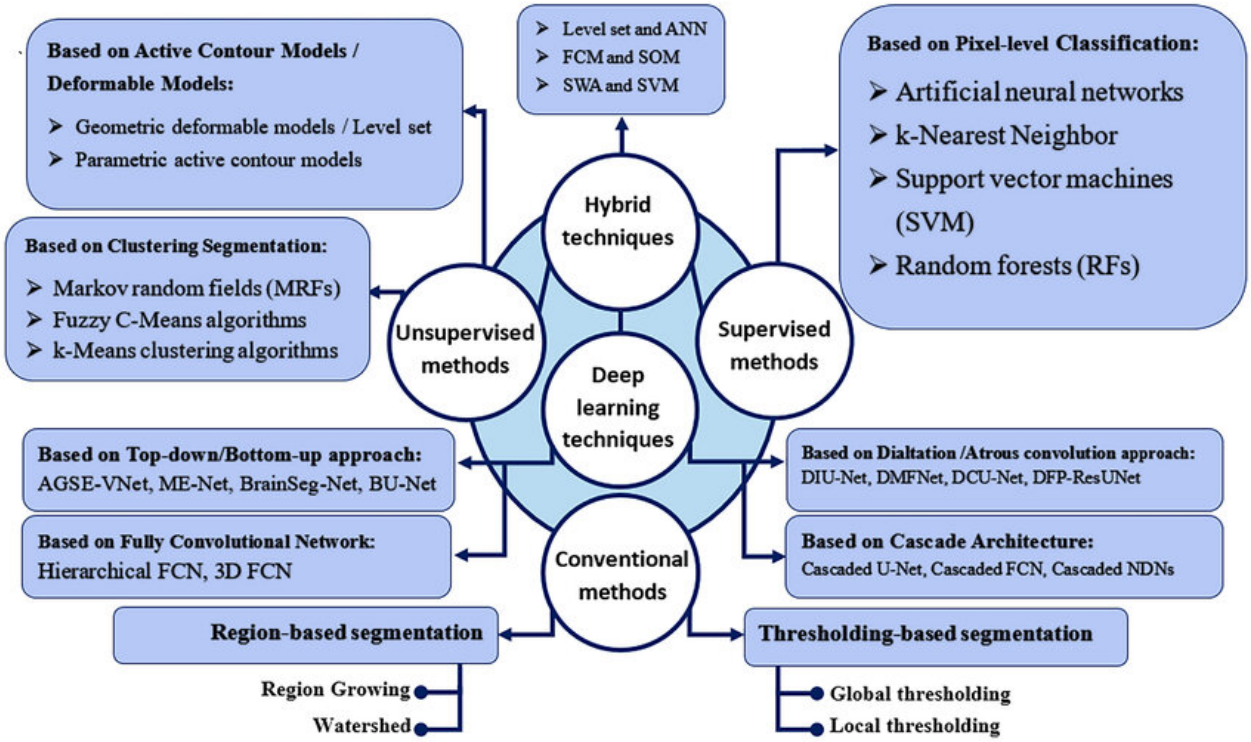

Figure 2.1: General methods and techniques for the segmentation of brain tumor images. [85]

good results and still required technique till now [67].

2.1.2 Supervised techniques

Supervised learning is a well-known category of machine learning. It is based on the use of labeled or named data in the training phase. there are a lot of commonly used supervised methods in the literature such that artificial neural networks(ANN), K nearest neighbors(kNN) [39], random forests(RF)[6], support vector machines(SVM)[27], and the superpixel-based segmentation which is a functional method to eliminate small regions by grouping similar pixels for a quick calculation. The simple linear iterative clustering (SLIC) method is used as a grouping mechanism. But it was not widely used with CNNs[2]. The superpixel is computationally efficient according to the sampling of image complexity and a large number of pixels combination. Researchers are working on brain

tumor segmentation as a supervised segmentation problem they use labels to identify normal and abnormal tumors[85]. But there are some key limitations like pre-processing of data and computation time are so big and others are shown in Table 2.1.

2.1.3 Unsupervised techniques

Unsupervised learning is used when the data is unlabeled. This mechanism is used to solve many problems and cover the miss data identification problem. There are many algorithms under unsupervised learning we will explore some of them.

1. **Clustering-based segmentation :** Image segmentation can be interpreted as a clustering approach in which pixels that satisfy a characteristic are grouped into a cluster while pixels that do not satisfy the set of criteria are placed in different groups[83]. The approach of grouping the data is generally termed clustering which is based on collecting the pixels which have similar features. The clustering technique can be divided into two main subcategories, Hard clustering, and soft clustering. Here we will take some examples of interesting clustering algorithms.

    - **FCM clustering :** Fuzzy C-Means is one of the most famous clustering techniques. For The Brain tumor segmentation domain, it is considered as one of the base models. Each pixel in the image is classified into all classes by assigning a membership value for each one corresponding to each class center[88] based on measuring the distance value between the classes center and the pixel value, which means that each pixel unit can be assigned to one or more classes. The suppleness of the FCM technique is another advantage when working with data that has numerous cluster solutions[85]. Also, the FCM clustering algo-

rithm has produced positive and encouraging results. Via the segmentation of tumors into numerous tissue classes using various membership classes. It provides comparatively better outcomes with overlapped pixels than k-means clustering.

- **k-Means clustering :** The k-Means clustering algorithm is an unsupervised algorithm for segmenting regions of interest from the background [86]. The theory of K-Means is based on partitioning a data set into k clusters by computing the distance between each point and the entire k centroid and later assigning it to the cluster with the minimum length. Also, this is a very expensive calculation that can lead to inaccurate and poor results, mainly if the value of k is selected inaccurately[29].

- **Markov random field :** The MRF also known as the Markov network or underacted graphical model is a group of random variables that is a useful method for incorporating spatial information, context, intensity, texture, and spectral qualities into the clustering process[85]. So The MRF algorithm is widely used in image segmentation as a probabilistic approach to reduce the likely problem of overlapping and the noise effect on the clustering result [122]. The MRF offers a straightforward and efficient method for representing spatial interdependence in image pixels. It is used to simulate how adjacent pixels are related to one another. The region is strongly characterized as either a brain tumor or a non-tumor in the case of brain tumor segmentation, and then MRF assesses whether the neighboring pixels have similar properties and belong to the same region (tumor or non-tumor) [85]. Also, Conditional random fields(CRFs) have been presented as a method to develop probabilistic models to segment and

label sequence data [62]. The MRF and CRF algorithms can express complex connections among data instances, providing high precision in the segmentation of brain tumors[66]. These methods are therefore increasingly being used in a variety of segmentation tasks[85].

2. **Active contour models/deformable models :** is one of the most recently applied segmentation techniques it is based on separating the pixel of interest in the image to do further operations. This method achieves great success in extracting the boundaries of brain tumors from 2D MRI images. The majority of 3D image segmentation methods use model-based segmentation approaches [19].

2.1.4 Conventional techniques

The conventional segmentation mechanism is based on thresholding, contour-based methods, and edge detection filters[31]. This technique is usually used for 2D image segmentation. For the tumor region's boundaries, they might be drawn with more accuracy. It consists of two common image-processing methods, mainly thresholding-based and region-based techniques. The conventional technique always asks for enough processing time and can not work well on 3D images.

1. **Thresholding-based technique** Its segmentation concept is based on intensity comparison it is a simple method to use an efficient way of segmenting data using one or more intensity thresholds[85]. Using some functions it converts the image from grayscale to a bi-level background and the object of interest. It is a very important method in the image segmentation domain Also, it is divided into two main subcategories and techniques global thresholding and local thresholding.

- **Global thresholding** When the intensity distribution between foreground and background objects is highly distinct, the global (single) thresholding approach is applied. When the distinctions between these items are highly distinct, a single threshold value can be used to separate them [8]. Consequently, in this type of thresholding, the value of threshold T is exclusively determined by the pixel attribute and the image's grey level value[57]. The critical problem with the threshold method is that it is based on the intensity rather than the relationship between the pixels. As a result, it might not work when the intensity levels of two or more tissue structures overlap [85].
- **Local thresholding** This method works a local segmentation for every region in the image by generating a small window and applying a local value of threshold that it means that it used multilevel thresholding. The full image multi-target segmentation and local threshold selection may both be accomplished using local threshold segmentation[36]. A single threshold cannot provide better segmentation results for any chosen image. Local thresholding algorithms can provide appropriate thresholds for local image areas including brightness, contrast, and texture[133].

2. **Region-based techniques** In comparison to other approaches, region-based segmentation is straightforward and noise-resistant. It separates an image into areas according to predefined criteria, such as color, intensity, or object[58]. Region-based segmentation methods are categorized as follows: region expanding, region splitting, and region merging[54]. One of the most interesting region-based techniques is the watershed technique, also known as the watershed transform, which is a mathematical morphology-based picture segmentation strategy[124]. The watershed is a

grayscale dedicated image[97].

2.1.5 Deep learning techniques

The importance of deep learning is increasing day by day and it is applied in many interesting domains and many new ones. In the medical image domain, especially the brain tumor segmentation widely using deep learning approaches and methods based on CNNs. Many important CNN-based methods have recently been developed which aim to develop accurate segmentation approaches that enhance the task of defining tumor locations. In this part, we will discuss some important architectural CNN-based model details.

1. **Based on fully convolutional network approach :** The fully convolutional network (FCN) was presented by[71], it is an architecture used in semantic segmentation where they shift the classifiers from basic CNNs to dense FCN layers by exchanging fully connected layers (dense layers). FCN tries to learn representations based on only local information, but it ignores the image's overall semantic context which produces segmentation quite unclear[85]. Consequently, the basic model VGG-16 [116] helped the FCN model to achieve effective results by skipping connections for fusing the downsampling path used for feature extractions and an upsampling path used for localization features[85].

2. **Based on cascaded CNN approach :** A cascaded architecture is one of the newest deep learning methods for image segmentation, its idea is based on the collection of some CNNs then the result of each model is combined with the other layers that it self concatenate a new CNN on the output of the previous network [107]and to

produce the final result. Cascaded architectures, also known, as are typically more expensive computationally[55]. This method is proposed to get better results from models and improve efficiency and accuracy.

3. **Based on the top-down/bottom-up approach:** In this approach, the base idea is the uses of encoder-decoder architectures [125]. The main idea in this approach is that for the segmentation process, we have two paths. The first path operates for feature extraction for the input image and where the image size will decrease throughout successive pooling and convolutional layers [128] so the feature map size will decrease. In the other part, the network will act the upsampling operations in an opposite manner from the first path increasing the image's dimensions using deconvolution layers to produce the localization of the target class and segments. So far, many essential architectures have been proposed for general medical segmentation based on a top-down/bottom-up approach, such as the U-Net architecture [102]. This architecture we will talk about in detail in the next section.

2.2 U-Net Network

The U-Net architecture is a U-shaped encoder-decoder network, the abstract representation of the U-Net contains four encoder-decoder blocks connected by the bridge[102].In the encoder part, the number of channels is increased and the spatial information will be reduced. The opposite is in the other part where it reduces the number of filters and increases the spatial dimensions of the feature maps, the idea is represented in Figure 2.2[121]. In the next section, we will explain well the different parts of the U-Net architecture.

Methods	Strengths	Limitations
Conventional-based segmentation methods		
Thresholding (global and local)	Simple execution whit fast computation time.	These techniques are ineffective for all types of brain MRI images due to the high-intensity variance in image intensity.
Region growing Watershed	Important for any type of segmentation and useful for image linear. Simple and capable of separating the regions that have similar attributes and high performance in the complicated regions. The perfect technique for grouping similar pixels in an image based on their intensity Divides multiple regions at the same time.	Limited applicability to improving different tumor regions, and lower performance in heterogeneous regions. Affected by noise and require user intervention May produce over-segmentation or holes due to noise or difference in intensity, and requires a post-processing step. The problem with the watershed is its sensitivity to intensity variations when the image is segmented into several large regions. Over-segmentation.
ML-based segmentation methods		
Artificial neural network Random forest	ANN is highly dependent on complex and multi-variate non-linear domains. RF is flexible in a classification problem and uses a rule-based technique. Helps to improve decision tree accuracy by reducing overfitting.	Collecting training samples is difficult, and the learning phases are generally slow. It requires more time for training and is computationally intensive
K-nearest neighbor	k-NN is easy to understand, very time-efficient, and easy to implement as the only thing that needs to be calculated is the distance between different points.	Since all the work is done at run-time, k-NN can have poor performance with a large dataset, and the computation would be very expensive. Sensitive to noise, and the value k affects the performance of the algorithm.
Support vector machine	It is more effective for training generalizable non-linear classifiers in high-dimensional spaces SVM reduces the number of misclassifications	Not suitable for large datasets because it takes a long time to train. Not suitable when the dataset has more noise
Superpixel-based segmentation	An efficient approach to minimize the number of image primitives subsequently required for further processing.	The standard convolution operation used on conventional grids is inefficient when applied to superpixels.
Fuzzy c-means (FCM) clustering	Less complex technique. Determines the degree of membership of the data for each category. It can converge the boundaries of tumors.	Not entirely appropriate for situations where regions do not have clearly specified boundaries. Relatively slow and sensitive to noise.
Markov random fields	may describe complicated relationships between data instances. More effective against noise.	It needs computationally expensive.

Hybrid techniques		
Hybrid techniques	Hybrid techniques are aimed at combining the benefits of many models into a single approach. More adaptable and flexible when dealing with high-dimensional data.	It requires a high and sophisticated computational effort. It depends on the combination of feature selections from different methods.
Deep learning techniques		
Deep learning techniques	High performance for intra-tumor segmentation. In deep learning techniques, problems are often solved on an end-to-end basis, whereas traditional methods are not able to do so.	Unreliable for real-time applications and often requires a post-processing step. Data labeling can be an expensive and time-consuming operation.

Table 2.1: Strengths and Limitations of Brain tumor segmentation methods[85].

1. **Encoder path** The encoder path is also known as the contracting path in this part the network must produce the feature map by the application of successive convolutions[39] each convolution is 3x3 in size, and it is followed by a ReLU activation function[102]. The output of the ReLU is called skip connection it is responsible for providing additional information or location information to the corresponding decoder block to obtain better results and improve the quality of the final segmentation[121]. In addition to the convolution, downsampling by 2x2 max-pooling is applied that's what makes the feature map dimensions reduced by half[102]. In general, this mechanism reduces the number of trainable parameters[121].

2. **Bridge** This part is responsible for the connection of the encoder and the decoder parts[120]. It consists of two 3x3 convolutions in addition to a ReLU activation function after each one [102].

3. **Decoder path** Also called the expanding path in this part the result obtained by the

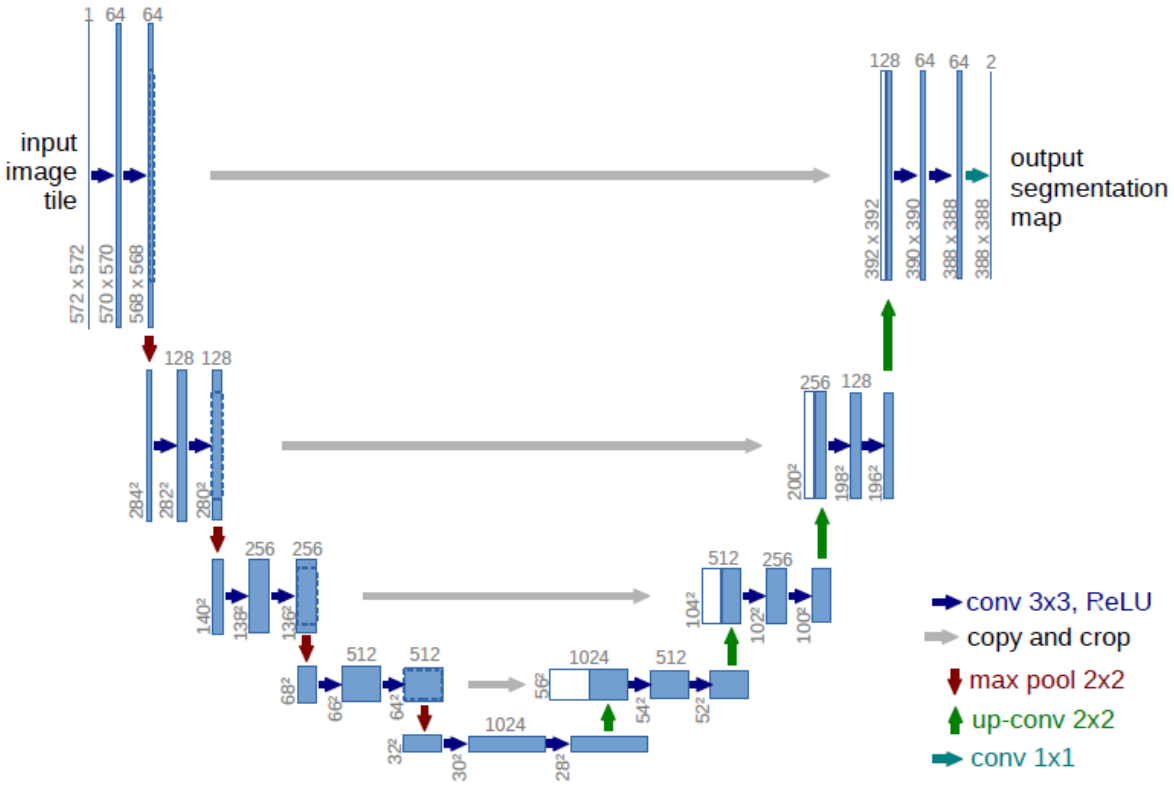

Figure 2.2: The U-Net architecture[102]

encoder part will be passed here to obtain the segmentation mask [102]. Starting with 2x2 transpose then the obtained skip connection feature map will be concatenated with each layer of this block [130]. In the end, a 1x1 convolution is applied to the output of the final layer, in addition to the sigmoid activation function which provides the segmentation mask of the pixel-wise classification[121].

4. **Skip Connections :** The skip connection in the U-Net architecture is really important, so it skips local information from the encoder part to the layers in the decoder part in a manner that the information will not be lost. It means that it supports the decoder part to achieve better results[130]. Moreover, the skip connections improve the network performance to get better result representation and help in the conver-

gence in a quick way[102].

U-Net has a contracting path to capture features and a symmetrical expanding path to enable precise information localization[70]. The concept of skip connections is one benefit of adopting U-Net versus traditional FCN[102]. where it directly concatenates the feature maps between the two paths and passes feature maps from the contracting path to the expanding path. The original image data through skip connections can help the layers in the contracting path recover details [70]. Several research works have been proposed on the U-Net architecture some of them will be discussed in the next section.

2.2.1 U-Net variants

What makes U-Net so valuable and mostly used for segmentation, is its ability to generate very detailed segmentation maps from extremely small trade samples. Because of its context-based learning, U-Net is also faster to train than most other segmentation models[113]. Given that the potential of U-Net keeps increasing, this section will look at the many advances in U-Net architecture and we will explain privily some U-Net variants.

1. **V-Net**

   The majority of medical data that is used in clinical practice is in 3D volumes such as MRI volumes, but most techniques can only analyze 2D images[115]. The V-Net is suggested to segment 3D images using a volumetric fully convolutional neural network and it could be used for 2D ones also the main difference in this architecture is that convolution layers are used instead of pooling layers. On their proposition[78], they based on the idea that max pooling will enforce the model to lose information so they change it to convolutions without padding to save more information. In this section, we will explain the V-Net architecture as it is represented in Figure 2.3.

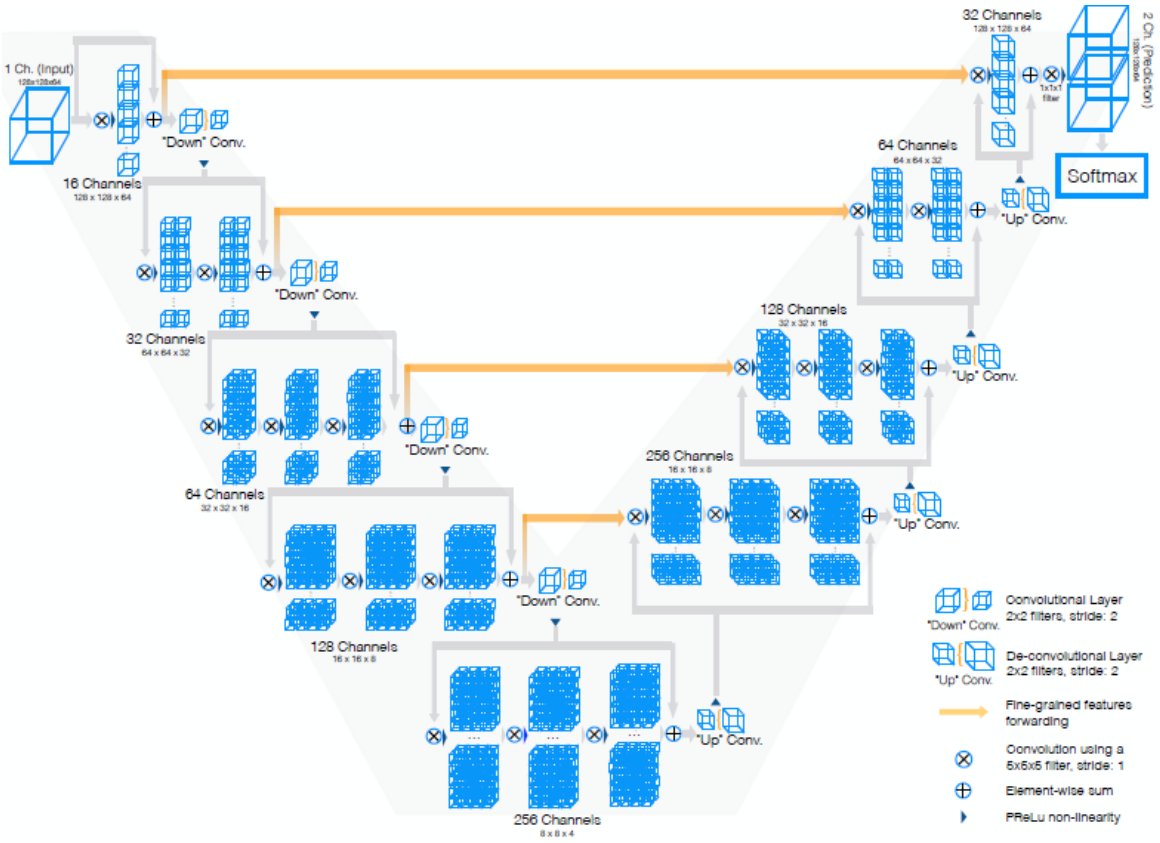

Figure 2.3: The V-Net architecture [78]

i. **The left side:**This part is divided into different levels each level is based on the running of one to three convolution layers[115]. Those convolutions used volumetric kernels of 5x5x5 voxel size [78]. Then from one level to another, the resolution of the input image is reduced by half using a convolution of 2x2x2 kernel with stride equal to two[18].

ii. **The right side:** The network uses feature extraction and increases the spatial support of the lower-resolution feature maps in order to gather and put together the necessary information to output a two-channel volumetric segmentation. The residual function is learned, similar to the left part of the network[78]. A deconvolution operation is adopted in order to increase

the size of the inputs at each stage, including half the number of 5x5x5 kernels used in the previous layer[115]. Finally, in the last convolutional layer, the last two feature maps are calculated using a 1x1x1 filter to produce an image of the same input image size and those feature maps will represent the probabilistic segmentation using the softmax function.

2. **Residual U-Net**

   The residual U-Net(ResU-Net) is a very interesting architecture so it is based on the benefit from the performance of residual networks[45]. It was founded with the goal of overcoming the challenges associated with training vastly deep neural networks such as overfitting and vanishing gradient [46]. At each block in ResU-Net, the input of each layer is connected to the output of the later layer using a skip connection. By the application of residual skip connection before downsampling and upsampling, the network will be allowed to better preserve feature maps in deeper neural networks as well as improved performance[140]. Each residual unit is represented by the following calculations let F be the function that represents layer i to layer i + n. Denote the input for layer i by x first, The residual connection operates identity mapping to x, then it accomplishes element-wise addition F(x) + x see Figure 2.4.

3. **Attention U-Net**

   The attention mechanism was introduced by Bahdanau et al. (2014)[10]. It is looking for solutions to two major questions: What to look for and where to look[119], to improve the performance of the encoder-decoder model for machine translation[25]. In the context of image segmentation, attention is the way of highlighting only the

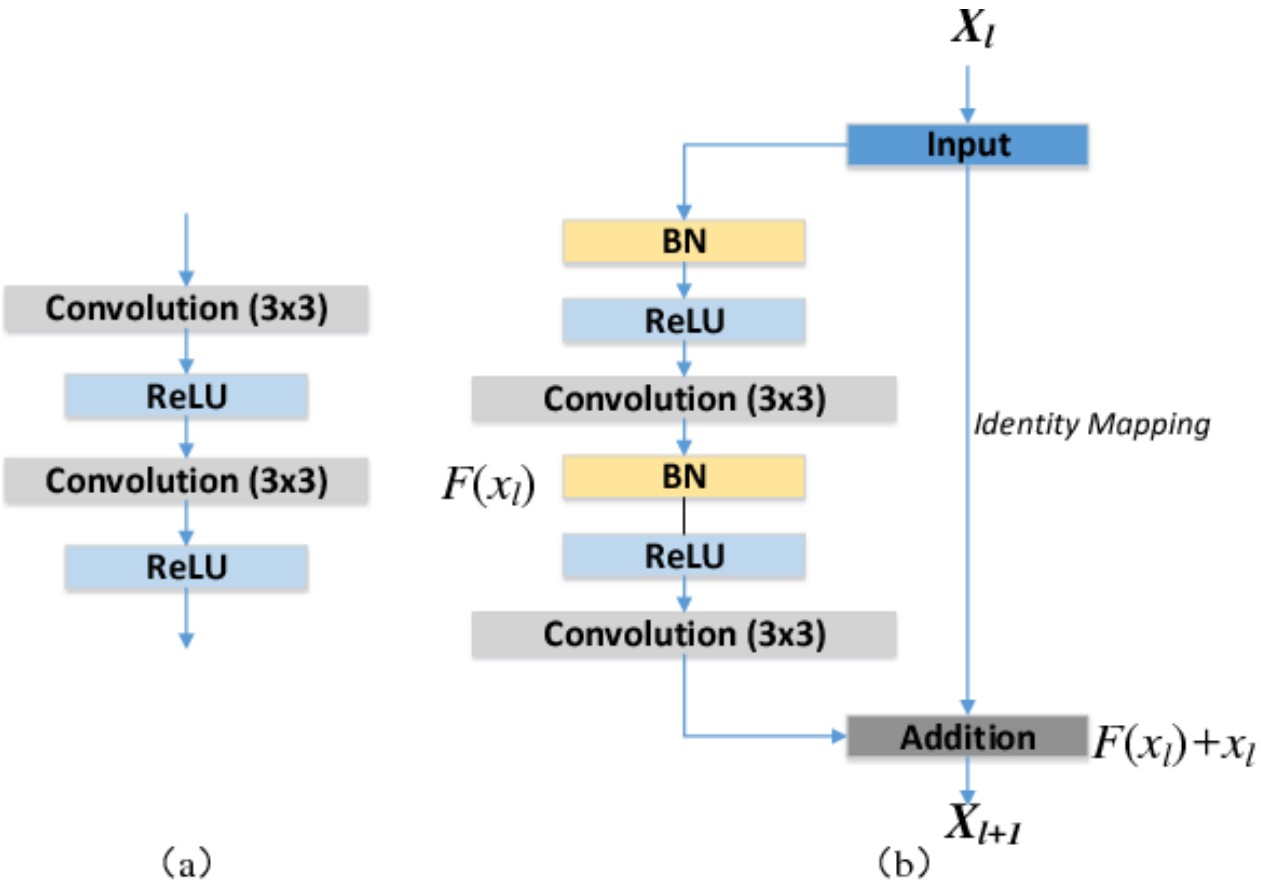

Figure 2.4: (a) Plain neural unit used in U-Net and (b) residual unit with identity mapping used in the proposed ResU-Net[140].

relevant activations during training or the network can pay attention to certain parts of the image[119]. This is what makes the computational resources reduced and is not wasted on such irrelevant activations, also it provides the network with better generalization efficiency[126]. The attention mechanism comes in two variants, hard[132] and soft[74] attention.

i. **Hard attention:** It bases on highlighting relevant regions by splitting the image or by iterative region proposal. Hard attention is used as the probability of a region getting selected and it can only choose one region of an image at a time[109].

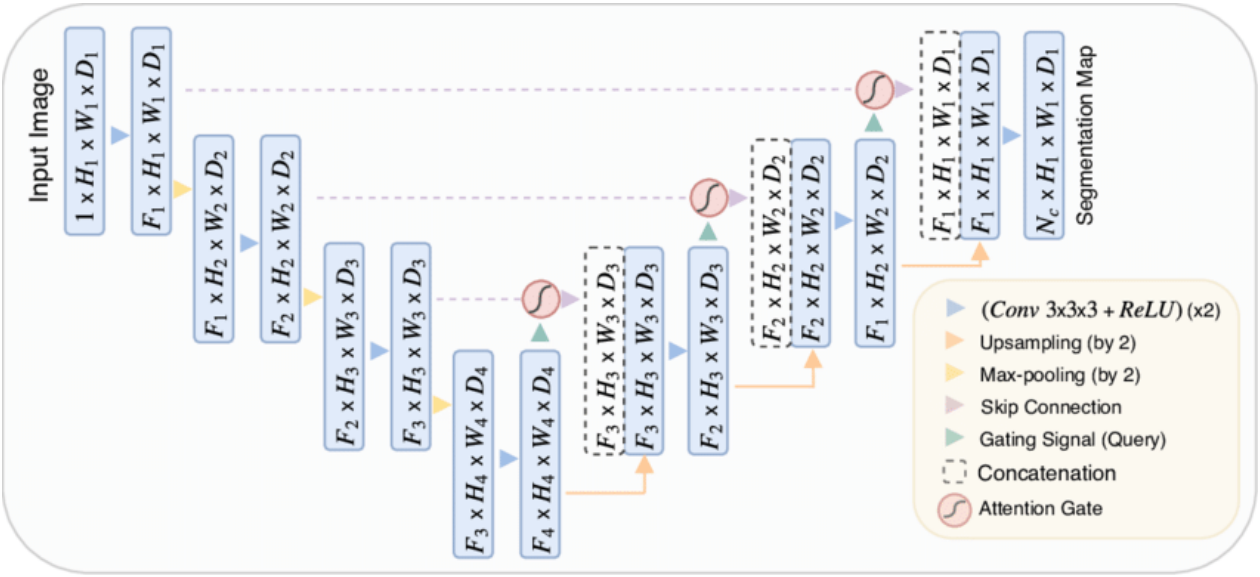

Figure 2.5: A block diagram of the Attention U-Net segmentation model[92].

ii. **Soft Attention:** For Soft attention, the attention score is used as a weight in the weighted average context vector calculation[28]. It works by weighing different regions of the image. regions with high relevance are multiplied with a larger weight and regions of low relevance are multiplied with smaller weights. More focus is given to the regions with higher weights as the model is trained[61].

A desirable treatment in an image processing network is the capability to focus on specific objects and areas with high importance while ignoring unnecessary areas. The attention U-Net achieves this by making use of the attention gate [92]. The attention gate has defined as a unit that learns to focus on target structures of varying shapes and sizes and trim features that are not relevant to the ongoing task[108]. The attention gate is associated with each layer in the expansive path through which the extracted features from the contracting path must pass before they are concatenated with the upsampled features in the expansive path as it is shown in Figure 2.5[82].

This successive reputation of the use of an attention gate after each layer will improve the segmentation performance significantly and without the model becoming overly computationally complex[112]. The network learns to focus on the desired region as training proceeds. The differentiable nature of the attention gate allows it to be trained during backpropagation, which means the attention coefficients get better at highlighting relevant regions[126]. The Attention U-Net has outperformed a plain U-Net in the overall Dice Coefficient Score. While the Attention U-Net has more parameters, the inference time is only marginally longer[92].

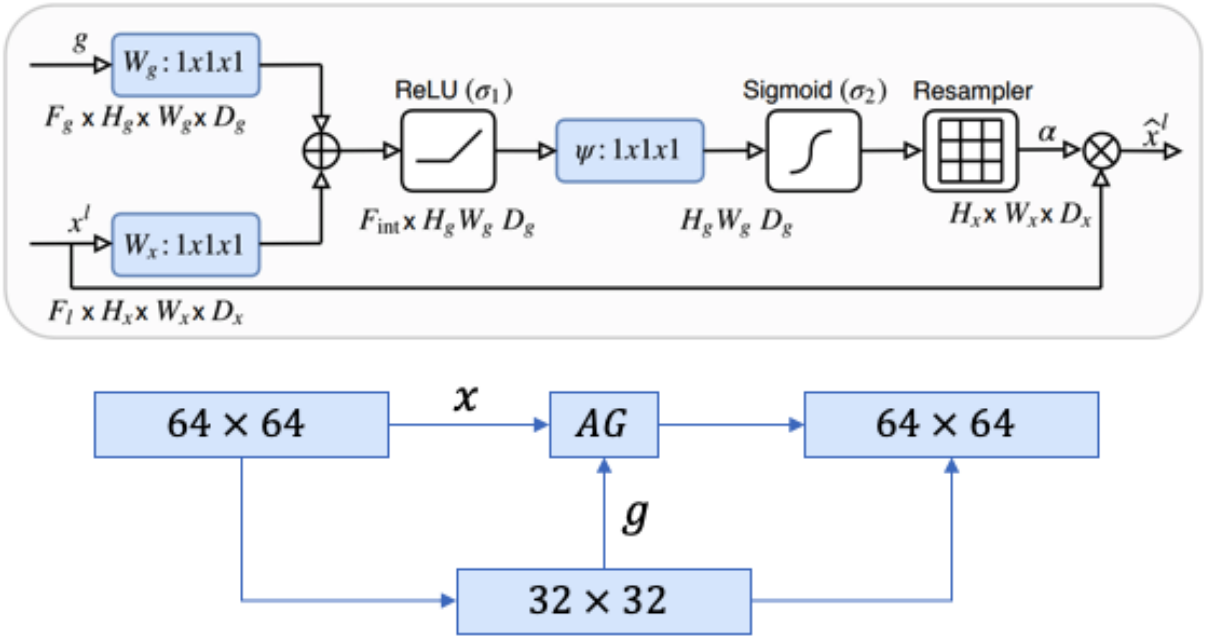

Figure 2.6: Schematic of the proposed additive attention gate (AG)[92].

The way the attention gate operates[98] is as follows, see Figure 2.6 Vectors x and g are the two inputs that the attention gate receives. The vector g, yields from the next lowest layer of the network. Because it comes from deeper in the network, the vector has smaller dimensions and better feature representation. In the example in

Figure 2.6, The dimensions of vector x would be 64x64x64 (filters x height x width) and vector g would be 32x32x32. The Vector x is convolved with striding so that its dimensions become 64x32x32 and vector g goes through a 1x1 convolution such that its dimensions become 64x32x32. The two vectors are summed element-wise, as a result of this process, aligned weights get larger while unaligned weights become relatively smaller. The resulting vector is subjected to a ReLU activation layer and a 1x1 convolution, which reduces the dimensions to 1x32x32. This vector is passed by a sigmoid layer, which scales it between [0,1]. producing the attention coefficients (weights), where coefficients closer to 1 indicate more relevant features. Using trilinear interpolation, the attention coefficients are upsampled to the original dimensions (64x64) of the x vector. The attention coefficients are multiplied element-wise to the original x vector, scaling the vector based on relevance. This is then passed along normally in the skip connection.

4. **Dense U-Net**

   The idea behind this architecture is that the Dense U-Nets employ DenseNet blocks instead of regular layers[112]. The most important characteristic of Dense U-Net is its ability to reduce the number of parameters and it is more efficient in parameter usage[51] also the vanishing-gradient problem[47] will be solved. Dense U-Net increases feature propagation and encourages feature reuse, the feature maps of all preceding layers are used as inputs and its own feature maps are used as inputs in all subsequent layers. This connectivity pattern facilitates feature reuse, strengthens information flow, enhances the performance of the network, and enables the network to learn discriminative features effectively as demonstrated by Figure 2.7.

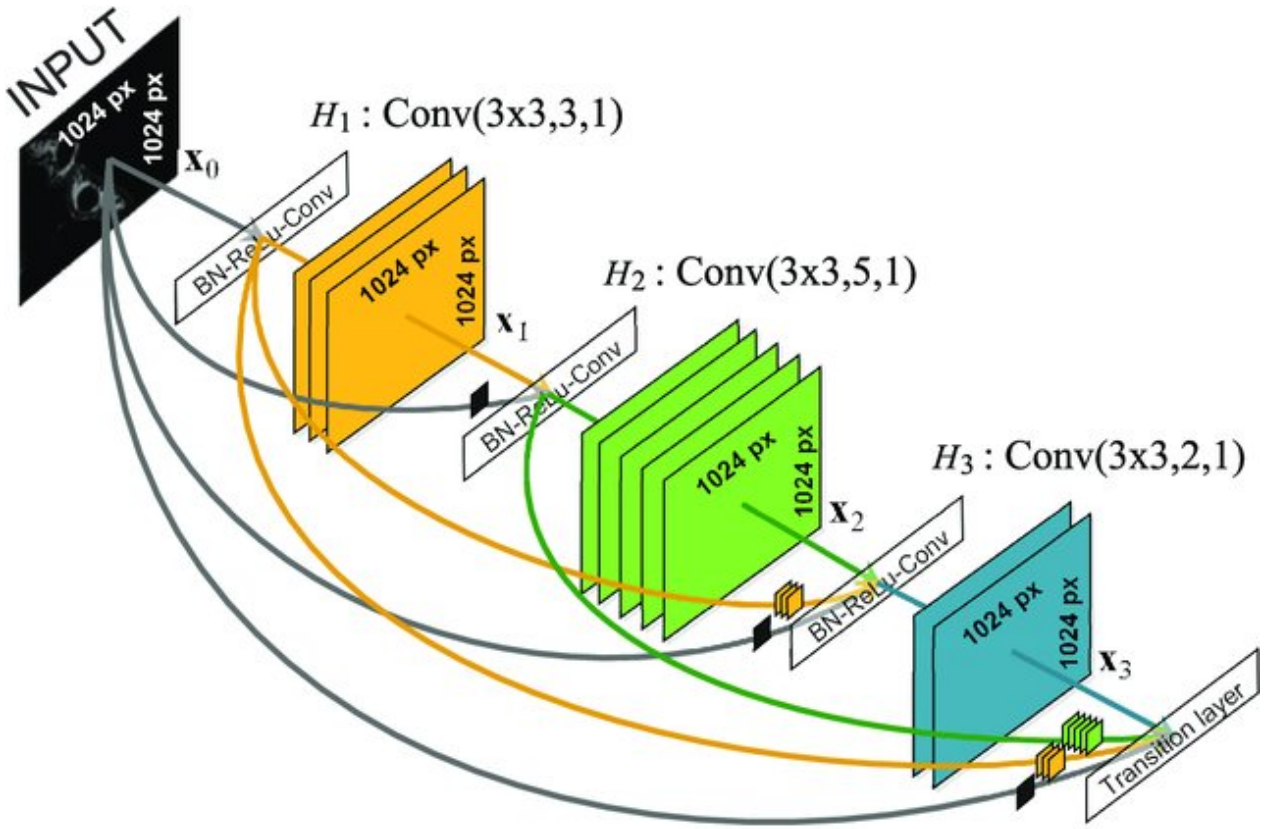

Figure 2.7: Illustration of dense connections in DenseNet [117]

5. **Attention Gate Residual U-Net**

   At the deeper step of downsampling, the U-Net network has a better feature representation capability as it learns more context information[102]. The process of continuously cascading output maps might cause lost spatial information and location specifics. To solve this problem they use attention gates that enable them to pay attention to small-scale regions and gather more information about them [92]. Thus The upsampling procedure is helpful in restoring the location information of small-scale regions. Additionally, to attention gates they add residual modules into U-Net architecture to pursue more promising segmentation performance which, when downsampled, may extract more precise dense feature information and when up-

sampled can effectively recover spatial information and position features[90]. The contracting path contains three residuals each one has two convolutional layers [137], and the PReLU[44] as an activation function instead of the ReLU function employed in the original U-Net architecture. The expanding path is made up of three residual blocks and three attention-gate units to improve the salient feature information.

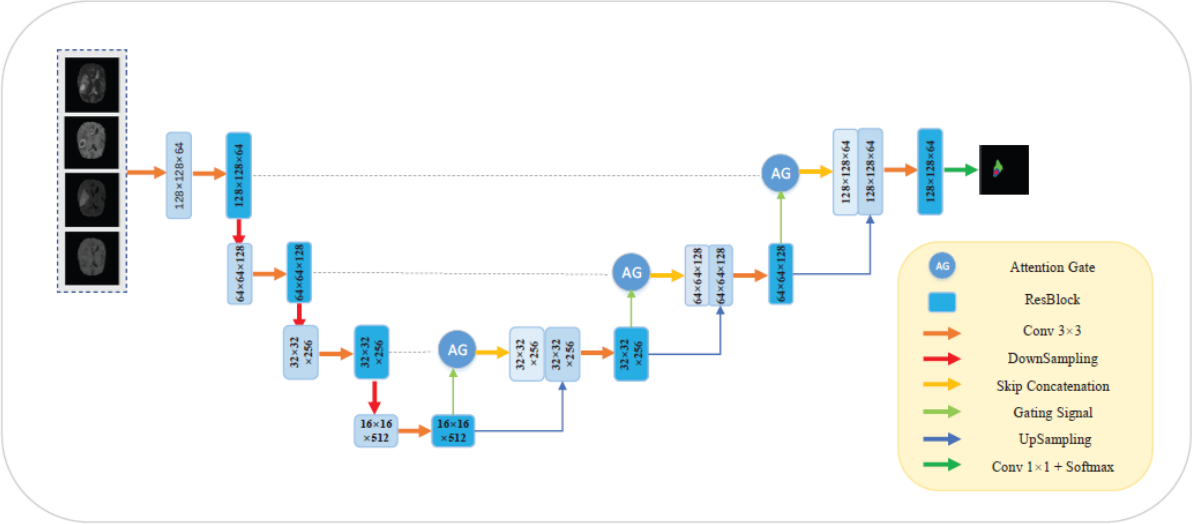

Figure 2.8: The end-to-end network architecture of Attention Gate Residual U-Net [137].

6. MobileNet U-Net

MobileNet is one of the most famous CNN models that is widely used in a lot of real-world applications. It was developed by Google scientists in 2017, they design a simple and easy architecture that is suitable to work on mobile devices which have a limited computation resource [106]. The idea of MobileNet is that it is based on an architecture that uses depthwise separable convolutions to reduce the complexity of computation while convolution operations and to build lightweight deep neural

networks[49]. The difference is that the convolutions apply the kernel on the whole image which produces a huge number of parameters while the depthwise separable uses another optimistic method and it is considered a key building block for many efficient neural network architectures[139]. The depthwise mechanism is a collection of two layers the depthwise convolution, and the pointwise convolution. As with so many other architectures, MobileNet is also combined with U-Net to get an effective network architecture[91] and computational power preserving. It works as the feature extractor in the encoder part[105] then the U-Net completes the process by upsampling to decode the feature maps.

As we see in this chapter the image segmentation domain is a really active one, especially in the applications of brain tumors. However, it is still an open problem due to the complexity of brain images as well as the limitation of computational resources, which provides valuable insights and guidelines for future progress, research, and studies. In the next chapter, we will go into more detail about very interesting methods which will be the base of our methodology.

CHAPTER 3

THE PROPOSED METHOD

For the image segmentation domain, it exists many important applications which show the importance of it in real life. For that, a lot of methods are used competitively. Thus, it varies from one application to another. For that, we chose to present some of the most useful pre-trained architectures such as MobileNet, ResNet, DeepLapV3+, and DenseNet. In this chapter, we will present those architectures in detail. Then we will show some advantages of those architectures and how we can improve the U-Net model segmentation based on them. Finally, we will introduce the ensemble models technique and how we use it in our work.

3.1 Pretrained Architectures

A pre-trained model represents a model or a saved network that has been trained on a big dataset and a group of learned weights to handle a similar problem[41]. Instead of training a network from scratch, it can be used as a starting point. Training deep learning models from scratch can be a time-consuming and computationally expensive process, particularly for large and complex datasets. This is why we have chosen to focus on pre-trained models.

3.1.1 Implementation details of the Dense U-Net

In The Dense U-Net architecture presented in Figure 3.1, we combine the DenseNet121 model as its backbone with an encoder section so the DenseNet blocks are used to gradually reduce the number of features by employing pooling layers within the dense blocks. These dense blocks are where each layer within a block receives inputs from all preceding layers refer to figure 2.7. In the decoder module, convolutional transpose layers (con-

vTranspose) are applied to upsample the reduced input image note Figure 3.1. The decoder blocks also incorporate skip connections from the encoder module, which allow the network to utilize low-level spatial information captured in the early stages of the encoding process. By combining the upsampled features from the decoder with the skip connections, the decoder module can reconstruct high-resolution feature maps.

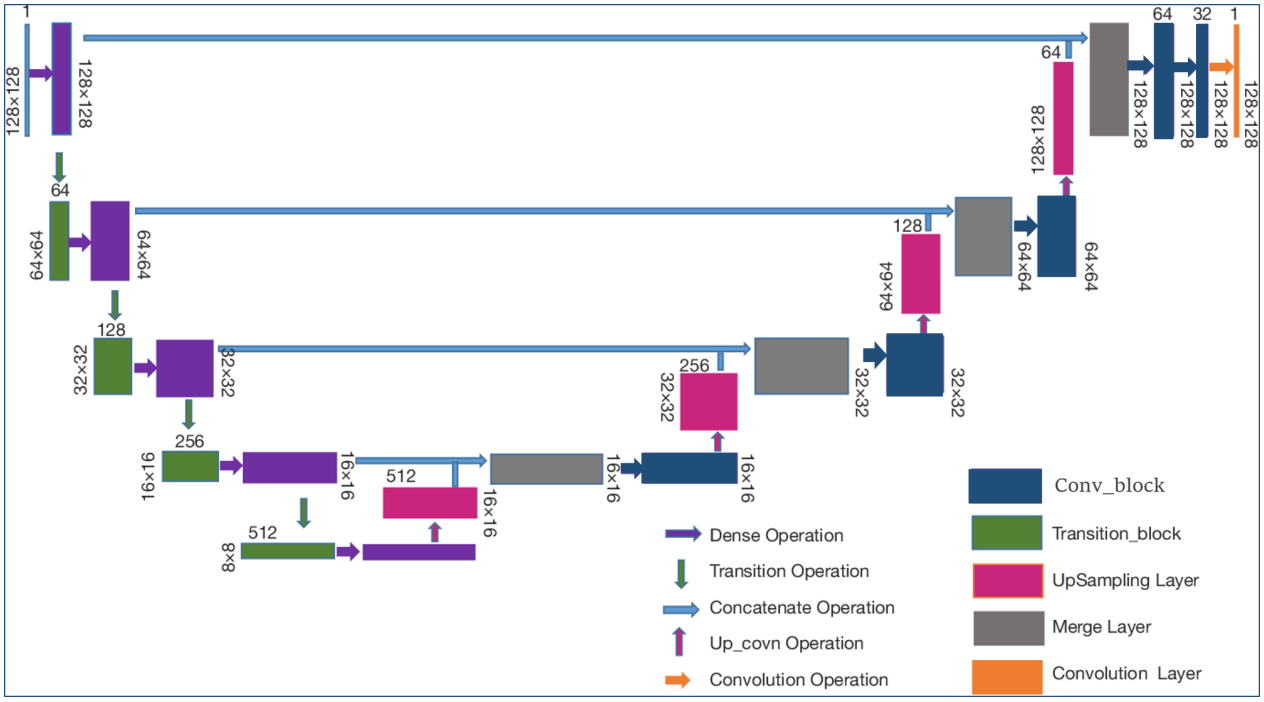

Figure 3.1: Schematic diagram of proposed Dense U-Net architecture.

Finally, a 1x1 convolutional layer is applied to the reconstructed image, which has the same dimensions as the input, to perform the segmentation process. The 1x1 conv layer reduces the number of channels to produce the final segmentation mask. The sigmoid activation function is used in this layer to ensure that the output values are within the range of 0 to 1, representing the probability of each pixel belonging to the target class.

Overall, the Dense U-Net architecture leverages the DenseNet121 model's feature extraction capabilities and incorporates skip connections to preserve spatial information,

allowing for accurate segmentation of objects in the input image.

3.1.2 Implementation details of the Residual U-Net

We combined the U-Net with ResNet blocks by a pre-trained ResNet50 model is instantiated with weights from the ImageNet dataset. The encoder Module of U-Net consists of four residual blocks extracted from different layers of the ResNet50 model. Each feature map captures different levels of abstraction, allowing the model to learn hierarchical representations of the input image as evident from Figure 3.2.

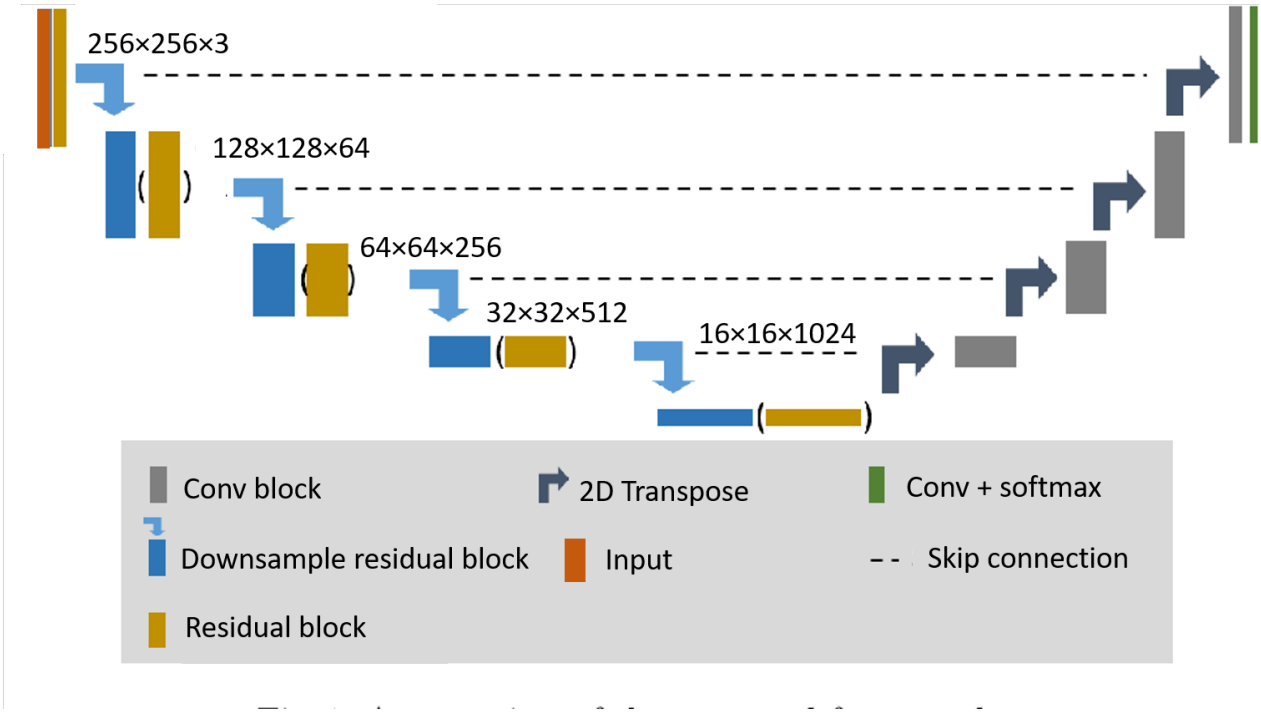

Figure 3.2: The Implemented Residual U-Net Architecture.

Following the encoding module, decoder modules are employed to up-sample the feature maps from the encoder section and concatenate them with the corresponding feature maps from the previous layer This enables the recovery of spatial details that were lost during the encoding process. Additionally, These decoder blocks incorporate a simple

convolutional layer applied at each stage as demonstrated by Figure 3.2, helping to refine the reconstructed image and capture fine-grained features. By combining upsampling, skip connections, and convolutional layers, the model ensures that the reconstructed image retains important spatial information and achieves better segmentation results.

3.1.3 Implementation details of the MobileNetU-Net :

The MobileNetU-Net architecture combines the lightweight MobileNet encoder with the U-Net decoder to create an efficient and effective model for image segmentation tasks see Figure 3.3. This architecture aims to provide accurate segmentation results while minimizing the computational resources required. Our MobileNet encoder utilizes layers from

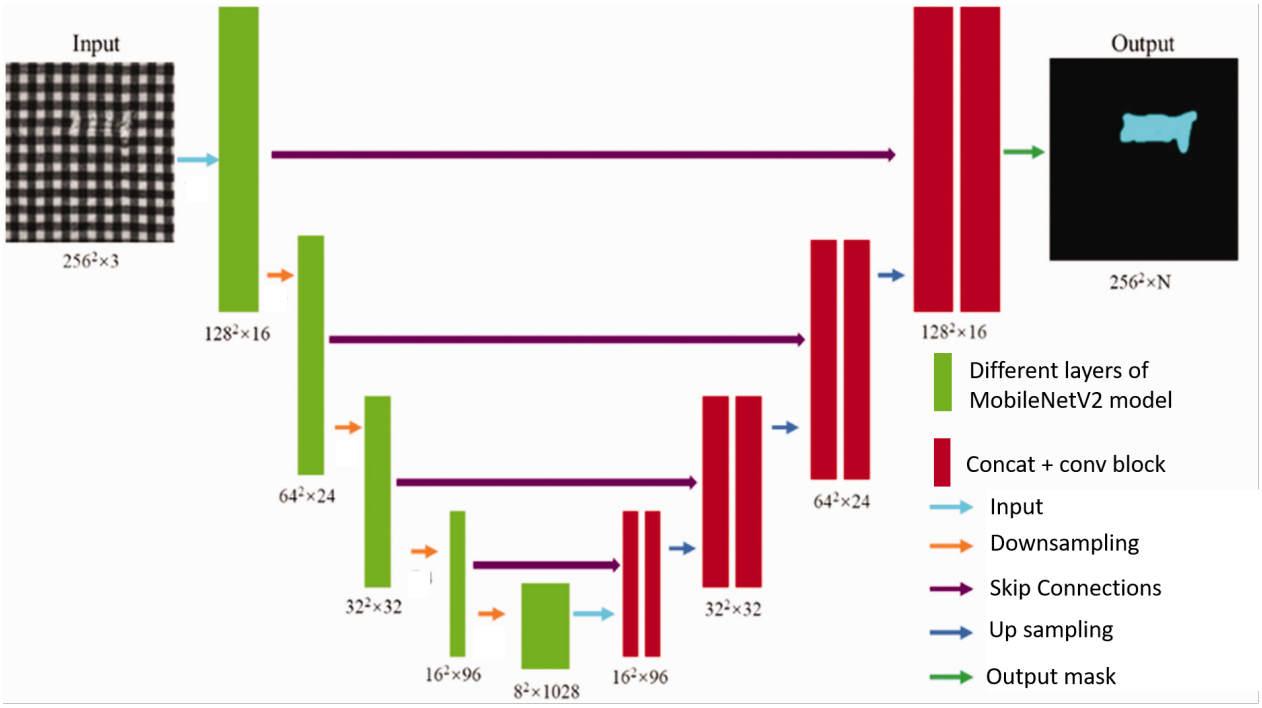

Figure 3.3: The Implemented MobileNet U-Net Architecture

the MobileNetV2 pre-trained model, which employs depthwise separable convolutions. These convolutions separate the spatial and channel-wise operations, leading to a reduc-

tion in the number of parameters and computational cost [50]. As a result, the MobileNet encoder is lightweight reducing the computational complexity and the number of parameters compared to traditional convolutional neural networks.

Depthwise convolution and pointwise convolution are the two layers that make up depthwise Separable Convolution. The depthwise convolution work by applying lightweight filtering for each input channel separately whereas the pointwise convolution is to collect or linearly combine the output of the depthwise for building new features that's why this methodology reduces the number of parameters[26]. In the depthwise convolutional layer, a single filter is applied individually to each input channel, as illustrated in Figure 3.4. This enables the model to efficiently capture spatial information. Following the depthwise convolution, a Pointwise convolutional layer then is applied to the depthwise convolution output. This layer involves a simple 1×1 convolution [1]. Both the depthwise and pointwise convolutions in MobileNets utilize batch normalization and ReLU nonlinearities [50] to enhance performance and introduce nonlinearities into the model as apparent from the Figure 3.6. The purpose of the pointwise convolution is to perform a linear combination [50] of the depthwise output channels which means convolving each pixel of the input as demonstrated by Figure 3.5. This combination of operations significantly improves the effectiveness and efficiency of the MobileNet architecture.

3.1.4 DeepLabV3+

DeepLabV3+ is an enhanced semantic segmentation architecture that extends DeepLabV3, by incorporating an effective decoder module to recover the object boundaries[23]. It combines the strengths of the encoder-decoder architecture and atrous convolution see Figure 3.7 to capture rich contextual information and improve the segmentation perfor-

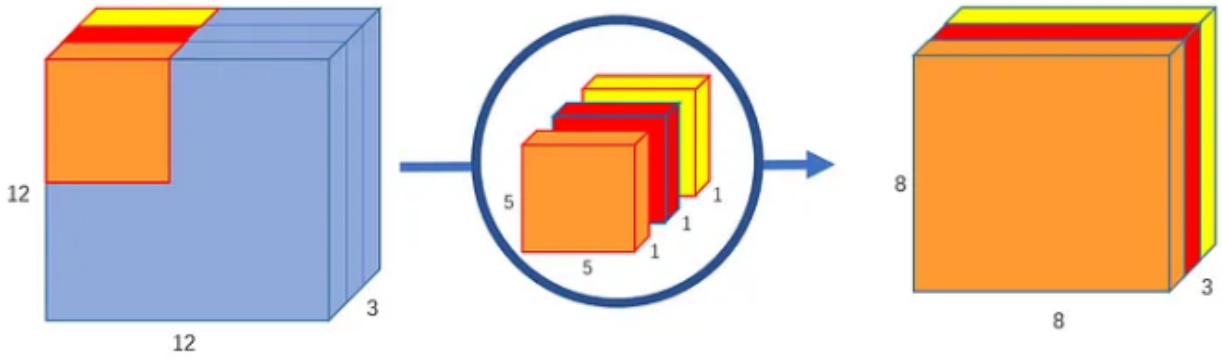

Figure 3.4: Illustration of depthwise Convolution. [1]

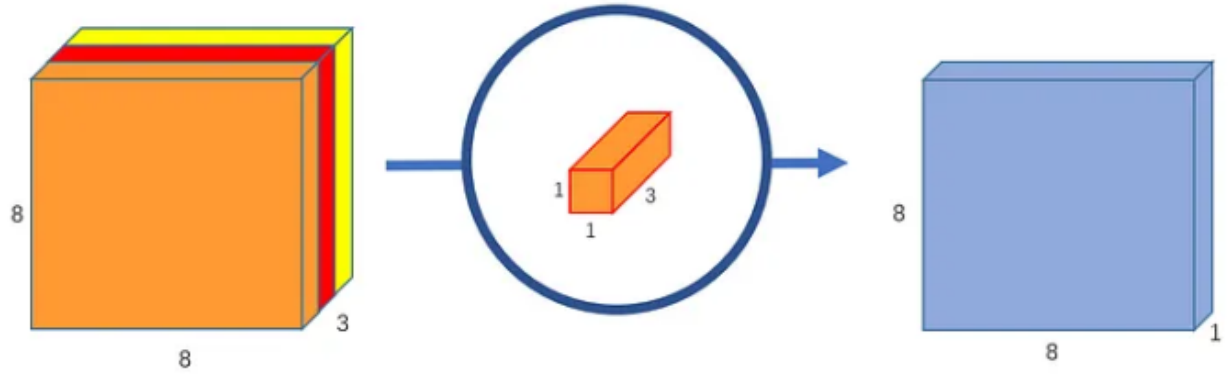

Figure 3.5: Illustration of applying pixel-wise convolution to the output of the depthwise convolution. [1]

mance, especially along object boundaries. In the next section, we are going to explore the DeepLabV3+ mechanism and concepts in detail.

1. **What is Atrous Convolution**

   Atrous convolution is a powerful technique used in deep convolutional neural networks that enables explicit control over the resolution of computed features as in Figure 3.8. It allows adjusting the receptive field without significantly increasing the number of parameters or computational complexity. By introducing dilations or gaps between the filter weights, Atrous convolution effectively captures multi-scale

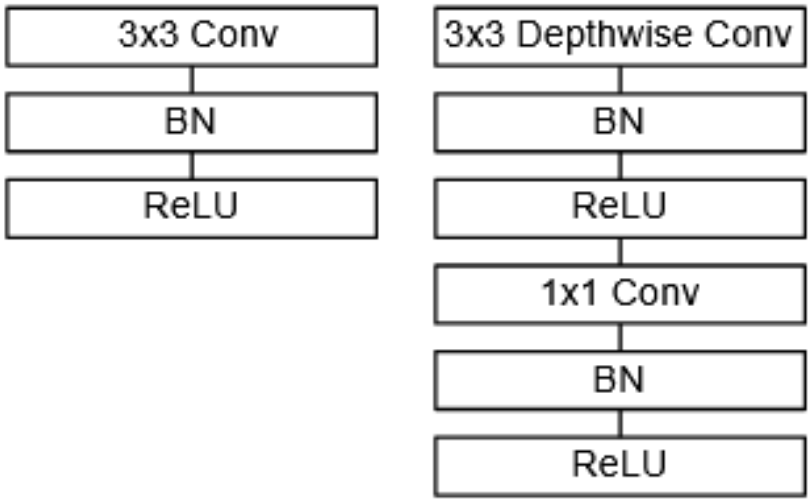

Figure 3.6: Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU. [49].

information and enhances the network's ability to analyze images at different levels of detail as Highlighted in Figure 3.8. This generalization of the standard convolution operation enables CNNs such as DeepLabV3+ to better capture contextual information and improve performance in tasks like semantic image segmentation. for Two-dimensional spaces, For each position i on the output feature map y, a convolution filter w and Atrous convolution are applied to the input feature map x. The process can be described as follows [23]:

$$Y(i) = \sum_{k} X(i + r \cdot k) \cdot W(k)$$

The sample stride at which we process the input signal is determined by the Atrous rate, written as r. It's important to remember that conventional convolution is a particular case where rate r = 1. We can modify the filter's field-of-view and adaptively control the receptive field size of the convolution operation [23].

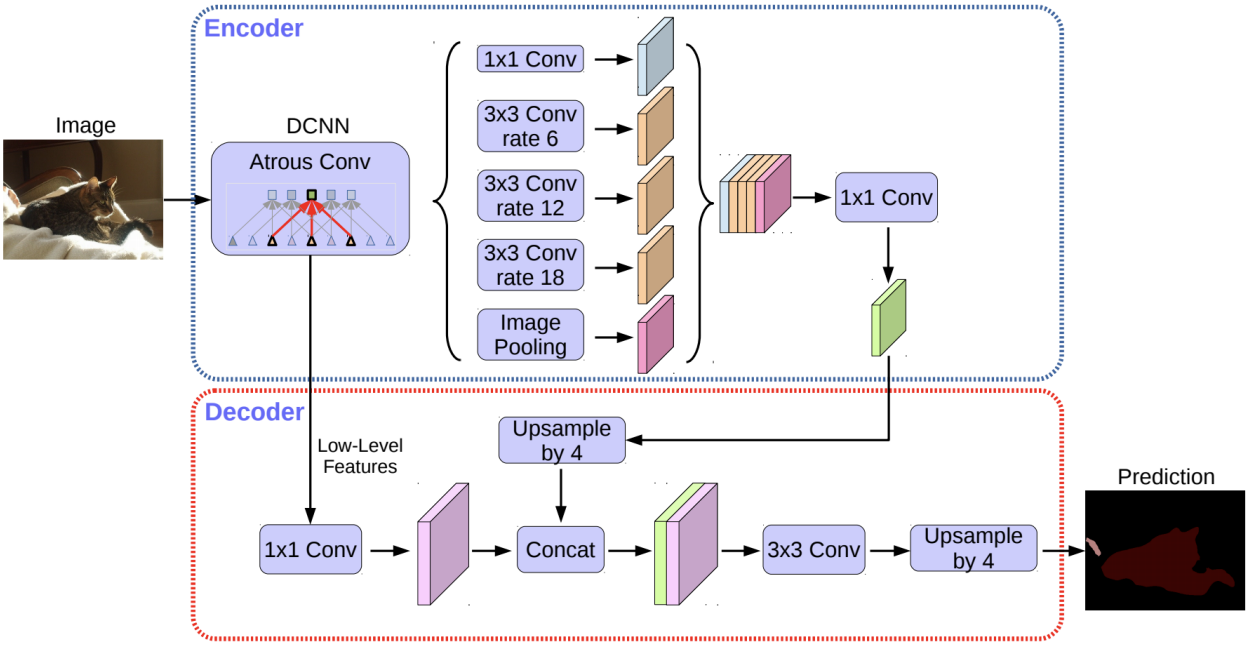

Figure 3.7: DeepLabV3+ architecture. [22]

2. **Encoder module**

As depicted in Figure 3.7 the encoder module is responsible for extracting high-level semantic features from the input image. It typically utilizes deep pre-trained models as the backbone, such as ResNet or Exception. These backbone networks are pre-trained on large-scale classification tasks and have learned to recognize and extract meaningful features from images. The encoder incorporates the Atrous Spatial Pyramid Pooling (ASPP) function [123], which utilizes Atrous (dilated) convolutions to capture information at multiple scales without losing spatial resolution. By using Atrous convolutions, the network can effectively expand the receptive field of each convolutional filter, allowing it to capture a larger spatial context while maintaining the original resolution. This multi-scale feature extraction capability is crucial for accurately capturing targets of different sizes in the image.

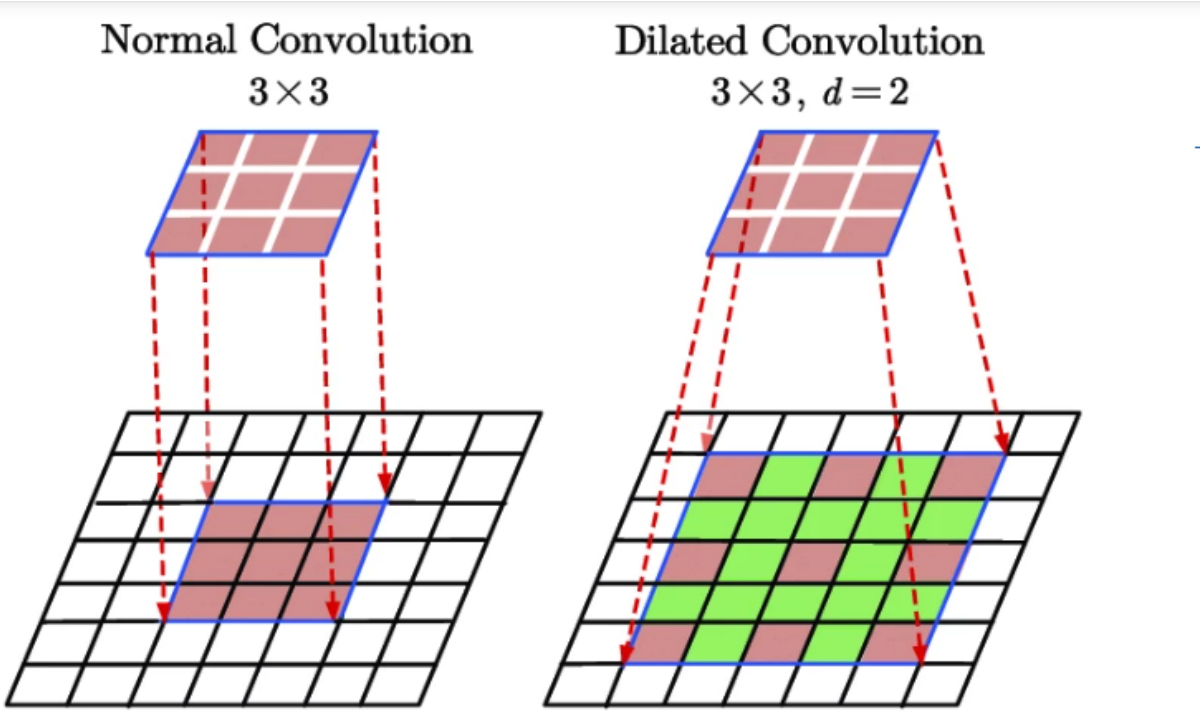

Figure 3.8: Illustration of Atrous/Dilated Convolution [42]

3. **Atrous Spatial Pyramid Pooling**

   As shown in Figure 3.9 The Atrous spatial pyramid pooling scheme is designed to utilize multiple Atrous convolutional layers. Each layer has a different sampling rate. These layers collectively enable the network to extract features at various scales and capture rich contextual information for accurate segmentation [21]. The output feature maps from the different Atrous convolutions are then concatenated to create a rich representation that incorporates contextual information at multiple scales. This fused feature representation can provide a more comprehensive understanding of the image and facilitate more accurate segmentation results.

4. **Decoder module**

   In the decoder module, as is shown in Figure 3.7, the encoder features are initially

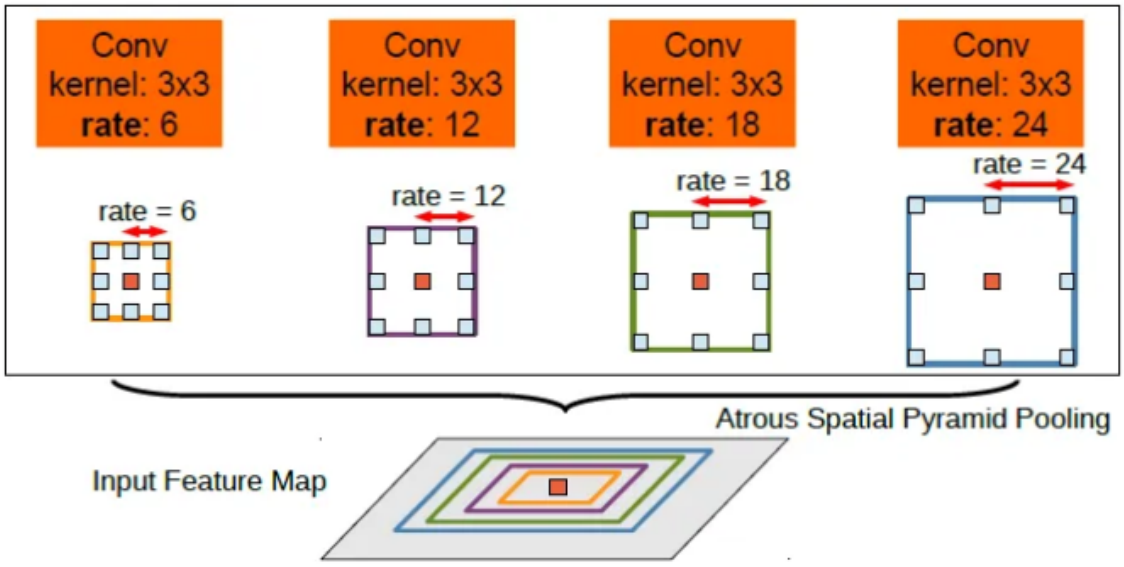

Figure 3.9: Atrous Spatial Pyramid Pooling (ASPP)[21].

upsampled by a factor of 4 using bilinear interpolation. These upsampled features are then concatenated with the corresponding low-level features obtained from the network backbone, which have the same spatial resolution. To reduce the computational complexity and balance the importance of the low-level and rich encoder features, a 1×1 convolution is applied to the low-level features, reducing the number of channels. This step helps prevent the low-level features, which often have a large number of channels, from dominating the valuable encoder features and potentially making the training harder. Following the concatenation, a series of 3×3 convolutions are applied to refine the combined features, enhancing the representation. Finally, the refined features are further upsampled by a factor of 4 using simple bilinear interpolation [23].

3.2 U-Net based on pre-trained Models

The U-Net framework has become widely used in computer vision for a variety of applications, including a significant contribution to image segmentation. As we see in the previous sections the U-Net model is composed of two principal parts encoder and decoder. It learns local features by convolutions and learns local information by skip connections. Our idea is to replace the encoder part with one of the previous pre-trained models' architectures means freezing the lower layers training of the model and continuing the training just for the decoder part as represented in Figure3.10.

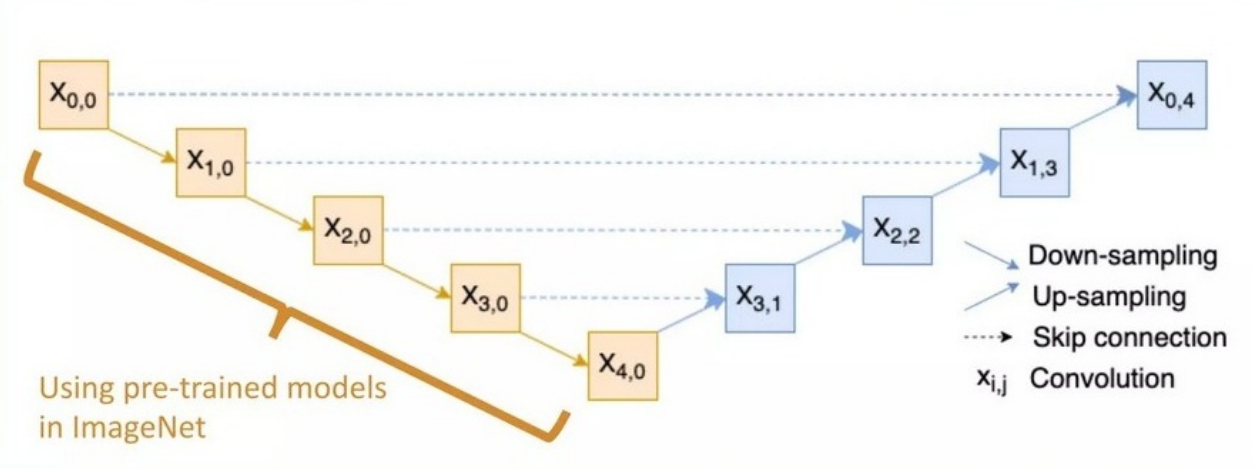

Figure 3.10: Illustration of how the Pre-trained Models are in conjunction with U-Net Architecture.

We aim to extract accurate predictions of the input image for each pre-trained model. To achieve this, we have conducted careful experimentation and made adjustments to the pre-trained architecture of the connected layers, as illustrated in algorithm 1 in "get-layer("model_layer").output". These adjustments ensure that the number of filters and spatial dimensions of the feature maps in the pre-trained model match the conditions required by U-Net. In addition, within the decoder module, we established skip connections and upsampling of the layers through the use of the DECODER_BLOCK function, to

Algorithm 1 Combination Algorithm

function COMBINATION-UNET($input$)
    $pret \leftarrow pretrained\_model(input)$
    $s1 \leftarrow pret.getlayer(\text{"}input\_model\_layer\text{"}).output$
    $s2 \leftarrow pret.getlayer(\text{"}model\_layer\_2\text{"}).output$
    $s3 \leftarrow pret.getlayer(\text{"}model\_layer\_3\text{"}).output$
    $s4 \leftarrow pret.getlayer(\text{"}model\_layer\_4\text{"}).output$
    $b1 \leftarrow pret.getlayer(\text{"}model\_layer\_3\text{"}).output$
    $d1 \leftarrow decoder\_block(b1, s4, 512)$
    $d2 \leftarrow decoder\_block(d1, s3, 256)$
    $d3 \leftarrow decoder\_block(d2, s2, 128)$
    $d4 \leftarrow decoder\_block(d3, s1, 64)$
    $output \leftarrow conv(1, 1, \text{"}sigmoid\text{"})(d4)$
    return $Model$
end function
function DECODER _BLOCK(inputs, skip _features , num _filters)
    $x \leftarrow convTranspose(num\_filters, (2, 2))(inputs)$
    $x \leftarrow concatinate(x, skip\_features)$
    $x \leftarrow CONV\_BLOCK(x, num\_filters)$
    return x
end function
function CONV _BLOCK (x , num _filters)
    $x \leftarrow conv(num\_filters, 3)(x)$
    $x \leftarrow activation(\text{"}relu\text{"})(x)$
    $x \leftarrow conv(num\_filters, 3)(x)$
    $x \leftarrow activation(\text{"}relu\text{"})(x)$
    return x
end function

enhance information flow. By aligning the architectures appropriately we can effectively leverage the pre-trained models and their capabilities for accurate predictions within the U-Net framework. It is worth noting that we have successfully trained each of ResNet, DenseNet, and MobileNet using the same concept algorithm for the U-Net combination. Algorithm 1 outlines the steps we followed to integrate these models. However, it is important to mention that DeepLabV3+ was not included in this combination as it utilizes a different concept of implementation.

3.3 Ensemble Models

Ensemble approaches can be considered the state-of-the reply to many machine learning problems. By training many models and combining their predictions in different manners, these methods increase the prediction accuracy of one model and obtain better generalization performance[104], the main idea is represented in Figure 3.11. Also, it is one of the most important techniques and its uses are increasing each year as shown in Figure 3.12. The primary concept is that evaluating and combining various individual viewpoints is better than selecting one individual opinion[95]. Deep learning architectures are now outperforming shallow or standard models in performance. They combine the advantages of ensemble learning as well as the deep learning models in a manner that the final result will be better[35]. There are many types of ensemble learning[129] such as bagging which is based on separating the dataset into many subsets and each model is trained on a specific subset. Also, the boosting mechanism which is based on the concept of each model will learn from the limitation of the other models in a sequential manner. Another method we have is RF which is a combination of bagging and RF it is based on the concept that each decision tree will learn from a specific subset, we have also stacking where the same dataset is used for the models training, Ada boosting, gradient boosting, weighted Average, and voting which the mechanism that we used in addition the average mechanism.

3.4 The proposed scheme

In this part, we will explain our approach to improve the segmentation results of U-Net through ensemble learning. After training each model separately, namely DenseNet,

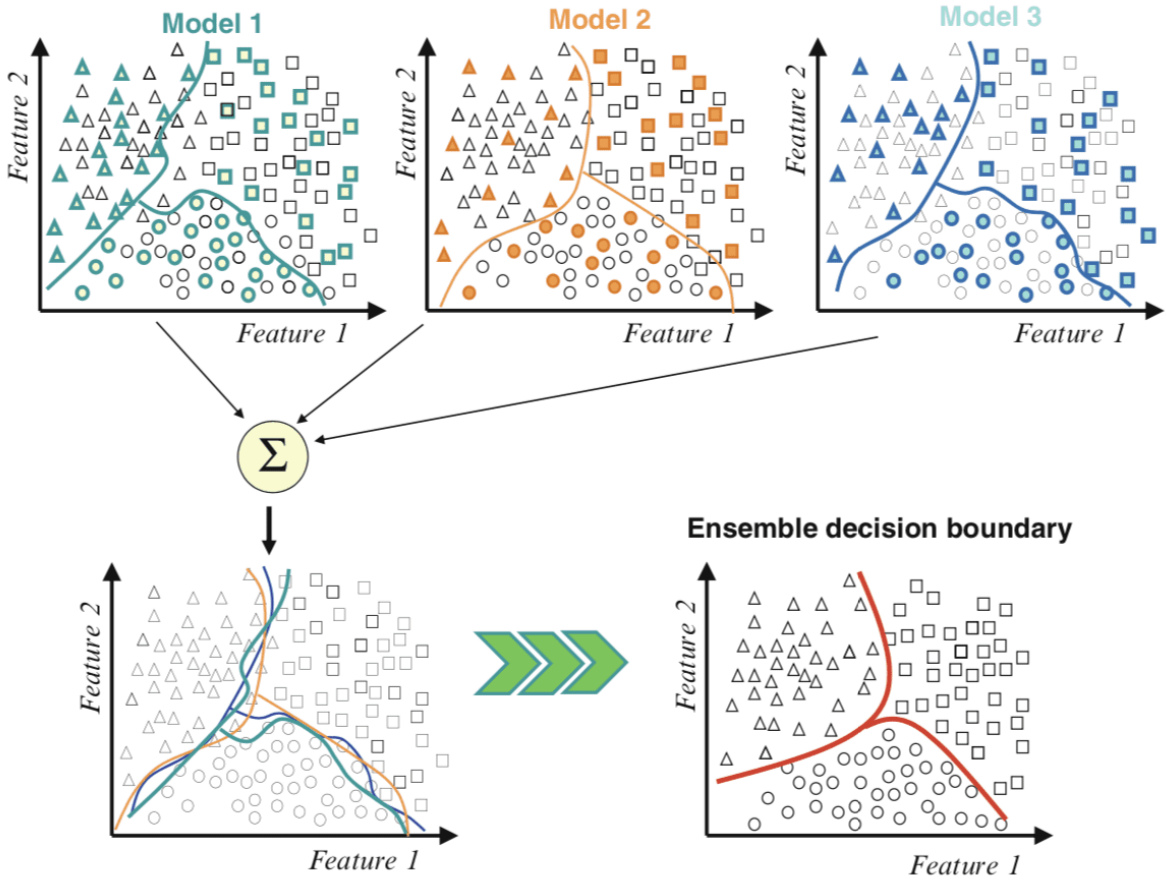

Figure 3.11: Example of combining decision boundaries [135].

ResNet, DeepLabV3+, and MobileNet, we utilize ensemble learning to obtain improved results. Our ensemble model consists of these individual models. We collect the predictions generated by each model and store them in a list, as outlined in the Algorithm 2. To make a final decision based on the predictions of these models, we apply a FUSION SCORE function. This function compares and aggregates the predictions from the individual models. By considering the decisions made by each model, we can generate a more robust and reliable segmentation result by employing ensemble learning with the chosen models (Dense U-Net, ResU-Net, DeepLabV3+, and MobileNetU-Net) the flowchart of our idea is presented in Figure 3.13, we leverage the strengths of each model and take

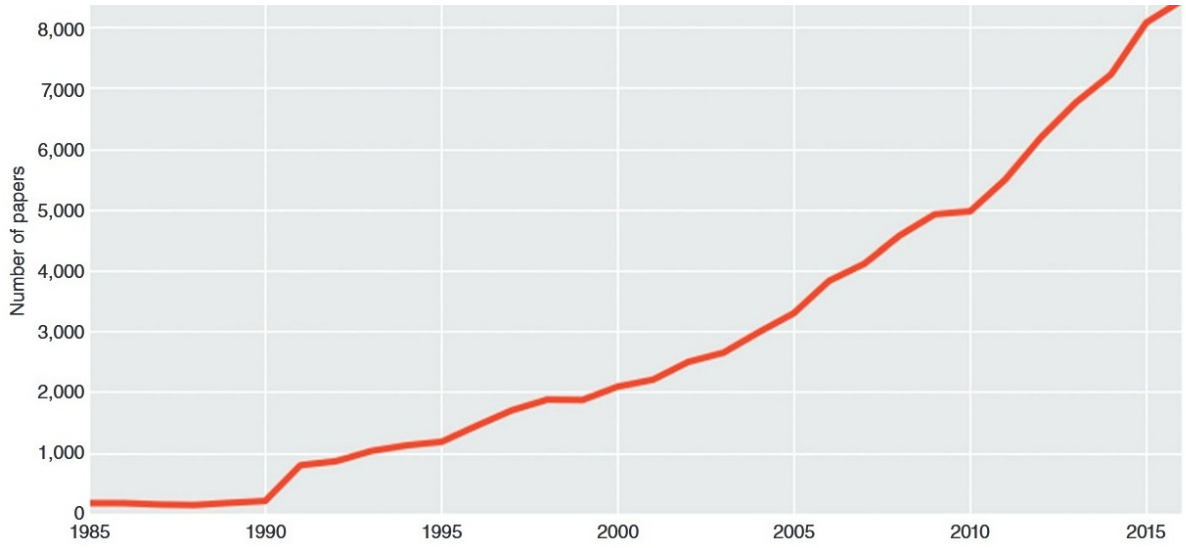

Figure 3.12: Number of published papers per year, based on searching the terms "ensemble" together with "machine learning" in the "web of science" database [104].

advantage of their diverse perspectives. This approach enhances the segmentation performance of U-Net and leads to more accurate and reliable results.

3.4.1 Score-level fusion techniques

The fusion score is a technique used in ensemble learning to combine and evaluate the predictions made by multiple models or classifiers and make a final decision. It helps us to effectively integrate the predictions and generate a more reliable and accurate result. In our approach, we have adopted several fusion techniques including:

1. **voting :** Each model's prediction is treated as a vote, and the final decision is based on the majority votes. For example, if three out of five models predict a certain class, that class is selected as the final prediction. [96]

2. **Averaging :** The predictions of individual models are averaged to obtain the final prediction. This can be done by taking the mean of the predictions.

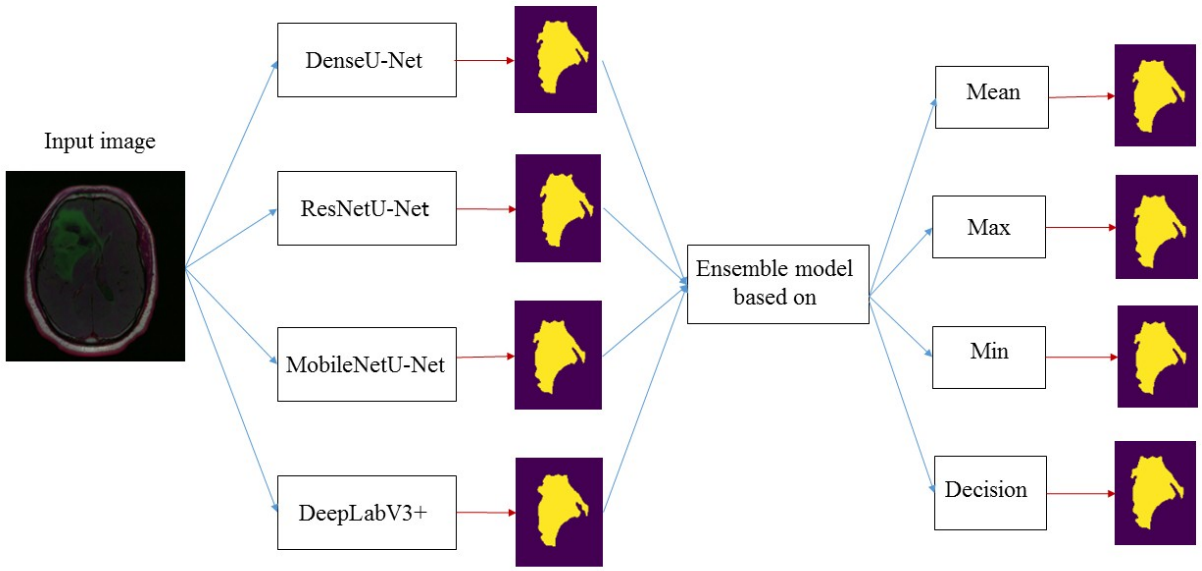

Figure 3.13: The Flowchart of our proposed method

3. **Min/ Max :** the minimum or maximum score among the predictions of the individual models is used as the final prediction. Depending on the characteristics of the problem and the ensemble method being used the minimum score represents the most conservative or cautious decision, and the maximum score represents the most confident or optimistic decision. Overall, The aim is to leverage the diversity of the individual models to obtain a more robust and accurate prediction.

Algorithm 2 Ensemble learning Algorithm

$Models \leftarrow [dense\_unet, Resnet\_unet, Deeplabv3+, Mobilenet\_unet]$
for $i = 0$ **to** $len(testdata)$ **do**
    $img \leftarrow read(testdata\_image[i])$
    $ensemble \leftarrow []$
    for model **in** Models **do**
        $pred \leftarrow model.predict(img)$
        $ensemble.add(pred)$
    end for
    $ensemble\_prediction \leftarrow ENSEMBLE\_FUSION\_SCORE(ensemble)$
    $ground\_truth \leftarrow read(testdata\_mask(i))$
    $accuracy \leftarrow Calculate\_accuracy(ground\_truth, ensemble\_prediction)$
    $ACCURACY.add(accuracy)$
end for
$Total\_accuracy \leftarrow mean(ACCURACY)$

In this chapter, we presented the main method that we work on. In addition to that, we explained our combination and training methodology in detail. In the coming chapter, we will show all the experiments that we did and all the used materials, and the results that we get.

CHAPTER 4

EXPERIMENTAL RESULT

After presenting our methodology and its functioning in the previous chapter, this chapter will focus on the experimental results we obtained, including the materials used, the dataset employed, the evaluation metrics utilized, and a comprehensive analysis of the results. We will delve into the details of our approach and the steps taken to obtain these results.

4.1 The Experimental Dataset

The dataset that we used in our research is the LGG segmentation dataset which is available on Kaggle[15]. The images which are contained in the dataset are MR images each one has three RGB channels and all images are represented in .tif format and they have a size of 256x256. All the images are collected from The Cancer Imaging Archive (TCIA) and they represent 110 patients registered in The Cancer Genome Atlas (TCGA). All patients are got from five different institutions Case Western St.Joseph's with 34 patients, Thomas Jefferson University with 16 patients, Case Western with 14 patients, UNC with 1 patient, and Henry Ford Hospital with 45 patients. Each patient has a number of images between 20 and 88 and the total number of images is 3929 which are divided into 1373 images of tumors and 2556 images without tumors. The patients have divided according to tumor grade categories 50 patients have grade II, 58 grade III, and 2 with unknown grade all those characteristics are shown in Table 4.1. In addition, manual and generated fluid-attenuated inversion recovery (FLAIR) segmentation masks are contained in the dataset and were performed by[16].

	Grade II N = 50	Grade III N = 58	Unknown N = 2
Tumor sub-types			
Astrocytoma	8	26	-
Oligoastrocytoma	14	14	1
Oligodendroglioma	28	18	-
Unknown	-	-	1
Gender			
Male	23	30	1
Female	27	28	-
Unknown	-	-	1
Age	Mean	STD	Range
	46	14	20-75

Table 4.1: Lower-grade glioma tumor (LGG) and patient characteristics[87].

4.1.1 Data Preprocessing

In the context of the LGG Segmentation Dataset, we encounter a data imbalance issue, characterized by a significant disparity in the number of samples between the healthy brain tissue class and the tumor region class. Specifically, the dataset contains a larger proportion of scans depicting healthy brain tissue compared to scans depicting tumor regions As shown in Figure 4.1.

The presence of data imbalance poses challenges when training accurate segmentation models. When one class has a much larger number of samples than the other the model tends to exhibit a bias towards the majority class during training. Consequently, the model's performance may suffer when it comes to identifying the minority class which in this case corresponds to the tumor regions. This bias can result in inaccurate or incomplete segmentation results as the model may struggle to capture the unique features and characteristics specific to the minority class.

Addressing data imbalance becomes crucial to ensure robust and accurate segmentation within the LGG Segmentation Dataset. One effective technique to mitigate the impact of class imbalance And overfitting is data augmentation. By employing data augmen-

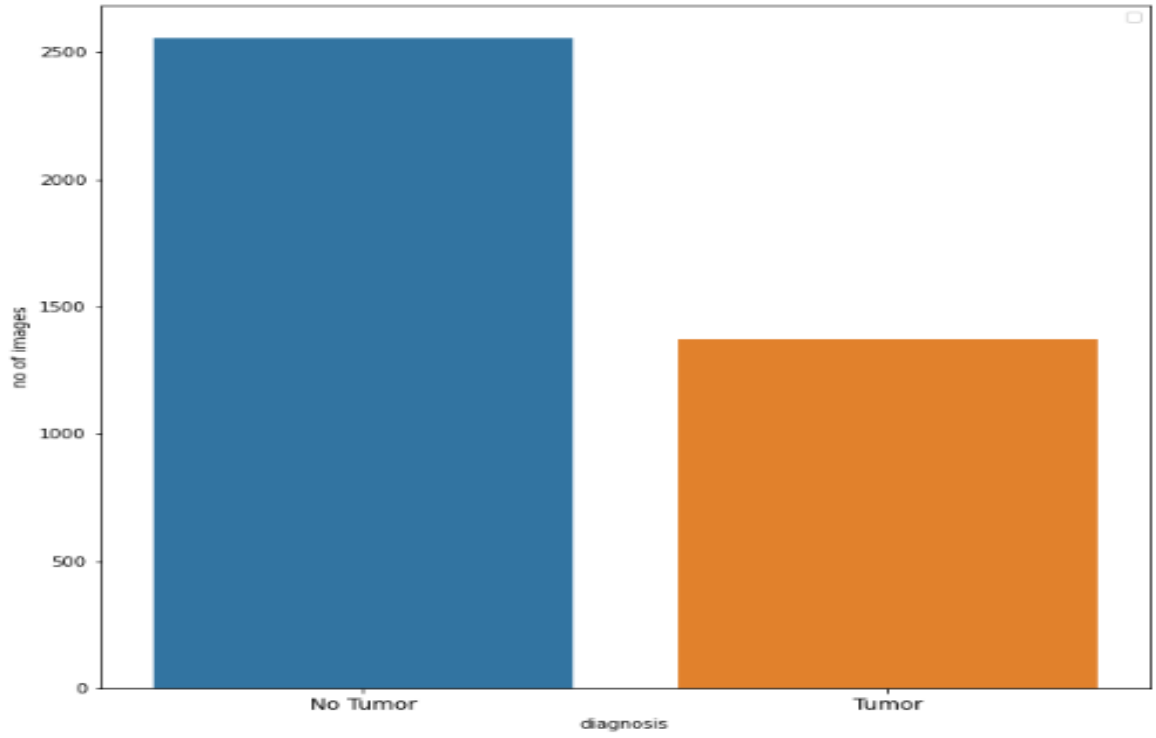

Figure 4.1: Imbalanced Distribution of Classes in the LGG Segmentation Dataset.

tation we can alleviate the imbalance by generating synthetic samples through various transformations and perturbations applied to the existing dataset.

4.1.2 Data Augmentation

The primary goal of data augmentation in the context of imbalanced datasets is to alleviate the bias towards the majority class and enable the model to learn more effectively from the minority class. By introducing variations in the existing samples we can create additional instances of the minority class thus reducing the imbalance. This augmentation process helps in capturing the variability associated with the tumor regions, enabling the model to generalize better and make accurate predictions on unseen data [103].

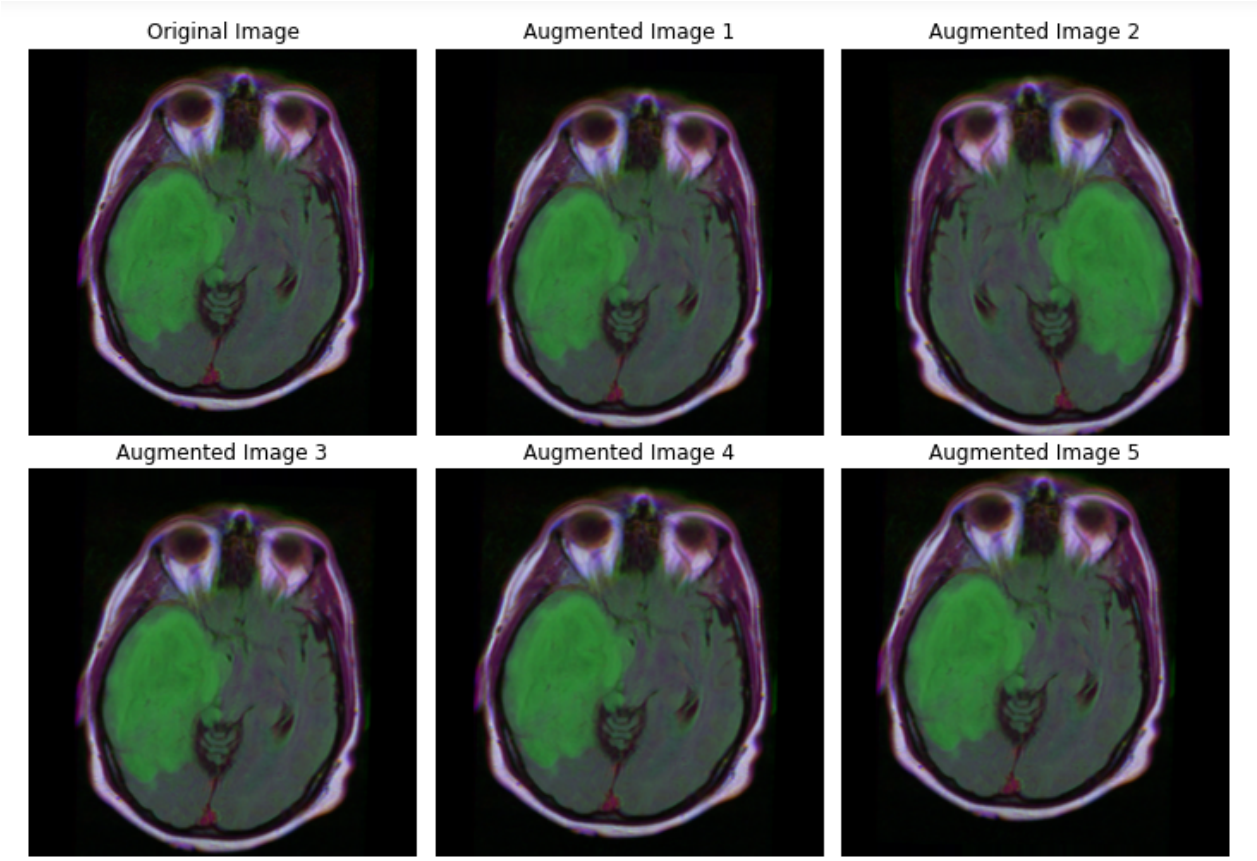

Figure 4.2: Augmented Images Generated for Addressing Class Imbalance in the LGG Segmentation Dataset.

4.2 The Experimental BraTs Dataset

We have adopted the BraTS2020 dataset for our experiment, This dataset consists of MRI images that are annotated with four distinct classes: necrotic, edema, enhancing tumor, and the healthy part of the brain. These annotations play a crucial role in understanding the spatial extent and characteristics of different tumor components, which is vital for the diagnosis, treatment planning, and monitoring of brain tumors. We have developed a DataGenerator class to facilitate the generation of training data for the BraTs dataset. Instead of utilizing 3D volumes directly, we opted to work with individual 2D slices. In

other words, tries to convert the 3d dataset to 2d. This decision enabled us to overcome the difficulties we encountered when utilizing 3D volumes to work with individual slices instead of 3D volumes was driven by computational resource limitations. The BraTs dataset contains a large number of volumetric data, and processing the entire 3D volumes simultaneously would have required significant computational power and memory. By focusing on individual slices, we were able to reduce the computational requirements and make the training process more manageable within the available resources. This approach allowed us to strike a balance between model performance and resource constraints, leading to successful segmentation results while efficiently utilizing the computational resources at hand. However, table 4.2 presents the quantitative results obtained from our experiments on the BraTs dataset. These results highlight the performance of our approach using the DataGenerator class and slice-based training with a traditional U-Net architecture. The table provides metrics such as accuracy, precision, recall, and F1-score, which evaluate the segmentation accuracy of our model. Furthermore, we have included figure 4.3 showcasing the predictions made by our model on the Brats dataset. These figures depict the segmentation outputs for different patients.

Accuracy Metrics		
Accuracy	**Mean IoU**	**Dice Coefficient**
0.9902	0.7893	0.6095
Precision	**Sensitivity**	**Specificity**
0.9938	0.9674	0.9587
Dice Coefficients		
Necrotic	**Edema**	**Enhancing**
0.4035	0.6204	0.5726

Table 4.2: Quantitative Results on BraTs Dataset.

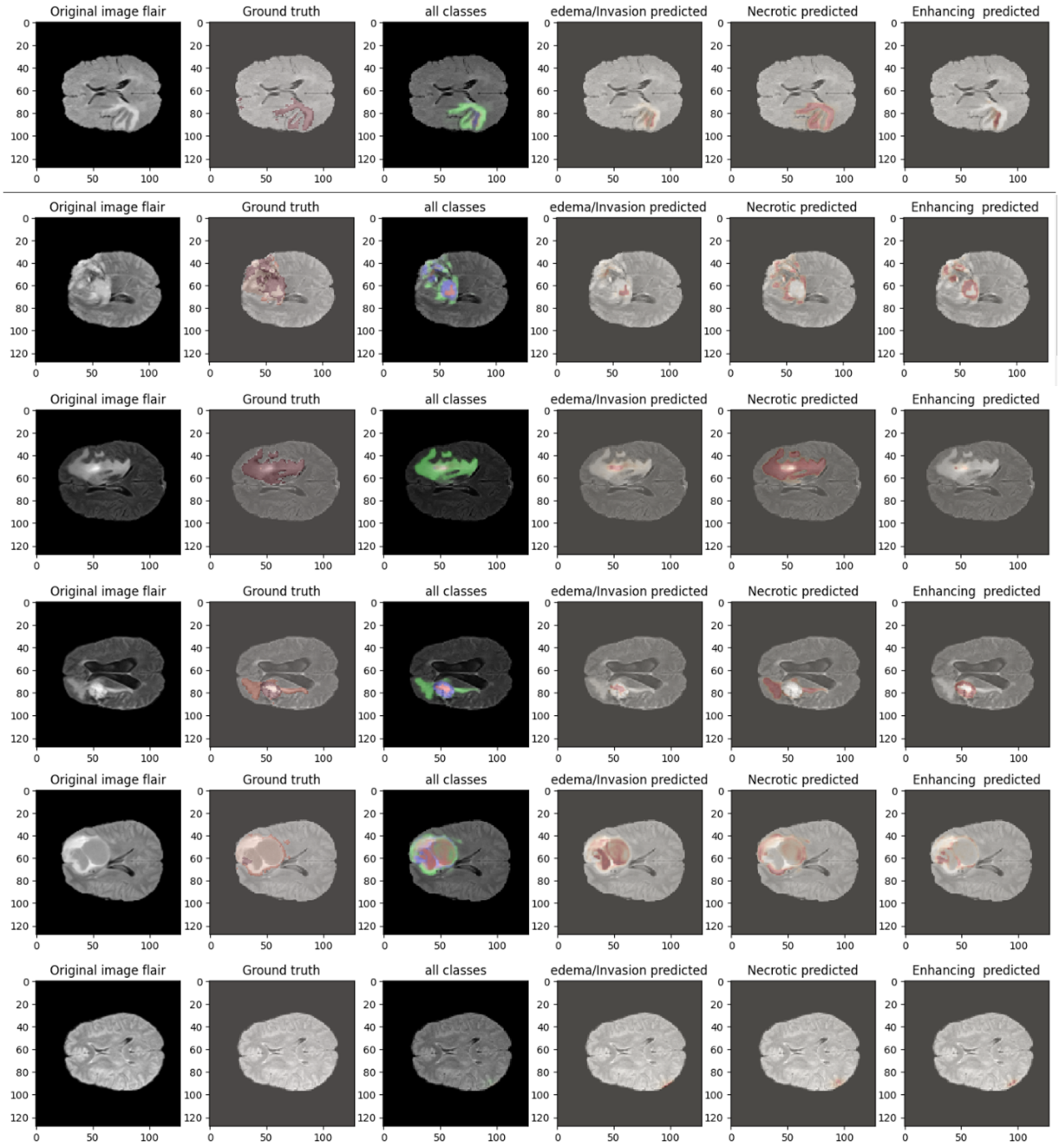

Figure 4.3: The predicted tumor segmentation showcases the different tumor components identified by the model.

4.3 Implementation

4.3.1 Software and libraries

1. **TensorFlow**

   TensorFlow is a versatile and widely-used software library for machine learning and artificial intelligence. With its extensive capabilities, it supports a broad spectrum of tasks, placing special emphasis on the training and inference of deep neural networks. As a free and open-source library, TensorFlow empowers researchers, developers, and practitioners to explore and leverage the potential of machine learning and AI effectively [134].

2. **Keras**

   Keras is a high-level deep learning API designed to build and train neural networks. It is designed to be user-friendly, modular, and efficient. Keras makes implementing neural networks easier by providing a simple and intuitive interface to define, compile, and train models. It is written in Python and supports several back-end libraries for neural network computations, such as TensorFlow, Theano, and CNTK. This allows users to select the back-end of their choice for neural network computations.

4.3.2 Tools

We have utilized the Kaggle platform, which offers cloud-based Jupyter notebooks and provides additional resources such as GPU P100, Disk(73,1GB), Ram(13GB), and CPU. These resources greatly aid us in performing the training process.

4.4 Evaluation Metrics

The evaluation metric is a measure that we use to calculate or measure our model work and performance. In the machine learning domain, there are several measurement techniques each of which is specified to one or more applications like classification, detection, and segmentation. Some of the most known metrics that we used in our work are the following:

4.4.1 Dice Coefficient

It is a famous statistical measure used in the image segmentation field and it is considered the most used validation tool in the domain[110]. In general, it is used to calculate the similarity between data more specifically is worked on to compare pixels of the ground truth and the predicted segmentation. The dice value is calculated as follows, dividing the intersection value of the predicted mask and the ground truth multiplied by two over the summation of the two areas for more understanding see Figure4.4. The dice value is between zero and one where one means that the prediction is excellent and zero means that the prediction is very poor.

4.4.2 Intersection Over Union

As abbreviation IOU or the Jaccard index, it is a well-known measure used in image segmentation and object detection problems[100]. It is a very simple concept since it calculates the overlap ratio between the ground truth(GT) and the predicted segments. To calculate this value the intersection is divided by the union as it is shown in Figure4.5, the output value is always between zero and one, where the zero value means that the

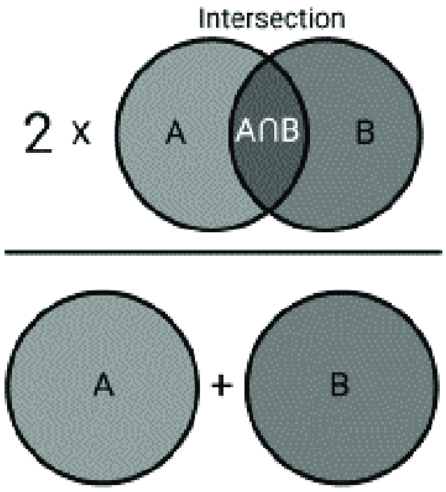

Figure 4.4: Calculation of segmentation quality metrics Dice similarity coefficient [60].

Metric	Purpose	Calculation form	Range
Dice	image segmentation	2 * Intersection / (Predi + GT)	0-1
IOU	image segmentation/Object detection	Intersection / Union	0-1
Accuracy	Overall correctness of predictions	correct predictions / Total number of predictions	0-1

Table 4.3: Summary of the main differences and purposes of our used evaluation metrics.

prediction is very bad and the one value means that the prediction is very good. And we can put a threshold value that makes us decide whether the prediction is well or not.

4.4.3 Accuracy

In machine learning each task need to be evaluated, the accuracy metric is very useful for many tasks such that classification, detection, and segmentation[53]. It is based on calculating the correctness of affecting each data point to the right class so it is calculated by dividing all the right and correct predictions over the total number of input samples. In general, it gives good results where we have the classes balanced.

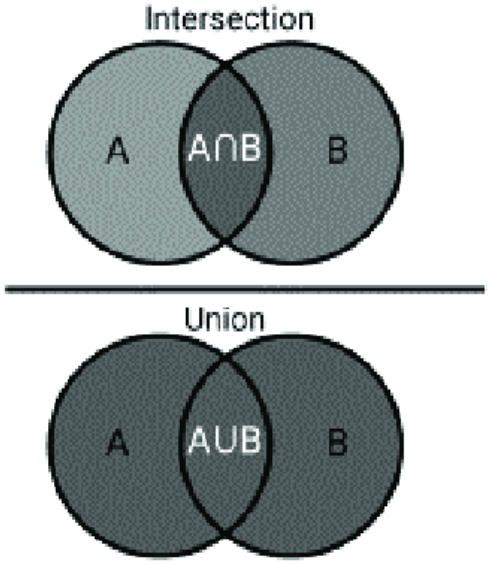

Figure 4.5: Calculation of segmentation quality metrics Intersection over Union [60].

4.4.4 Confusion Matrix

In the context of image segmentation, the confusion matrix is computed at the pixel level to evaluate the performance of the segmentation model. Each pixel in the confusion matrix represents a prediction for a specific class.

In the case of binary classification (such as tumor vs. non-tumor), the confusion matrix has a 2x2 structure:

Actual/Predicted	Negative	Positive
Negative	TN	FP
Positive	FN	TP

- **True Positive (TP):** The number of pixels correctly classified as the positive class (e.g., tumor).
- **True Negatives (TN):** The number of pixels correctly classified as the negative class

(e.g., non-tumor).

- **False Positives (FP):** The number of pixels incorrectly classified as the positive class when they actually belong to the negative class.
- **False Negatives (FN):** The number of pixels incorrectly classified as the negative class when they actually belong to the positive class.

4.4.5 Cross-validation

Cross-validation is a technique used in machine learning and statistical modeling to evaluate the performance and the generalizability of a predictive model. It involves partitioning the available dataset into multiple subsets or "folds" as visualized in figure 4.6. The model is then trained on a subset of the data called the training set and evaluated on the remaining data known as the validation set. This process is repeated multiple times [99], with different subsets serving as the validation set each time. By averaging the performance across these iterations, cross-validation provides a more reliable estimate of the model's performance and helps identify potential issues such as overfitting. At the same, describe the generalization ability of the model.

4.5 Comparison and Discussion

The table 4.4 provides a comparison of different models used for segmentation tasks on the test images, based on their accuracy metrics. The models evaluated include U-Net, Dense U-Net, Residual U-Net, DeepLabV3+, MobileNet U-Net, Ensemble mean, Ensemble-min, Ensemble-max, and Ensemble based on the most frequent vote.

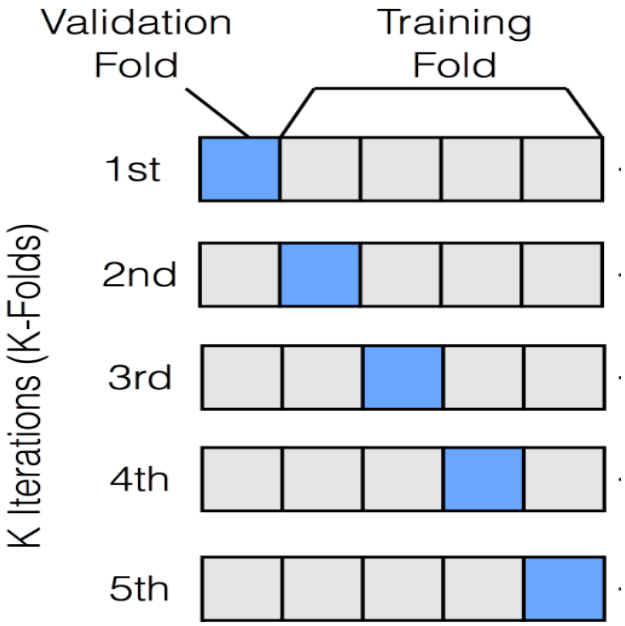

Figure 4.6: A visual representation of dividing the dataset into multiple subsets for cross-validation.

Among the individual models evaluated using our dataset, Residual U-Net demonstrates the highest performance in terms of IOU. This can be attributed to several factors, one of which is the utilization of residual connections in Residual U-Net. The main difference between Residual U-Net, which incorporates ResNet Blocks in the encoder module, and Dense U-Net, which incorporates DenseNet blocks in the encoder module, lies in their approach to utilizing preceding feature maps. ResNet uses a single preceding feature map [136], while DenseNet incorporates features from all the preceding convolutional blocks [51]. However, both Residual U-Net and Dense U-Net share a common philosophy of connecting to the feature maps of all preceding convolutional blocks, allowing for better information flow and improved gradient propagation throughout the network.

In Figure 4.7, it is evident that Residual U-Net consistently outperforms other models, showcasing its superiority in capturing meaningful features and achieving accurate semantic segmentation.

Model	IOU	Dice-coeff	Accuracy	Number of Params
U-Net	0.78364	0.86560	0.99804	31,043,521
Dense U-Net	0.82803	0.87840	0.99867	16,352,961
Residual U-Net	0.86768	0.91821	0.99888	20,642,113
DeepLabV3+	0.85668	0.91341	0.99886	17,795,425
MobileNet U-Net	0.85774	0.91102	0.99887	**11,725,617**
Ensemble Mean	**0.869282**	0.914094	0.999058	-
Ensemble Min	0.835959	0.889702	0.998853	-
Ensemble Max	0.86676	**0.92027**	**0.99901**	-
Ensemble Freq	0.86331	0.911681	0.99891	-

Table 4.4: Accuracy metrics comparison of different segmentation models on the test images.

Model	IOU	Dice-coeff	Accuracy
Residual U-Net	0.78852	0.85962	0.99828
MobileNet U-Net	0.78018	0.85494	0.99813
Ensemble_Bagging	0.80515	0.87498	0.99846

Table 4.5: Results of Bagging Ensemble Implementation.

In our evaluation, we found that Dense U-Net had a lower performance compared to Residual U-Net. This could be attributed to the fact that Dense U-Net has a smaller number of parameters. The DenseNet architecture offers the option of feature reuse, allowing the layers to generate fewer parameters while representing the same information as simpler layers. However, this reduction in the number of parameters might have affected the overall performance of Dense U-Net compared to Residual U-Net in our specific dataset. We present the performance during training of the individual Dense U-Net models in the following Figure 4.8.

On the other hand, DeepLabV3+ achieved an IOU score of 0.85668, indicating the accuracy of its segmentation results. Although this score is slightly lower than that of Residual U-Net, DeepLabV3+ stands out due to its unique architecture and the concept of Atrous

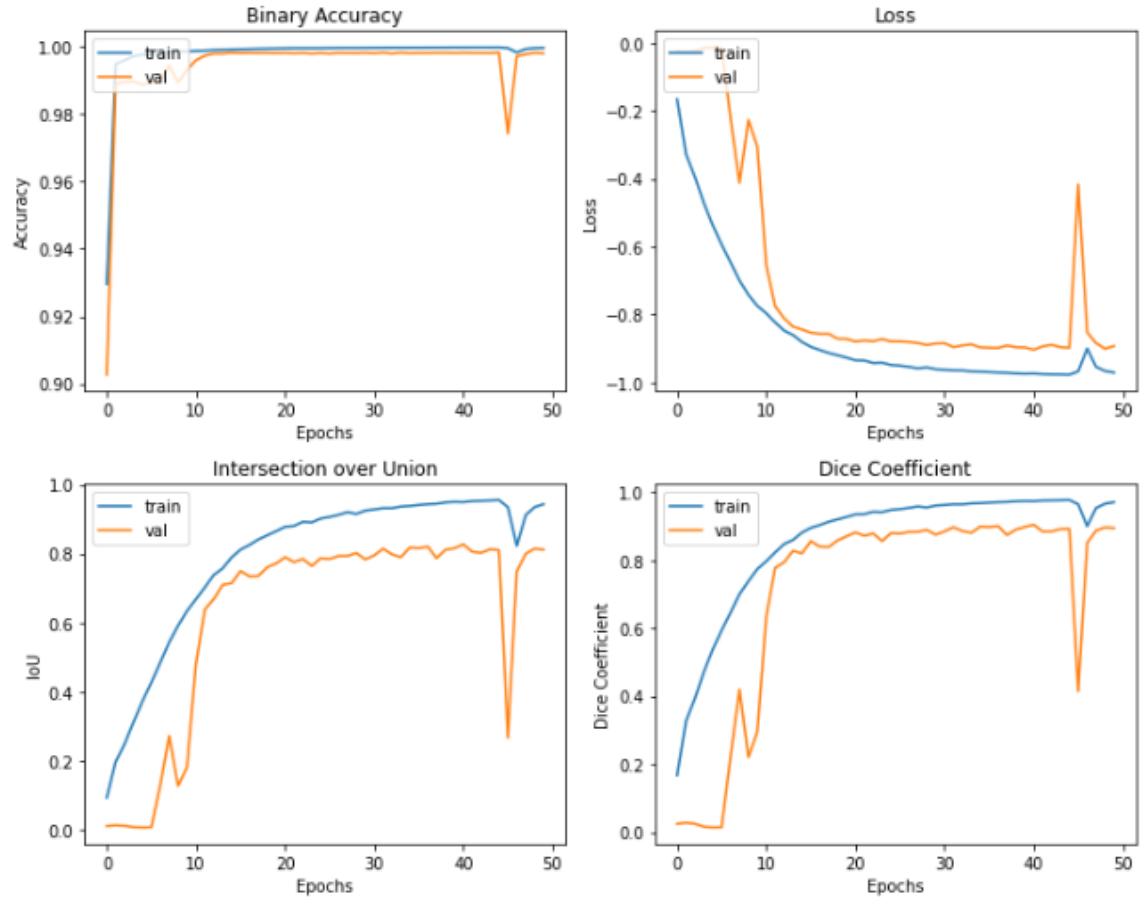

Figure 4.7: Residual U-Net performance.

convolution. The encoder-decoder architecture employed in DeepLabV3+ enables it to effectively capture and represent contextual information [23], leading to improved segmentation performance. In general, DeepLabV3+ remains one of the top-performing models As shown in Figure 4.9, leveraging its innovative design to achieve accurate and reliable segmentation results in our work.

MobileNetU-Net offers an excellent balance between performance and efficiency with its lower number of parameters. Residual U-Net achieves higher IOU and Dice coefficient scores but has more parameters. DeepLabV3+ performs well across metrics, similar to MobileNetU-Net but with a slightly larger number of parameters. The number of

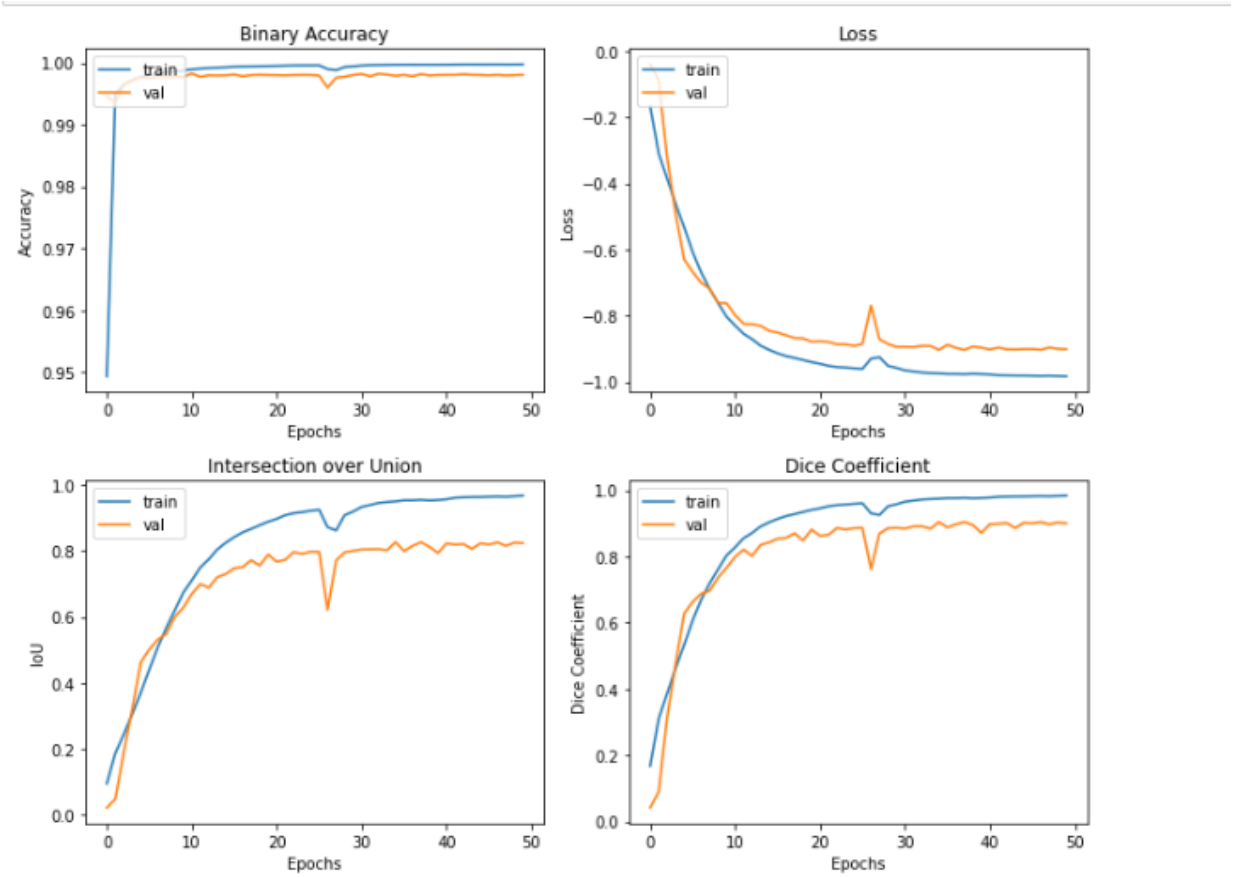

Figure 4.8: Dense U-Net performance.

parameters in the MobileNetU-Net model is 11,725,617, which is relatively lower compared to other models in the table, This emphasizes the lightweight and efficient nature of the MobileNetU-Net architecture. Overall, MobileNetU-Net showcases the benefits of its efficient architecture by achieving impressive results with fewer parameters as it is highlighted in Figure 4.10. It seems that the idea of depth separable convolution has a significant benefit and leads to excellent results in our dataset compared to the traditional convolution used in U-Net. The depth separable convolution, consisting of depthwise convolutions and Pointwise convolutions, allows for more efficient and lightweight computation while preserving important spatial information. This architectural choice in the

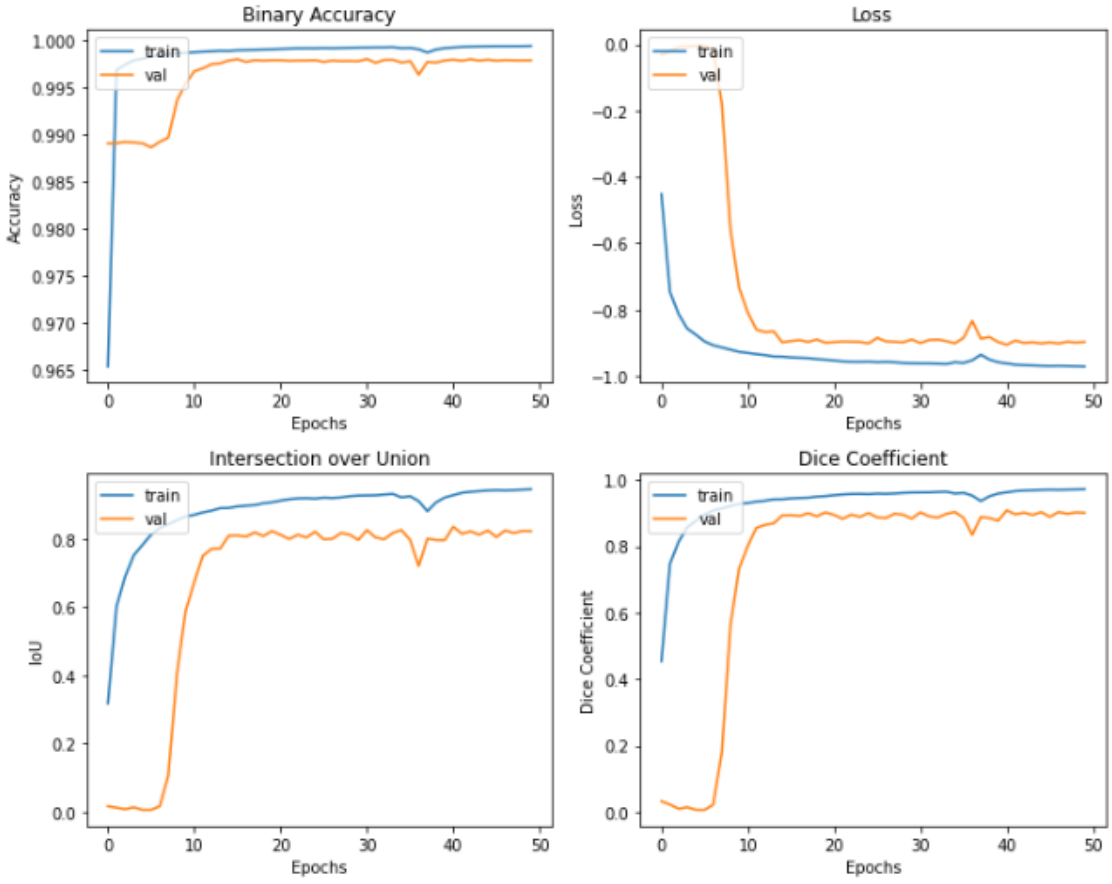

Figure 4.9: DeepLabV3+ performance.

MobileNetU-Net model likely contributes to its superior performance in terms of accuracy, IOU, and Dice coefficient on our dataset.

To assess the feasibility and stability of the constructed model, we performed cross-validation on the training data. A k-fold cross-validation scheme was employed, where k represents the number of splits for training and validation data. This approach ensured that the dataset for each class was randomly divided into k exclusive subsets, avoiding data grouping.

Remarkably, each individual model exhibited outstanding performance across the entire dataset, demonstrating proficiency in various aspects or segments of the data, as de-

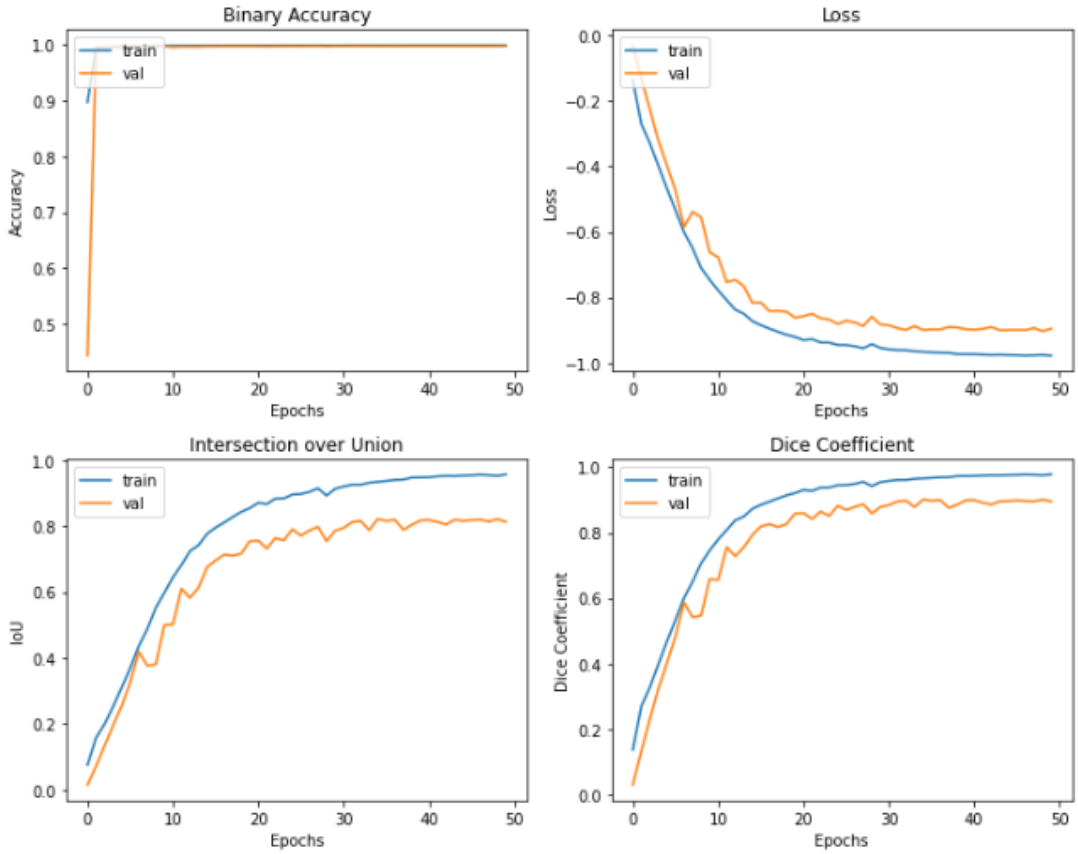

Figure 4.10: MobileNetU-Net performance.

picted in Figure 4.18. These results provide evidence for the successful creation of unbiased and stable models. The exceptional performance of the chosen models distinguishes them from other suggested models, confirming our decision to adopt them for our approach.

The Ensemble models, which combine the predictions of multiple models, show even better performance. Among them, the Ensemble-max model stands out by selecting the maximum score between the chosen models for each pixel. It achieves the highest Dice-coeff score of 0.92027, and overall accuracy of 0.99901. we introduce diversity in the results "Min, Max, Frequency" to capture a wide range of predictions and select the best-fitting one for our study. This approach enhances the robustness and accuracy of the

segmentation. The Ensemble-max model leverages the strengths of the individual models and maximizes the segmentation quality by selecting the maximum score. Overall, the Ensemble-max models demonstrate superior performance, showcasing the effectiveness of ensemble modeling in segmentation tasks.

In contrast, ensemble bagging demonstrates excellent performance compared to individual models refer to Table 4.5. However, due to the concept of using subsets of data in the bagging ensemble, where the data is divided in half for both ResidualU-Net and MobileNetU-Net models instead of utilizing the entire dataset, resulting the performance is lower than that of our proposed ensemble models in our specific dataset.

- **Predictions of individual models and ensemble learning.**

After comparing the performance of different models on the test set, we provide visual representations of the predictions made by the Dense U-Net, Residual U-Net, DeepLabV3+, and MobileNetU-Net models see Figure 4.12, Figure 4.13 and Figure 4.14, as well as the predictions of the ensemble models on the test images with different patients can be observed in Figure 4.15, Figure 4.16 and Figure 4.17. These images allow for a comprehensive evaluation of the segmentation capabilities of each model. By observing the predictions, we can clearly see the effectiveness of the individual models in accurately delineating the objects of interest in the test dataset. Additionally, the ensemble models further enhance the segmentation results by combining the predictions of multiple models.

Furthermore, we have presented a confusion matrix of the mean ensemble model With reference to Figure 4.11, which allows us to evaluate the performance of the ensemble models at the pixel level. By analyzing the confusion matrix, we can assess how well the ensemble models perform in accurately classifying each pixel. This evaluation provides

insights into the effectiveness of the ensemble models in accurately delineating the objects of interest in the test dataset, showcasing their performance, and highlighting their ability to make precise pixel-level predictions.

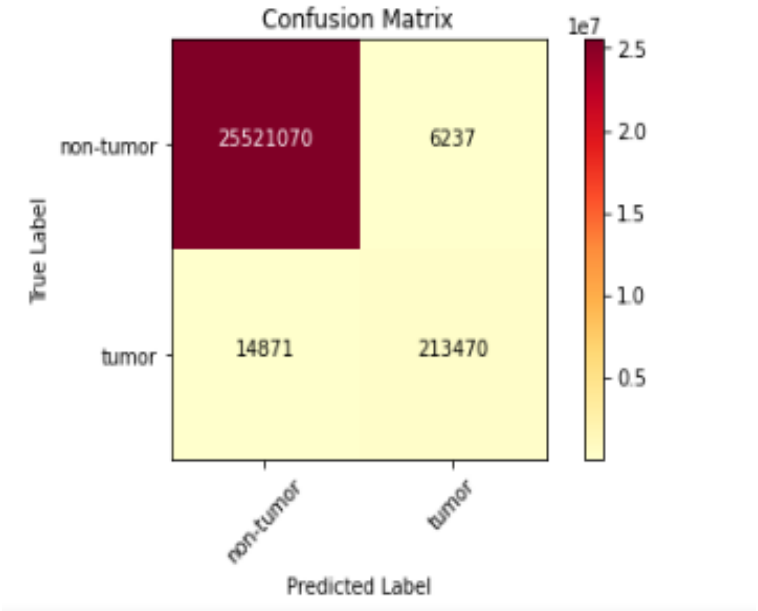

Figure 4.11: Confusion matrix of The mean Ensemble models.

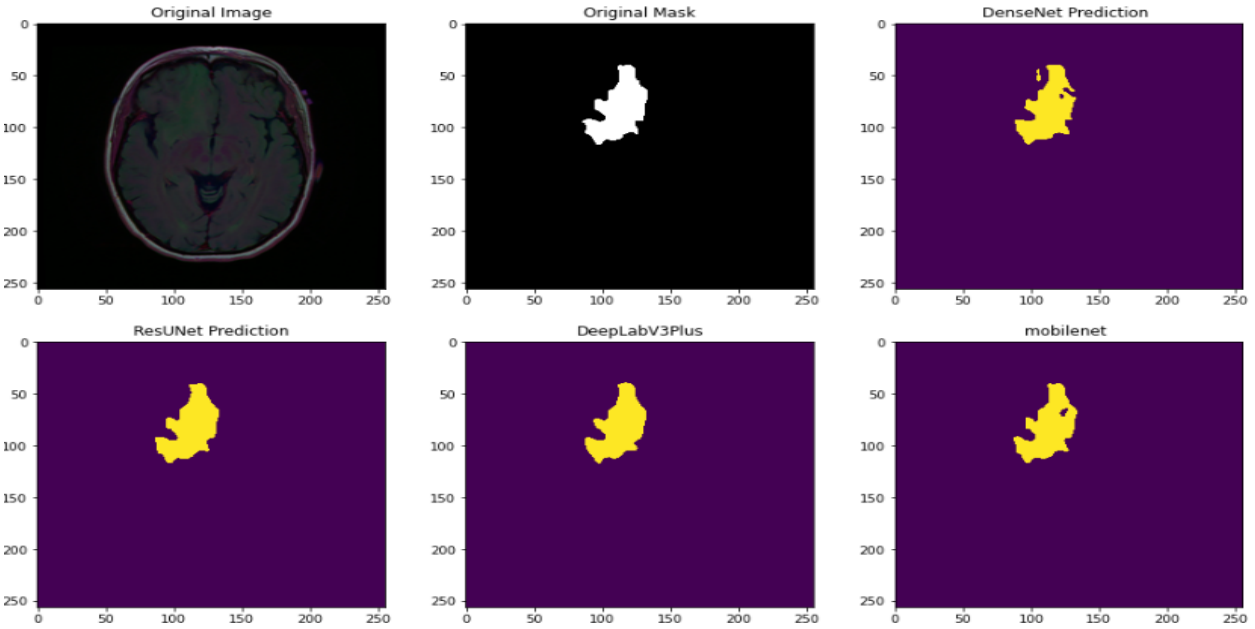

Figure 4.12: Predictions of individual models on Patient TCGA_DU_7300_19910814: Segmentation results showcasing the outputs generated by the U-Net, Dense U-Net, Residual U-Net, DeepLabV3+, and MobileNetU-Net.

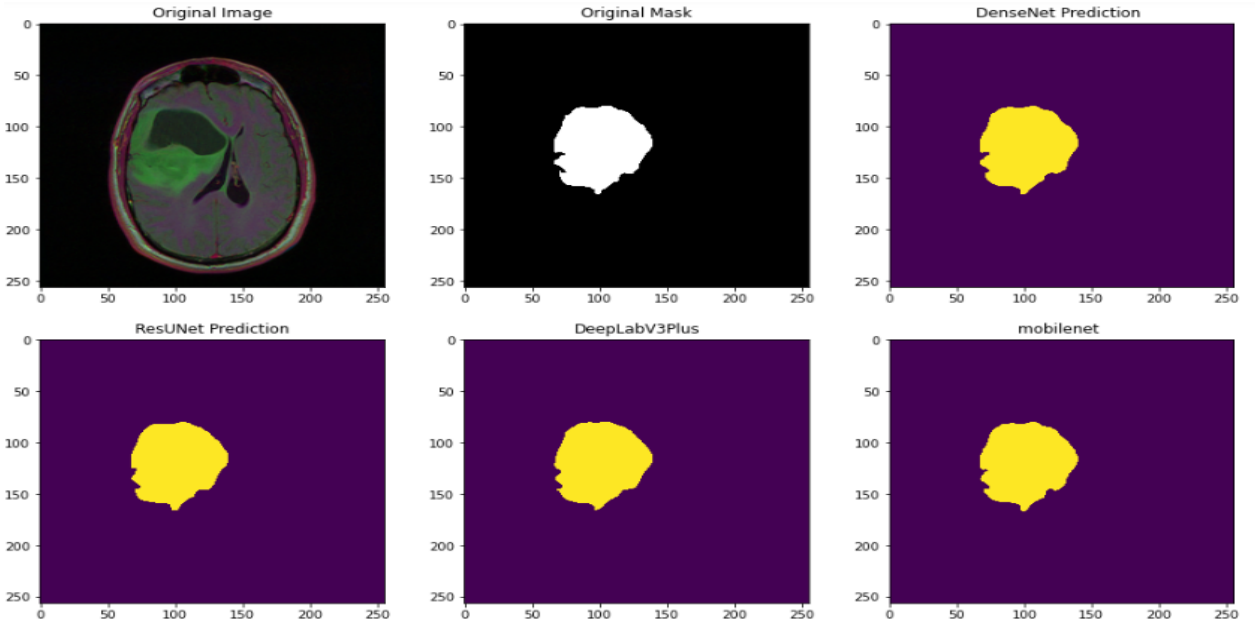

Figure 4.13: Predictions of individual models on Patient TCGA_FG_6689_20020326: Segmentation results showcasing the outputs generated by the U-Net, Dense U-Net, Residual U-Net, DeepLabV3+, and MobileNetU-Net.

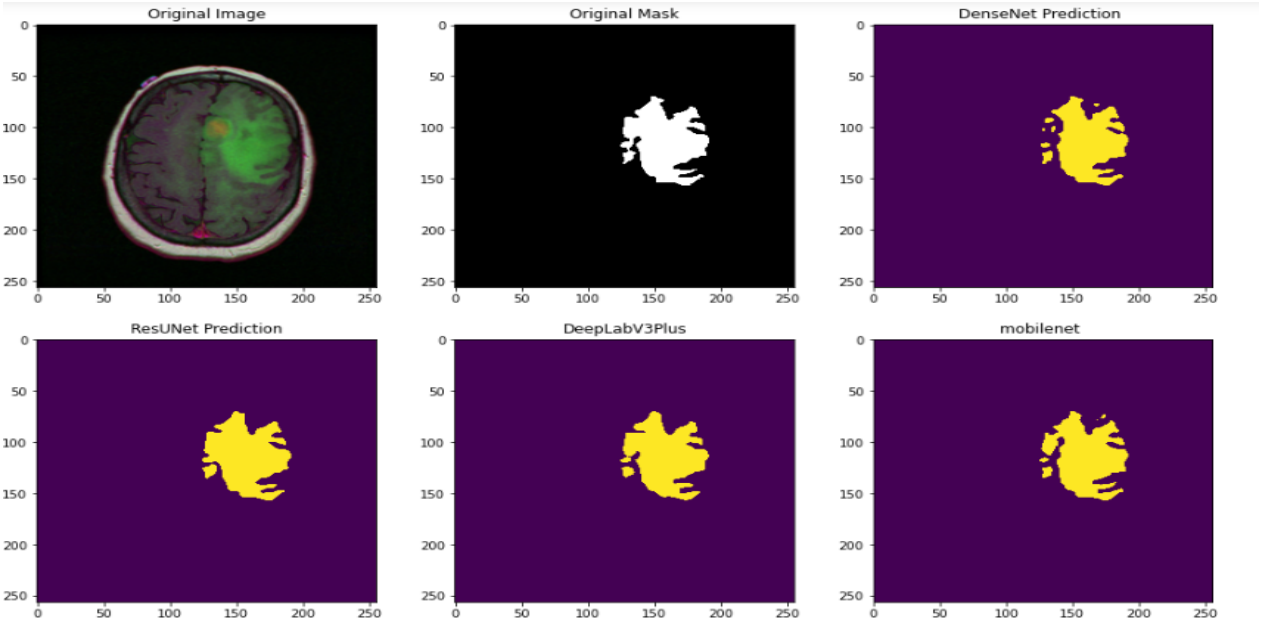

Figure 4.14: Predictions of individual models on Patient TCGA_DU_7010_19860307: Segmentation results showcasing the outputs generated by the U-Net, Dense U-Net, Residual U-Net, DeepLabV3+, and MobileNetU-Net.

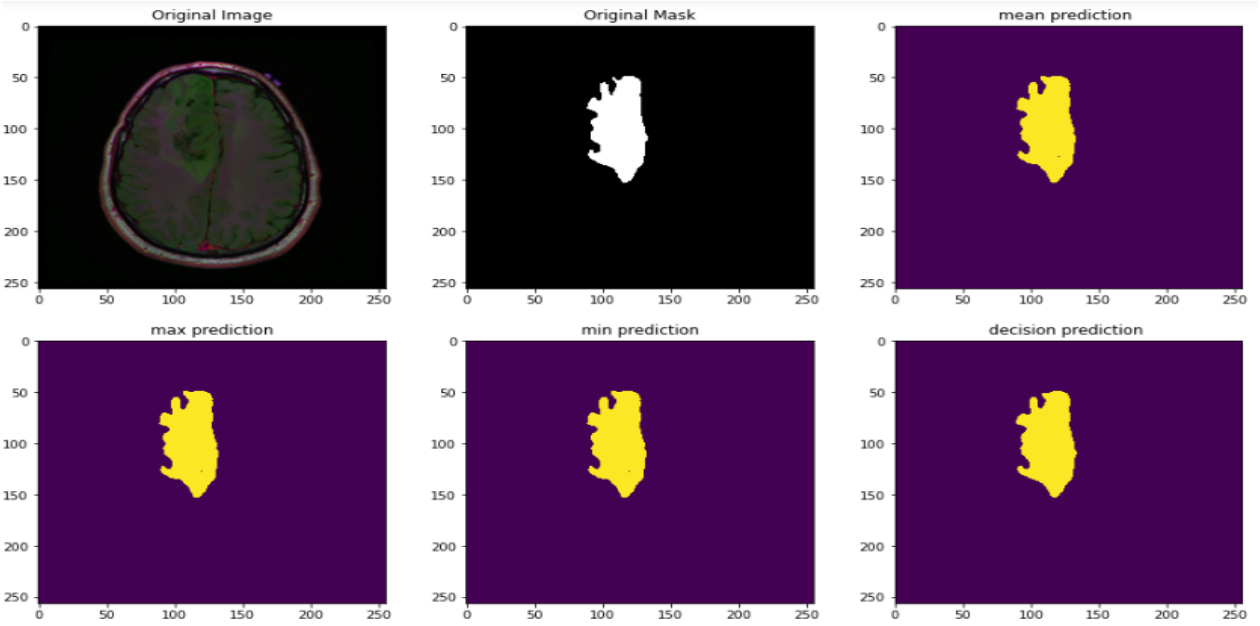

Figure 4.15: Predictions of Ensemble Models on Patient TCGA_DU_8164_19970111 The image displays the segmentation results generated by the ensemble models.

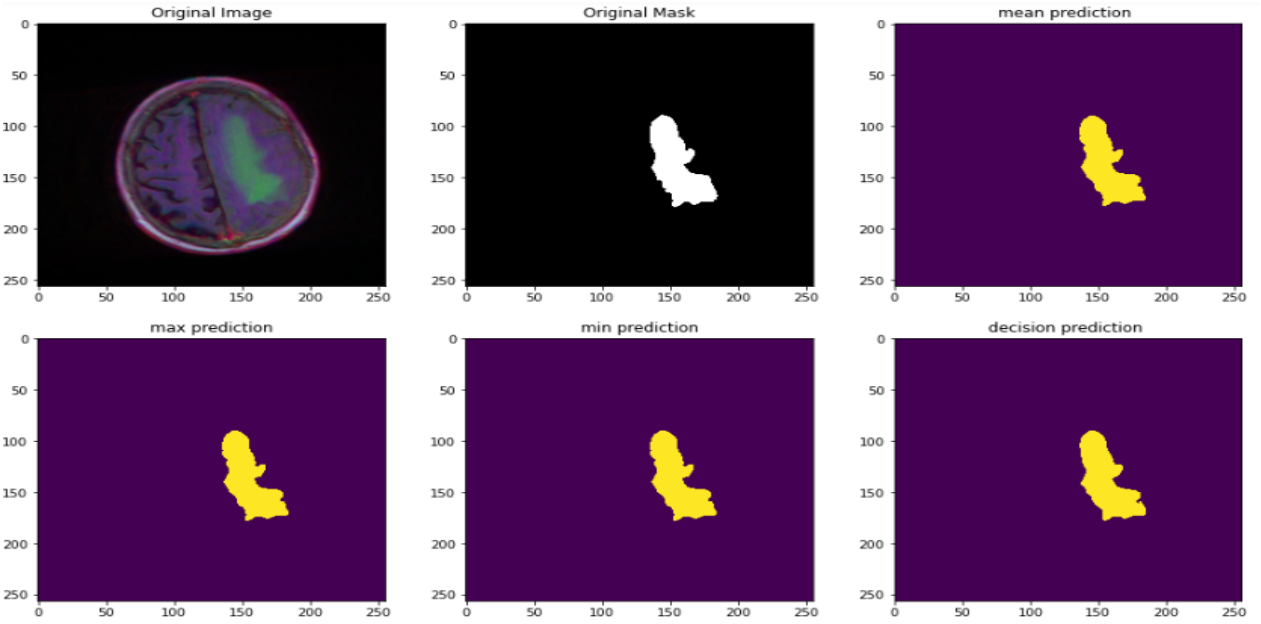

Figure 4.16: Predictions of Ensemble Models on Patient TCGA_HT_7882_19970125 The image displays the segmentation results generated by the ensemble models.

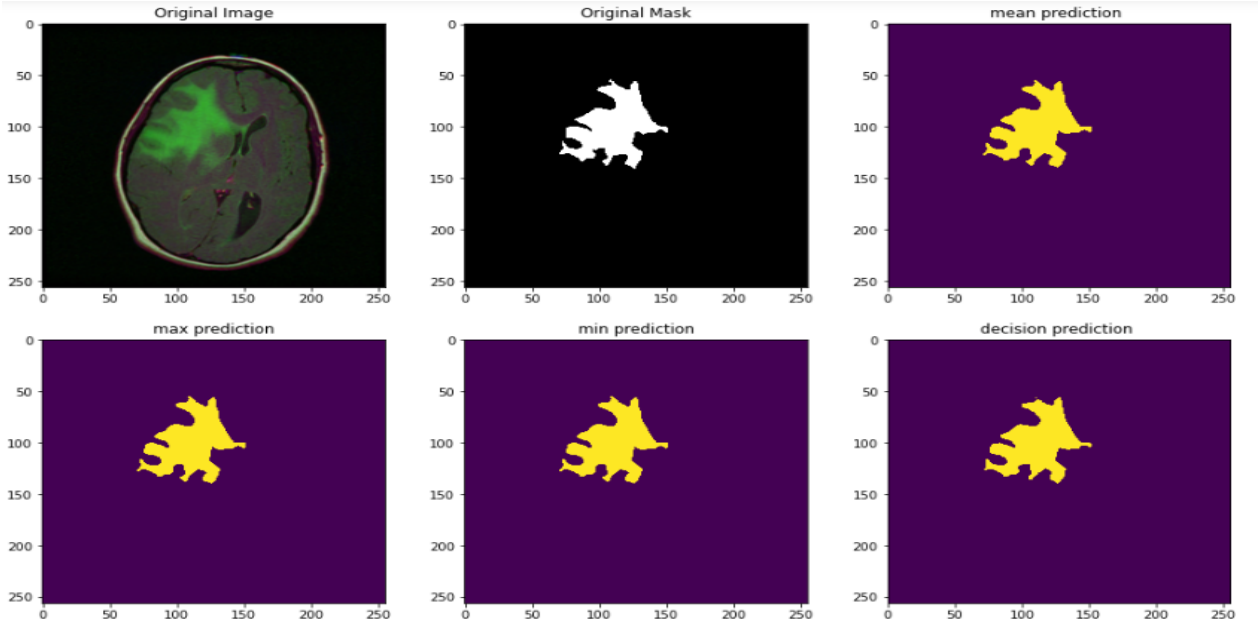

Figure 4.17: Predictions of Ensemble Models on Patient TCGA_DU_6408_19860521 The image displays the segmentation results generated by the ensemble models.

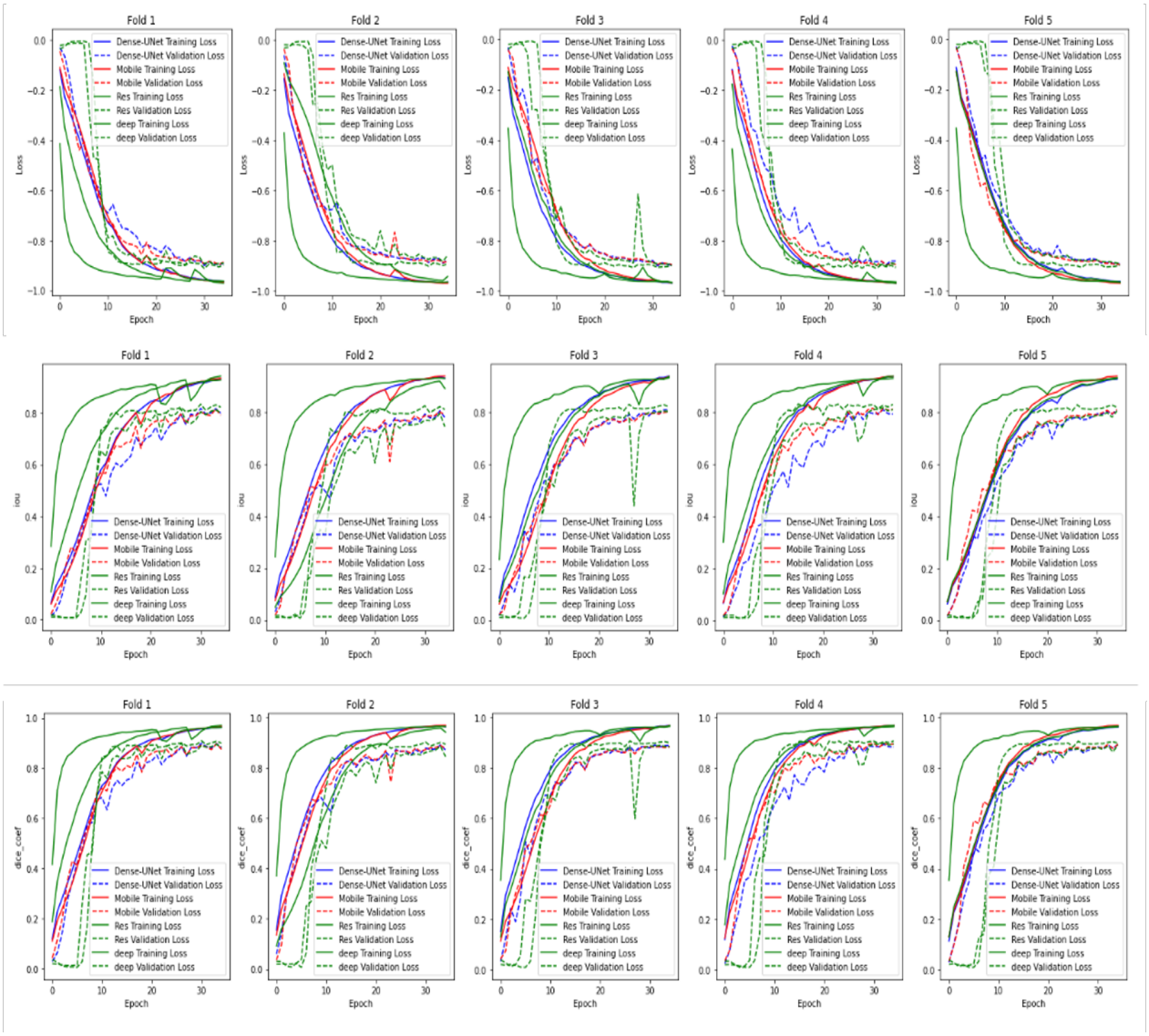

Figure 4.18: Performance comparison through cross-validation of Residual U-Net, Dense U-Net, MobileNetU-Net, and DeepLabV3+ models.

CHAPTER 5

GENERAL CONCLUSION

In This project, we focused basically on developing an approach for segmenting LGG brain MRI images and getting better results compared to what is exist based on U-Net architecture. According to our in-depth review of the literature and reading about many existing methods in the domain we decided to work and explore the U-Net architecture. Also, we reviewed many U-Net variants to take an overview of how researchers think to work with U-Net. since the U-Net is an encoder-decoder architecture we started by trying many architectures as encoder parts with U-Net such as VGG19, ResNet, DenseNet, MobileNet, DeepLabV3+, and many others. Some of the architecture works well and some give us less accurate results, based on that we select the best ones to apply in the ensemble learning technique to get the final result. Also, we presented our methodology step by step and explain all the work that we did and all the concepts that we work on. We present the architectures that we worked on to get the ensemble learning results.

Next, we reviewed the obtained results and all the used materials starting with the used dataset details and information also all the preprocessing steps that we did, and all the tools that we used. Then we presented the evaluation metrics that we used to calculate the accuracy of the obtained segment and to compare the final result of each experiment.

Finally, in the result chapter we present all our obtained results in detail for each architecture. The obtained results were promising thus, the proposed method gives us more accurate segmentations as the ensemble methods using Mean achieve 0.869 IOU, and ensemble methods using Max achieve0.92 Dice-coeff.

5.1 Perspectives

After all the experiments that we did and the obtained results that we get, we have noticed that several ideas can be realized in order to increase the accuracy of segmentation in brain images which consists of building our own model from scratch and trying to train it in a specific context using our own dataset. Also, if we have a big computation station we will try our work in other big datasets such as BraTs, and try many other new architectures like transformers. Also, we can go furthermore and generalize our work to not just do a binary segmentation but for many brain tumor types segmentation and apply it for many applications. Those are some perspectives that may be involved in our future works and research. Finally, we are preparing an interesting scientific paper to close our work with, hoping that we can achieve our goal as soon as possible.

BIBLIOGRAPHY

[1] Depth-wise separable convolutional neural networks, Accessed: 19/05/2023. https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728/.

[2] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[3] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30:449–459, 2017.

[4] Zainab T. Al-Sharify, Talib A. Al-Sharify, Noor T. Al-Sharify, and Husam Yahya naser. A critical review on medical imaging techniques (ct and pet scans) in the medical field. *IOP Conference Series: Materials Science and Engineering*, 870(1):012043, jun 2020.

[5] Esther J. Alberts. *Multi-modal Multi-temporal Brain Tumor Segmentation, Growth Analysis and Texture-based Classification*. PhD thesis, TUM School of Computation, Information and Technology Technische UniversitÃ¤t MÃ¼nchen, 2019.

[6] Jehad Ali, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. Random forests and decision trees. *International Journal of Computer Science Issues(IJCSI)*, 9, 09 2012.

[7] Saqib Ali, Jianqiang Li, Yan Pei, Rooha Khurram, Khalil Ur Rehman, and Tariq Mahmood. A comprehensive survey on brain tumor diagnosis using deep learning and emerging hybrid techniques with multi-modal mr image. *Archives of Computational Methods in Engineering*, 29(7):4871–4896, 2022.

[8] D Anithadevi and K Perumal. A hybrid approach based segmentation technique for brain tumor in mri images. *arXiv preprint arXiv:1603.02447*, 2016.

[9] Reza Azad, Nika Khosravi, Mohammad Dehghanmanshadi, Julien Cohen-Adad, and Dorit Merhof. Medical image segmentation on mri images with missing modalities: A review, 2022.

[10] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[11] Min Bai and Raquel Urtasun. Deep watershed transform for instance segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2858–2866, 2017.

[12] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Ad-

vancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13, 2017.

[13] Hmrishav Bandyopadhyay. An introduction to image segmentation: Deep learning vs. traditional, 2023.

[14] P Gokila Brindha, M Kavinraj, P Manivasakam, and P Prasanth. Brain tumor detection from mri images using deep learning techniques. *IOP Conference Series: Materials Science and Engineering*, 1055(1):012115, feb 2021.

[15] Mateusz Buda. Brain mri segmentation, 2017.

[16] Mateusz Buda, Ashirbani Saha, and Maciej A. Mazurowski. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in Biology and Medicine*, 109:218–225, 2019.

[17] Senay Cakir, Marcel Gauß, Kai Häppeler, Yassine Ounajjar, Fabian Heinle, and Reiner Marchthaler. Semantic segmentation for autonomous driving: Model evaluation, dataset generation, perspective comparison, and real-time capability. *arXiv preprint arXiv:2207.12939*, 2022.

[18] Enrico Calleris. The hourglass network.

[19] Herng-Hua Chang and Daniel J Valentino. An electrostatic deformable model for medical image segmentation. *Computerized Medical Imaging and Graphics*, 32(1):22–35, 2008.

[20] Arkapravo Chattopadhyay and Mausumi Maitra. Mri-based brain tumour image detection using cnn based deep learning method. *Neuroscience Informatics*, 2(4):100060, 2022.

[21] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[22] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR*, abs/1802.02611, 2018.

[23] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[24] Aichi Chien, Bin Dong, and Zuowei Shen. Frame-based segmentation for medical images. *Communications in Mathematical Sciences*, 9(2):551–559, 2011.

[25] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

[26] FranÃ§ois Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, 2017.

[27] Nello Cristianini and Elisa Ricci. *Support Vector Machines*, pages 928–932. Springer US, Boston, MA, 2008.

[28] Soumyya Kanti Datta, Mohammad Abuzar Shaikh, Sargur N Srihari, and Mingchen Gao. Soft attention improves skin cancer classification performance. In *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4*, pages 13–23. Springer, 2021.

[29] Nameirakpam Dhanachandra, Khumanthem Manglem, and Yambem Jina Chanu. Image segmentation using k -means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54:764–771, 2015. Eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Data Mining and Warehousing, ICDMW 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Image and Signal Processing, ICISP 2015, August 21-23, 2015, Bangalore, India.

[30] Khaoula Drid, Mebarka Allaoui, and Mohammed Lamine Kherfi. Object detector combination for increasing accuracy and detecting more overlapping objects. In *Image and Signal Processing: 9th International Conference, ICISP 2020, Marrakesh, Morocco, June 4â6, 2020, Proceedings*, page 290â296, Berlin, Heidelberg, 2020. Springer-Verlag.

[31] Somayeh Ebrahimkhani, Mohamed Hisham Jaward, Flavia M. Cicuttini, Anuja Dharmaratne, Yuanyuan Wang, and Alba G. Seco de Herrera. A review on segmentation of knee articular cartilage: from conventional methods towards deep learning. *Artificial Intelligence in Medicine*, 106:101851, 2020.

[32] Omar Elharrouss, Somaya Al-Maadeed, Nandhini Subramanian, Najmath Ottakath, Noor Almaadeed, and Yassine Himeur. Panoptic segmentation: a review. *arXiv preprint arXiv:2111.10250*, 2021.

[33] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61:55–79, 2004.

[34] U.S. Food and Drug Administration(FDA). medical-imaging, 2023.

[35] M.A. Ganaie, Minghui Hu, A.K. Malik, M. Tanveer, and P.N. Suganthan. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115:105151, 2022.

[36] Shangshang Gao, Yuanyuan Wang, Zhaofeng Chen, Feng Zhou, Rugang Wang, and Naihong Guo. Design and implementation of local threshold segmentation based on fpga. *Journal of Electrical and Computer Engineering*, 2022, 2022.

[37] Mina Ghaffari, Arcot Sowmya, and Ruth Oliver. Automated brain tumor segmentation using multimodal brain scans, a survey based on models submitted to the brats 2012-18 challenges.

[38] Nelly Gordillo, Eduard Montseny, and Pilar Sobrevilla. State of the art survey on mri brain tumor segmentation. *Magnetic resonance imaging*, 31(8):1426–1438, 2013.

[39] Gongde Guo, Hui Wang, David Bell, and Yaxin Bi. Knn model-based approach in classification. 08 2004.

[40] Bhat G.M. Hafiz, A.M. A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval volume*, 9:171â189, 2020.

[41] Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Yuan Yao, Ao Zhang, Liang Zhang, Wentao Han, Minlie Huang, Qin Jin, Yanyan Lan, Yang Liu, Zhiyuan Liu, Zhiwu Lu, Xipeng Qiu, Ruihua Song, Jie Tang, Ji-Rong Wen, Jinhui Yuan, Wayne Xin Zhao, and Jun Zhu. Pre-trained models: Past, present and future. *AI Open*, 2:225–250, 2021.

[42] Hasty AI. Deeplabv3+. Website, Accessed: May 14, 2023. https://hasty.ai/docs/mp-wiki/model-architectures/deeplabv3.

[43] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[44] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[45] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[46] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[47] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6:107–116, 04 1998.

[48] Komal R Hole, Vijay S Gulhane, and Nitin D Shellokar. Application of genetic algorithm for image enhancement and segmentation. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2(4):1342, 2013.

[49] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, abs/1704.04861, 2017.

[50] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[51] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[52] Ahmed Iqbal, Muhammad Sharif, Mussarat Yasmin, Mudassar Raza, and Shabib Aftab. Generative adversarial networks and its applications in the biomedical image segmentation: a comprehensive survey. *International Journal of Multimedia Information Retrieval*, 11(3):333–368, 2022.

[53] Alipujiang Jierula, Shuhong Wang, Tae-Min OH, and Pengyu Wang. Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data. *Applied Sciences*, 11(5), 2021.

[54] Hassana Grema Kaganami and Zou Beiji. Region-based segmentation versus edge detection. In *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 1217–1221. IEEE, 2009.

[55] Siva Teja Kakileti, Geetha Manjunath, and Himanshu J. Madhu. Cascaded cnn for view independent breast segmentation in thermal images. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6294–6297, 2019.

[56] Wen-Xiong Kang, Qing-Qiang Yang, and Run-Peng Liang. The comparative research on image segmentation algorithms. In *2009 First International Workshop on Education Technology and Computer Science*, volume 2, pages 703–707, 2009.

[57] Nirpjeet Kaur and Rajpreet Kaur. A review on various methods of image thresholding. *International Journal on Computer Science and Engineering*, 3(10):3441, 2011.

[58] Waseem Khan. Image segmentation techniques: A survey. *Journal of image and graphics*, 1(4):166–170, 2013.

[59] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9404–9413, 2019.

[60] Ihor Konovalenko, Pavlo Maruschak, Janette BrezinovÃ¡, Olegas Prentkovskis, and Jakub Brezina. Research of u-net-based cnn architectures for metal surface defect detection. *Machines*, 10(5), 2022.

[61] Mandar Kulkarni and Aria Abubakar. Soft attention convolutional neural networks for rare event detection in sequences. *arXiv preprint arXiv:2011.02338*, 2020.

[62] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[63] Lagergren and Rosengren. Masterâs thesis in biomedical engineering, 2020.

[64] Joseph S. Lappin and Herbert H. Bell. Form and function in information for visual perception. *i-Perception*, 12(6):20416695211053352, 2021. PMID: 35003612.

[65] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[66] Chi-Hoon Lee, Mark Schmidt, Albert Murtha, Aalo Bistritz, Jöerg Sander, and Russell Greiner. Segmenting brain tumors with conditional random fields and support vector machines. In *Computer Vision for Biomedical Image Applications: First International Workshop, CVBIA 2005, Beijing, China, October 21, 2005. Proceedings 1*, pages 469–478. Springer, 2005.

[67] Bing Nan Li, Chee Kong Chui, Stephen Chang, and Sim Heng Ong. Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation. *Computers in biology and medicine*, 41(1):1–10, 2011.

[68] Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji, and Yichen Wei. Fully convolutional instance-aware semantic segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4438–4446, 2017.

[69] Jin Liu, Min Li, Jianxin Wang, Fangxiang Wu, Tianming Liu, and Yi Pan. A survey of mri-based brain tumor segmentation methods. *Tsinghua science and technology*, 19(6):578–595, 2014.

[70] Zhihua Liu, Lei Tong, Long Chen, Zheheng Jiang, Feixiang Zhou, Qianni Zhang, Xiangrong Zhang, Yaochu Jin, and Huiyu Zhou. Deep learning based brain tumor segmentation: a survey. *Complex & Intelligent Systems*, pages 1–26, 2022.

[71] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[72] David N Louis, Arie Perry, Guido Reifenberger, Andreas Von Deimling, Dominique Figarella-Branger, Webster K Cavenee, Hiroko Ohgaki, Otmar D Wiestler, Paul Kleihues, and David W Ellison. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta neuropathologica*, 131:803–820, 2016.

[73] JW Lu, C Ma, C Zuo, JP Guillemin, P Gouton, and JC Coquille. Distinguishing onions and weeds in field by using color image. *Transactions of the CSAE*, 17(5):153–158, 2001.

[74] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[75] Tirivangani Magadza and Serestina Viriri. Deep learning for brain tumor segmentation: a survey of state-of-the-art. *Journal of Imaging*, 7(2):19, 2021.

[76] Priyanka Malhotra, Sheifali Gupta, Deepika Koundal, Atef Zaguia, and Wegayehu Enbeyle. Deep neural networks for medical image segmentation. *Journal of Healthcare Engineering*, 2022, 2022.

[77] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[78] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[79] Kyoungwook Min, Seung-Jun Han, Dongjin Lee, Dooseop Choi, Kyungbok Sung, and Jeongdan Choi. Sae level 3 autonomous driving technology of the etri. *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 464–466, 2019.

[80] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey, 2020.

[81] Ajay Mishra, Yiannis Aloimonos, and Cornelia Fermuller. Active segmentation for robotics. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3133–3139, 2009.

[82] Dhanunjaya Mitta, Soumick Chatterjee, Oliver Speck, and Andreas NÃ¼rnberger. Upgraded w-net with attention gates and its application in unsupervised 3d liver segmentation. 02 2021.

[83] Himanshu Mittal, Avinash Chandra Pandey, Mukesh Saraswat, Sumit Kumar, Raju Pal, and Garv Modwel. A comprehensive survey of image segmentation: clustering

methods, performance parameters, and benchmark datasets. *Multimedia Tools and Applications*, pages 1–26, 2021.

[84] Yujian Mo, Yan Wu, Xinneng Yang, Feilin Liu, and Yujun Liao. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493:626–646, 2022.

[85] Yahya Mohammed, Said El Garouani, and Ismail Jellouli. A survey of methods for brain tumor segmentation based mri images. *Journal of Computational Design and Engineering*, 2022.

[86] Shi Na, Liu Xumin, and Guan Yong. Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, pages 63–67, 2010.

[87] Mohamed A. Naser and M. Jamal Deen. Brain tumor segmentation and grading of lower-grade glioma using deep learning in mri images. *Computers in Biology and Medicine*, 121:103758, 2020.

[88] Janmenjoy Nayak, Bighnaraj Naik, and HSr Behera. Fuzzy c-means (fcm) clustering algorithm: a decade review from 2000 to 2014. In *Computational Intelligence in Data Mining-Volume 2: Proceedings of the International Conference on CIDM, 20-21 December 2014*, pages 133–149. Springer, 2015.

[89] Tri Nguyen and Myungsik Yoo. Fusing lidar sensor and rgb camera for object detection in autonomous vehicle with fuzzy logic approach. In *2021 International Conference on Information Networking (ICOIN)*, pages 788–791, 2021.

[90] Zhen-Liang Ni, Gui-Bin Bian, Xiao-Hu Zhou, Zeng-Guang Hou, Xiao-Liang Xie, Chen Wang, Yan-Jie Zhou, Rui-Qi Li, and Zhen Li. Raunet: Residual attention u-net for semantic segmentation of cataract surgical instruments, 2019.

[91] Jakhongir Nodirov, Akmalbek Bobomirzaevich Abdusalomov, and Taeg Keun Whangbo. Attention 3d u-net with multiple skip connections for segmentation of brain tumor images. *Sensors*, 22(17), 2022.

[92] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[93] Jay Patel and Kaushal Doshi. A study of segmentation methods for detection of tumor in brain mri. *Advance in Electronic and Electric Engineering*, 4(3):279–284, 2014.

[94] Chiara Piazzalunga, Linda Greta Dui, Cristiano Termine, Marisa Bortolozzo, Matteo Matteucci, and Simona Ferrante. Investigating visual perception impairments through serious games and eye tracking to anticipate handwriting difficulties. *Sensors*, 23(4), 2023.

[95] R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45, 2006.

[96] Robi Polikar. Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, 6(3):21–45, 2006.

[97] SV Kasmir Raja, A SHAIK ABDUL KHADIR, and SS Riaz Ahamed. Moving toward region-based image segmentation techniques: A study. *Journal of Theoretical & Applied Information Technology*, 5(1), 2009.

[98] Vishal Rajput. Attention u-net, resunet, many more, 2022.

[99] Payam Refaeilzadeh, Lei Tang, and Huan Liu. *Cross-Validation*, pages 532–538. Springer US, Boston, MA, 2009.

[100] Seyed Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian D. Reid, and Silvio Savarese. Generalized intersection over union: A metric and A loss for bounding box regression. *CoRR*, abs/1902.09630, 2019.

[101] Giulia Rizzoli, Francesco Barbato, and Pietro Zanuttigh. Multimodal semantic segmentation in autonomous driving: A review of current approaches and future perspectives. *Technologies*, 10(4), 2022.

[102] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[103] Spencer Rose. *An evaluation of deep learning semantic segmentation for land cover classification of oblique ground-based photography*. PhD thesis, 2020.

[104] Omer Sagi and Lior Rokach. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1249, 2018.

[105] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[106] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2019.

[107] Jo Schlemper, Jose Caballero, Joseph V. Hajnal, Anthony N. Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2018.

[108] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, 53:197–207, 2019.

[109] Joan Serra, Didac Suris, Marius Miron, and Alexandros Karatzoglou. Overcoming catastrophic forgetting with hard attention to the task. In *International Conference on Machine Learning*, pages 4548–4557. PMLR, 2018.

[110] Reuben R. Shamir, Yuval Duchin, Jinyoung Kim, Guillermo Sapiro, and Noam Harel. Continuous dice coefficient: a method for evaluating probabilistic segmentations. *CoRR*, abs/1906.11031, 2019.

[111] Amit Sharma, Pradeep Kumar Singh, and Yugal Kumar. An integrated fire detection system using iot and image processing technique for smart cities. *Sustainable Cities and Society*, 61:102332, 2020.

[112] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9:82031–82057, 2021.

[113] Nahian Siddique, Paheding Sidike, Colin Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: theory and applications. *arXiv preprint arXiv:2011.01118*, 2020.

[114] Erika Siegel, Jolie Wormwood, Karen Quigley, and Lisa Barrett. Seeing what you feel: Affect drives visual perception of structurally neutral faces. *Psychological Science*, 29:095679761774171, 02 2018.

[115] sik Ho Tsang. Review: V-net â volumetric convolution (biomedical image segmentation).

[116] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[117] Pavel Škrabánek and Alexandra Zahradnikova Jr. Automatic assessment of the cardiomyocyte development stages from confocal microscopy images using deep convolutional networks. *Plos one*, 14(5):e0216720, 2019.

[118] Sovi Guillaume Sodjinou, Vahid Mohammadi, Amadou Tidjani Sanda Mahama, and Pierre Gouton. A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images. *Information Processing in Agriculture*, 9(3):355–364, 2022.

[119] Derya Soydaner. Attention mechanism in neural networks: where it comes and where it goes. *Neural Computing and Applications*, 34(16):13371–13385, 2022.

[120] Jasjit Suri, Mrinalini Bhagawati, Sushant Agarwal, Sudip Paul, Amit Pandey, Suneet Gupta, Luca Saba, Kosmas Paraskevas, Narendra Khanna, John Laird, Amer Johri, Manudeep Kalra, Mostafa Fouda, Mostafa Fatemi, and Desineni Naidu. Unet deep learning architecture for segmentation of vascular and non-vascular images: A microscopic look at unet components buffered with pruning, explainable artificial intelligence, and bias. *IEEE Access*, PP:1–1, 01 2022.

[121] Nikhil Tomar. What is unet?, 2021.

[122] Thanh N Tran, Ron Wehrens, and Lutgarde MC Buydens. Clustering multispectral images: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 77(1-2):3–17, 2005.

[123] Shih-Yao Tseng. Review: Deeplabv3+ (atrous separable convolution semantic segmentation), [Online; Accessed: May 14, 2023]. https://sh-tsang.medium.com/review-deeplabv3-atrous-separable-convolution-semantic-segmentation-a625f6e83b90.

[124] Ranjith Unnikrishnan, Caroline Pantofaru, and Martial Hebert. A Measure for Objective Evaluation of Image Segmentation Algorithms. 1 2005.

[125] Vinorth Varatharasan, Hyo-Sang Shin, Antonios Tsourdos, and Nick Colosimo. Improving learning effectiveness for object detection and classification in cluttered backgrounds. pages 78–85, 11 2019.

[126] Robin Vinod. A detailed explanation of the attention u-net, 2020.

[127] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation. *Frontiers in computational neuroscience*, 13:56, 2019.

[128] Yu Wang, Quan Zhou, Jia Liu, Jian Xiong, Guangwei Gao, Xiaofu Wu, and Longin Jan Latecki. Lednet: A lightweight encoder-decoder network for real-time semantic segmentation. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1860–1864, 2019.

[129] Li Wen and Michael Hughes. Coastal wetland mapping using ensemble learning algorithms: A comparative study of bagging, boosting and stacking techniques. *Remote Sensing*, 12(10):1683, 2020.

[130] Dongxian Wu, Yisen Wang, Shu-Tao Xia, James Bailey, and Xingjun Ma. Skip connections matter: On the transferability of adversarial examples generated with resnets. *arXiv preprint arXiv:2002.05990*, 2020.

[131] Jialian Wu, Jiale Cao, Liangchen Song, Yu Wang, Ming Yang, and Junsong Yuan. Track to detect and segment: An online multi-object tracker. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12352–12361, 2021.

[132] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057. PMLR, 2015.

[133] Z Ye, H Li, W Zha, et al. A visual detection method of tool damage using local threshold segmentation. *Hsi-An Chiao Tung Ta Hsueh/Journal of Xiâan Jiaotong University*, 55(4):52–60, 2021.

[134] Serdar Yegulalp. What is tensorflow? the machine learning library explained, [Online; accessed May 18, 2023]. https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html.

[135] Cha Zhang and Yunqian Ma. *Ensemble machine learning: methods and applications*. Springer, 2012.

[136] Chaoning Zhang, Philipp Benz, Dawit Mureja Argaw, Seokju Lee, Junsik Kim, Francois Rameau, Jean-Charles Bazin, and In So Kweon. Resnet or densenet? introducing dense shortcuts to resnet. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3550–3559, 2021.

[137] Jianxin Zhang, Zongkang Jiang, Jing Dong, Yaqing Hou, and Bin Liu. Attention gate resu-net for automatic mri brain tumor segmentation. *IEEE Access*, 8:58533–58545, 2020.

[138] Shihao Zhang, Huazhu Fu, Yuguang Yan, Yubing Zhang, Qingyao Wu, Ming Yang, Mingkui Tan, and Yanwu Xu. Attention guided network for retinal image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 797–805. Springer, 2019.

[139] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. *CoRR*, abs/1707.01083, 2017.

[140] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.