Identification of Non-Enhancing Brain Tumor Regions Using an Upscaling Pre-Activation U-Net with Depth Regularization

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
der Fakultät für Physik der Universität Regensburg

vorgelegt von

Tobias Schaffer

aus Mühldorf am Inn

im Jahr 2024

Diese Arbeit wurde angeleitet von Prof. Dr. rer. nat. Elmar Lang

Das Promotionsgesuch wurde eingereicht am 15.01.2024

Prüfungsausschuss

Vorsitzender:	Prof. Dr. K. Rincke
1. Gutachter:	Prof. Dr. E. Lang
2. Gutachter:	Dr. S. Solbrig
Weiterer Prüfer:	Prof. Dr. T. Wettig

Abstract

The challenge of identifying and analyzing non-enhancing tumor regions (NET) presents a critical frontier in radiology. The focus of this work is on the identification of these NET tumor areas, which can be crucial in understanding tumor progression and therapeutic response. This study addresses the detection of structural aspects of NET in MRI records of glioblastoma, a special form of brain cancer, known for its aggressive behavior and challenging treatment. Leveraging an innovative machine learning model, this work aims to enhance the detection and characterization of the NET regions. The approach will work towards improving diagnostic accuracy and can aid in developing more effective treatment strategies.

The presented neural net architecture will be based on the classical U-Net, a successful segmentation model in medical domains and beyond. This work will add an upscaling extension in the decoder path which enables the neural net to generate resolution enhanced segmentation masks. The resulting upscaling pre-activation U-Net will further be trained on a subset of the BraTS 2018 dataset after reconstructing ground-truth data for NET tumor regions.

Acknowledgements

First and foremost, I would like to give my gratitude to my advisor Prof. Dr. Elmar Lang for his encouragement, assistance and supervision during all the time of this work. He gave me the great opportunity to work in his rich field of research and supported my way towards academic work. His strong interdisciplinary expertise combined with his patience and care had a deep impact on me. Thank you so much for giving me your trust, your time and your exceptional support.
I also want to express my deep thanks to Prof. Dr. Alexander Brawanski for his motivation towards this field of research and his invaluable insights into the medical and clinical domain. Thank you for sharing your deep expertise and knowledge in the neurological field, for your advice in the statistical analysis as well as for your feedback on the machine learning and coding applications.
A special thank goes to my parents for their mental support and for believing in this project. Thank you for your strong faith and confidence in all my way here.
Also I want to thank my children for giving me their charm and trust. Thank you so much for your optimism, your interest and your encouragement, all of this gave me a continuous reason for working towards the goal.
Finally, I want to give a major thank to my wife for her long lasting support during the last years, for having me throughout demanding times and for giving me the motivation to finalize this work. Without your strong backup and patience, this and many other things would never be possible.

Table of Contents

List of Tables

List of Figures

1 Introduction

This work will present a comprehensive and detailed exploration of non-enhancing tumor (NET) regions in glioblastoma, utilizing advanced MRI techniques and machine learning models.

Therefore Section 2 will introduce the relevant pathological regions, which can typically be observed in brain tumors. A focus throughout this work and also in Section 2 will be the non-enhancing brain tumor parts. These are demanding in their identification and are gaining increased attention when evaluating tumors in prognostic and treatment aspects.

In Section 3 the physical foundation of magnetic resonance imaging (MRI) together with different conventional MRI sequences will be introduced. Further advancements in magnetic resonance imaging have opened new ways for the detailed study of tumors and their structures. Section 3 will also introduce state-of-the-art MRI techniques, which aim to uncover subtle but significant aspects of NET regions, which traditional imaging methods might miss. This advanced detection capability is an important step towards a more accurate diagnosis and has the potential to verify and facilitate automated detection based on conventional imaging methods together with sophisticated machine learning methods.

Furthermore, latest machine learning models to analyze the vast and complex data obtained from MRI scans will be presented and discussed in Section 4. These models are designed to identify patterns and characteristics within the NET regions that are not immediately apparent to the human eye. This approach not only aids in better understanding of the tumor's structure and progression but also has the potential to predict or interpret treatment responses more accurately.

The MICCAI BraTS challenges will serve as a major resource of training data for these machine learning models. Section 5 will describe the datasets of the recent years and will outline NET data being implicitly contained in some of this data. Reconstructing and cleaning NET segmentation data from BraTS challenge records offers a promising potential for obtaining training data for the automated identification of non-enhancing brain tumor parts. Generally, machine learning models and neural networks have high demand of training data for providing adequate results and thus the availability of extensive amount of training data is essential.

An enhanced U-Net model will be presented in Section 6 which will be modified towards an improved detection of small scaled NET parts. The architecture will be developed and tested using the original BraTS data. For building a training data basis serving the NET detection, the BraTS data will be processed using morphological filters for extracting and isolating non-enhancing ground truth data. This will allow to refine BraTS labels and to make BraTS data from different years compatible to each other. The formerly incompatible and strictly separated BraTS datasets of different

years can then be unified to a large combined dataset.

Section 7 will show and discuss the results of the previous section. Selected examples segmentations will be discussed and compared towards the characteristic properties of NET in MRI images given in Sections 2 and 3. Statistical analysis of the volume distributions of the automatically segmented NET regions will be presented, as well as an ANOVA analysis of the intensity distributions of the different tumor region classes, which will give significant results. The novel U-Net model will be trained on the new unified large dataset and will be able to reach competitive segmentation accuracy while also detecting NET regions. Finally, 3D models based on the automatically generated segmentation data will be presented showing typical characteristic features of non-enhancing tumor areas.

2 Brain Tumor Types and Regions

Brain tumors are abnormal growths of cells in the brain and can be categorized into various types based on their nature, origin, and malignancy level. Gliomas are one type of tumor that starts in the glial cells, which are cells in the brain and spinal cord that support and protect nerve cells. They are classified into grades based on their rate of growth and how the cells appear on a microscopic level. High-grade gliomas (Grade III and IV) are more malignant and grow faster than low-grade gliomas (Grade I and II). Among these, glioblastomas represent the most aggressive form of brain cancer, categorized as a high-grade tumor due to its rapid growth and the strength of symptoms they can cause. The World Health Organization's classification of tumors of the central nervous system [75] is a widely accepted resource for organizing and understanding the different types of brain tumor regions, as well as radiology textbooks such as "Osborn's Brain" by Anne G. Osborn [93]. Glioblastoma tumors are classified as Grade IV under this classification system and are characterized by distinct regions that are identified on medical imaging, such as MRI. These regions are:

- Edema *ED*
- Necrosis *NCR*
- Enhancing Tumor *ET*
- Non-Enhancing Tumor *NET*

The treatment of glioblastoma typically consists of multiple approaches including surgery, radiation therapy, and chemotherapy. However, due to their highly invasive nature, complete surgical removal is challenging, and the disease often recurs. Understanding the different parts and MRI imaging characteristics of glioblastoma is crucial for clinicians to plan the best treatment strategies for patients. The complexity of glioblastoma treatment and management underscores the need for ongoing research and the development of novel diagnostic options.

The following paragraphs will summarize each of the above tumor region's medical characteristics. Section 3.4 in the later course, will show up radiology imaging features helping to identify and differentiate them.

2.1 Edema

In general there are various types of edema, but the main type of edema associated with brain tumors is the so called vasogenic edema. This is mainly caused by the disruption of the blood-brain barrier, which is a protective layer that normally prevents the entry of substances from the bloodstream into the brain tissue. Brain tumors, particularly high-grade and malignant tumors, can disrupt this barrier, allowing plasma proteins and fluid to leak into other parts of the brain.

The edema area generally surrounds the tumor and can extend into the neighbouring white matter, following the natural pathways of fluid movement within the brain. Image (A) in Figure 1 shows a typical example of edema. The higher quantity of fluid in the brain tissue leads to an increase in volume, which can result in an increase in intracranial pressure. This can cause a variety of symptoms, including headaches, nausea or vomiting. Additionally, the edema can compress surrounding brain structures, which can lead to neurological deficits, such as weakness, numbness, or difficulty speaking, depending on the area of the brain that is affected.

For the treatment of brain tumor-related edema, corticosteroids are commonly used to reduce inflammation and decrease the edema extent. These medications can lead to a rapid improvement in symptoms, although they do have potential side effects, especially with long-term use. In some cases, surgical intervention to resect the tumor or relieve pressure may also be necessary. Furthermore, addressing the underlying tumor through surgery, radiation therapy, or chemotherapy is a crucial part of managing brain tumor-related edema.

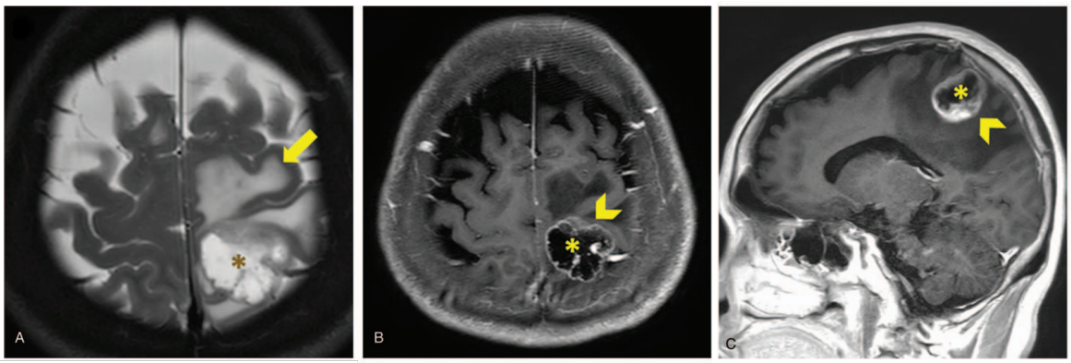

Figure 1: Representative MR images showing tumor necrosis, rim enhancement, peri-tumoral edema, and hemorrhage. Axial T2-weighted MRI image (A) shows central necrosis (asterisk) and peritumoral vasogenic edema (arrow). Axial (B) and sagittal (C) contrast-enhanced T1-weighted MRI image shows central necrosis (asterisk) and rim enhancement (arrowhead). Figure from [68].

2.2 Necrosis

Necrosis within a brain tumor refers to the death of cells in the tissue, which can occur for various reasons, including insufficient blood supply and lack of oxygen, or the effects of a tumor treatment like radiation. Necrosis is a characteristic for high-grade, aggressive tumors, but it can also be seen in lower-grade tumors under certain conditions. Figure 1 image (A) gives an example of necrosis showing in an axial T2 image.

The presence of necrosis in a brain tumor is generally an unfavorable prognostic factor, as it is associated with more aggressive tumor behavior and a poorer response to treatment. High-grade gliomas mostly have extensive necrosis, and their prognosis is typically poor. Necrosis can also have implications for treatment planning.

For example, in radiation therapy, the presence of necrosis may influence the spatial limits of the target area for treatment. Additionally, in cases where surgery is considered, the surgeon must carefully plan the approach to avoid causing additional damage to the surrounding healthy brain tissue. The treatment of brain tumors with necrotic regions typically involves a combination of surgery, radiation therapy, and chemotherapy. The goal of surgery is often to reduce the tumor's size, while radiation and chemotherapy aim to target the remaining tumor cells.

In some cases, particularly in high-grade gliomas, the treatment itself can cause necrosis in the tumor, which can sometimes make it challenging to distinguish between treatment effects and tumor progression on medical imaging.

2.3 Enhancing Tumor

The enhancing parts of a brain tumor on radiology imaging refer to regions that show reception of contrast material, indicating a disruption of the blood-brain barrier. These areas typically represent active and viable tumor tissue and their evaluation is important for diagnosis, treatment planning, and monitoring the response to applied therapy.

On magnetic resonance imaging, enhancing tumor regions become more visible after the administration of intravenous contrast substances, such as gadolinium. The enhancement effects can be different and my be grouped as follows:

- **Solid Enhancement**: This pattern is typical for dense cellular tumor regions with a disrupted blood-brain barrier. It can be seen in high-grade gliomas, metastases and several other types of brain tumors.
- **Rim Enhancement**: This occurs when there is central necrosis surrounded by a ring of enhancing tumor. This pattern is often associated with high-grade gliomas, but can also be seen in brain abscesses, necessitating careful differential diagnosis, as shown in Figure 1 image (B).
- **Heterogeneous Enhancement**: Some tumors show a combination of solid and rim enhancement or an irregular enhancement pattern, reflecting variable tumor density and vascularity.

The presence and pattern of enhancement have significant clinical implications. Enhancing tumor parts are often associated with higher-grade, more aggressive tumors, similar to what was said about necrosis. Understanding the extent of the enhancing tumor is crucial for surgical planning, as the goal is often to achieve a maximized, yet safe resection. It also helps in planning radiation therapy. Generally, tumors with more extensive enhancement have a worse prognosis, as they are likely to be higher-grade and more aggressive. In the treatment phase, monitoring changes in the enhancing parts of a tumor on MRI can help to judge the response to therapy. Decreased enhancement might suggest a positive response, while increased or new

enhancement could indicate tumor progression or recurrence.

The treatment is often to resect as much of the enhancing tumor as possible without causing significant neurological damage. In some cases, complete resection of enhancing tumor can lead to improved outcomes. Additionally, radiation therapy is used to target residual enhancing tumor after surgery (or as a primary treatment in cases where surgery is not possible). Also chemotherapy can target enhancing tumor regions, and this can be an important part of the treatment, especially for specific tumor types like glioblastoma.

2.4 Non-Enhancing Tumor

The non-enhancing part of a brain tumor (NET) on medical imaging refers to regions that do not show consumption of contrast material and thus do not show enhancement on contrast-enhanced magnetic resonance imaging scans. These areas may represent a variety of tissue types and tumor cells, and their interpretation requires careful consideration of the clinical context and imaging characteristics.

On MRI scans, high-grade gliomas typically show a heterogeneous pattern of contrast enhancement, which indicates the presence of a disrupted blood-brain barrier in the tumor. However, low-grade gliomas may not show any enhancement on MRI scans, and instead, they may have a non-enhancing component that appears hypointense on T1-weighted images and hyperintense on T2-weighted images (see section 3.4). Moreover, in all glioma types non-enhancing regions may represent edema, infiltrative tumor cells, or a combination of both. Distinguishing between these possibilities can be very challenging, but is crucial for accurate tumor grading and treatment planning.

The non-enhancing component of gliomas represents a region of the tumor that has a lower density of tumor cells and a higher density of normal brain tissue. The non-enhancing region may contain infiltrative tumor cells that are diffusely spread out and cannot be detected by contrast-enhanced MRI scans. These infiltrative cells can migrate into the surrounding brain tissue and cause tumor recurrence after surgery [65]. This makes the the non-enhancing parts very crucial for the evaluation of cancer progression and also for it's potential recurrence after resection.

The clinical implications of non-enhancing parts of a brain tumor depend on their underlying nature. If the non-enhancing region represents infiltrative tumor growth, it is indicative of a more diffuse and potentially aggressive tumor. For these areas a complete surgical resection might not be possible, and they can significantly influence the treatment plan and prognosis. The non-enhancing region is an important area to consider in the treatment of brain tumors because it can be a source of diagnostic and therapeutic challenges. Biopsies of the non-enhancing region can be difficult to obtain because the region may not show any contrast enhancement. Therefore, image-guided biopsy or functional imaging techniques may be hard to use to target

the non-enhancing region for sampling.

The treatment of brain tumors that involve a non-enhancing part can also be challenging because the region may contain infiltrative tumor cells that cannot be removed by surgery. Radiation- or chemotherapy may be used to target the non-enhancing region and reduce the risk of tumor recurrence.

For patients undergoing a brain tumor treatment, the reduction or at least stability of non-enhancing regions can be indicative of a positive response to therapy. Conversely, the expansion of non-enhancing areas might suggest tumor progression. The treatment of non-enhancing parts of a brain tumor is tailored to the individual patient and depends on factors such as the tumor type, location, and overall extent of the disease:

- **Observation**: In cases where the non-enhancing region is believed to represent a lower-grade or less aggressive part of the tumor, and it is not causing significant symptoms, observation with regular follow-up imaging might be an appropriate approach.
- **Surgery**: If the non-enhancing region is accessible and causing symptoms, surgical intervention might be considered. The goal of surgery is often to achieve maximum safe resection while preserving neurological function.
- **Other therapy**: Radiation therapy and chemotherapy might be used to target infiltrative tumor cells in non-enhancing regions, particularly if these areas are not accessible to surgical resection.

In order to ensure appropriate treatment, **reliable detection of non-enhancing tumor parts** must be possible and requires a comprehensive understanding of the imaging findings. For surgical brain tumor treatment, the role of the non-enhancing tissue comes into focus as a border region for resection, as stated by Brenner et al. [17]:

> *We found that new contrast enhancement appeared within the residual, non-enhancing tumor mass in 21 of 24 patients (87.5%). The location of new contrast enhancement within the residual tumor region was nonrandom; it occurred adjacent to the wall of the resection cavity in 12 of 21 patients (57.1%).*

and also by Lasocki et. al [65]:

> *There is a growing understanding of the prognostic importance of noncontrast-enhancing tumor in glioblastoma, and recent attempts at more aggressive management of this component using neurosurgical resection and radiosurgery have been shown to prolong survival.*

For finding appropriate resection borders of a brain tumor, the detection of the non-enhancing area, which typically adjoins and extends the necrotic and enhancing tumor core regions is therefore crucial.

3 Magnetic Resonance Imaging MRI

3.1 Physics of MRI Imaging

Magnetic Resonance Imaging (MRI) is a medical imaging technique that uses a magnetic field and radio waves to create detailed images of the inside of the body. The basic principle of MRI is that when a person is placed in a strong magnetic field, the nuclear spin moments of the hydrogen atoms in their body align towards the magnetic field. Radio waves are then used to disrupt this alignment, causing the atoms to release energy in the form of a radio frequency signal. This signal is then detected by a receiver coil and used to create an image of the body. The produced image can be highly detailed and is used to identify a wide range of abnormalities, including injuries, diseases of inner organs or, as in the scope of this work, tumors with a focus on brain tumors and especially glioblastomas [77, 80].

Magnetic fields can be described by two distinct, but closely related physical entities, magnetic flux density B (measured in Tesla T) and magnetic field strength H (measured in A/m). In vacuum they are connected through the vacuum magnetic permeability μ_0

$$B = \mu_0 H \tag{1}$$

while in other materials they can be related linearly or non-linearly depending on the material's magnetic properties. In a less-strict sense the term *magnetic field* in this work will refer to both interchangeably.

For creating image data, MRI utilizes the magnetic properties of hydrogen atoms, which are present in the body's water and fat molecules. When a person is placed in a strong magnetic field, the randomly aligned hydrogen atoms' spins tilt towards the direction of the B-field. As a consequence of the collective behavior of quantum spins the macroscopic magnetization is then pointing in B direction instead of being zero without the strong external field [19, 42]. This alignment is called **longitudinal magnetization** and is the basis for the resonance imaging process.

The magnetic field's flux density used in MRI is typically about 1.5 Tesla (T) to 3 T, which is much stronger than that of the earth's magnetic field (about 0.00005 T). This strong magnetic field causes the hydrogen nuclei to absorb energy from the radiofrequency (RF) pulses sent by the MRI machine. The RF pulses are sent at a specific frequency, known as the Larmor frequency [112], which depends on the flux density of the magnetic field and the type of nucleus being imaged. For hydrogen nuclei, the Larmor frequency is around 63.86 MHz (at a B-field of 1.5 T). Depending on their frequency and duration, the radiofrequency pulses can cause the hydrogen nuclei to flip their magnetic moment by 90 degrees, which is known as **transverse magnetization** [91].

After the RF pulses, the magnetization begins to relax back to its equilibrium state, emitting a signal that is detected by the receiver coil and used to construct the image. The process of relaxation is described by the T1 (spin-lattice) and T2 (spin-spin) relaxation times, which depend on the properties of the tissue being imaged. The different T1 and T2 relaxation times of the various tissues in the body allow the MRI machine to create images with different contrasts, highlighting different structures in the body.

The T1 relaxation time is the time it takes for the longitudinal magnetization to return to 63% of its initial value and is longer for fats than for water. The T2 relaxation time is the time it takes for the transverse magnetization to decay to 37% of its initial value and is longer for water than for fats, for example.

To examine and measure relaxation times, so called spin echo sequences are used which repeatedly send and receive RF pulses to determine relaxation times. Spin echo is commonly associated with T2-weighted imaging due to its effectiveness in highlighting T2 relaxation differences, but also offers advantages when applied in T1-weighted imaging. Therefore both, T2 as well as T1 will be covered when outlining the use of conventional spin echo sequences below.

3.2 Conventional MRI Sequences

When a patient is positioned within the MRI scanner, the hydrogen nuclei, primarily found in water and fat molecules, are exposed to a strong and uniform magnetic field with magnetic flux density B_0. This causes the hydrogen protons' spin to align with the magnetic field in the z-component.

The alignment of these spins with the field is, however, not perfect. Due to their intrinsic angular momentum, the protons precess about the direction of the magnetic field at a specific frequency, the Larmor frequency. This frequency is given by the Larmor Equation:

$$\omega_0 = \gamma B_0 \tag{2}$$

where

- ω_0 represents the Larmor frequency,
- γ is a constant, the gyromagnetic ratio (gyromagnetic ratios for different isotopes are shown in Table 1),
- B_0 is the magnetic flux density.

To generate an MRI signal, a radiofrequency pulse with the Larmor frequency is applied, tipping the net magnetization out of alignment with the magnetic field. Once the RF pulse is turned off, the system begins to relax back to its equilibrium state.

Isotope	Gyromagnetic Ratio (MHz/T)
^{1}H (Hydrogen-1)	42.576
^{2}H (Deuterium)	6.536
^{3}He (Helium-3)	-32.434
^{7}Li (Lithium-7)	16.546
^{13}C (Carbon-13)	10.705
^{15}N (Nitrogen-15)	-4.315
^{17}O (Oxygen-17)	-5.772
^{19}F (Fluorine-19)	40.052
^{23}Na (Sodium-23)	11.262
^{31}P (Phosphorus-31)	17.235

Table 1: Gyromagnetic ratios of commonly measured isotopes. The gyromagnetic ratio [MHz/T] is used in the Larmor equation to determine the resonance frequency of the nuclei in a magnetic field.

This relaxation process is characterized by two time constants: $T1$ and $T2$.

In order to spatially encode the MRI signal and create an image, gradient fields are applied. These gradient fields vary the magnetic field linearly across the imaging volume. The Larmor frequency at a particular location $\mathbf{r}$ within the body is then modified according to the equation:

$$\omega(\mathbf{r}) = \gamma(B_0 + \mathbf{G} \cdot \mathbf{r}) \tag{3}$$

where

- $\mathbf{G}$ is the gradient field vector,
- $\mathbf{r}$ is the position vector.

The received MRI signals are complex wave forms that are then transformed from the time domain to the frequency domain using the Fourier Transform. This transformation is crucial for reconstructing the final image, and it can be expressed with the equation:

$$I(\mathbf{k}) = \int \rho(\mathbf{r}) e^{-2\pi i \mathbf{k} \cdot \mathbf{r}} d\mathbf{r} \tag{4}$$

with terms

- $I(\mathbf{k})$ representing the image in k-space (frequency domain),
- $\rho(\mathbf{r})$ denoting the proton density at position $\mathbf{r}$,
- $\mathbf{k}$ being the spatial frequency.

By applying an inverse Fourier Transform to the data in k-space, the final image in the spatial domain can be obtained, revealing the internal structures of the body. T1-

and T2-weighted images are used to visualize anatomical details, such as the brain's gray and white matter or changes in tissue structure caused by disease, injury or, as needed here, tumors.

3.2.1 T1-Weighted Imaging

The $T1$ time constant, or longitudinal relaxation time, represents the time it takes for the longitudinal magnetization to recover to approximately 63% of its original magnitude. Assuming that an RF pulse has tipped the net magnetization vector away from its alignment with the magnetic field, the recovery of the longitudinal magnetization M_z back to its equilibrium value M_0 can be described by the model equation:

$$M_z(t) = M_0 - (M_0 - M_z(0))e^{-\frac{t}{T1}} \tag{5}$$

where

- $M_z(t)$ is the longitudinal magnetization at time t,
- M_0 is the equilibrium magnetization,
- $M_z(0)$ is the longitudinal magnetization just after the RF pulse
- $T1$ is the longitudinal relaxation time constant.
- t is the time,

Just after the RF Pulse (at $t = 0$) the longitudinal magnetization $M_z(0)$ would be significantly less than M_0, depending on how much the magnetization was tipped and could even be negative if the magnetization was flipped by 180°. Long time after the RF pulse (with $t \to \infty$) $M_z(t)$ approaches M_0, meaning the magnetization has fully recovered to its equilibrium state. The rate at which $M_z(t)$ recovers is determined by $T1$. Smaller $T1$ values mean faster recovery to equilibrium.

The spin echo sequence used in T1-weighted MRI imaging is performed in a repeated application of RF pulses as described below. The key difference to T2 spin echo sequence lies in the selection of how the sequence is timed, which is determined by repetition and echo time:

1. **Initial magnetization and RF pulse:** The body is subjected to a strong magnetic field and a 90-degree RF pulse is applied to disturb the alignment of the hydrogen protons.
2. **T1 and T2 relaxation processes:** After this RF pulse, both T1 and T2 relaxation processes begin. For T1-weighted imaging, the focus is on how quickly protons realign with the main magnetic field.

3. **T1 relaxation (spin-lattice relaxation):** T1 relaxation is the process by which the protons return to their original state, aligned with the main magnetic field. This process is called *spin-lattice relaxation* because it involves transferring energy from the protons to the surrounding 'lattice' or tissue environment.

4. **180-degree RF pulse:** A 180-degree pulse is applied. This pulse creates an echo by reversing the phase dispersion of the spinning protons.

5. **Timing for T1 weighting:** The key to achieving T1 weighting in a spin echo sequence lies in the timing of the pulses and the intervals between them. To emphasize T1 contrast, the repetition time (TR) (time between 90-degree pulses) is kept short. A short TR does not allow tissues with long T1 relaxation times to fully recover their longitudinal magnetization between pulses, resulting in lower signal intensity from these tissues.

6. **Echo time (TE):** The echo time (TE), which is the time between the application of the 90-degree pulse and the peak of the echo signal, is usually kept relatively short in T1-weighted imaging. This minimizes the T2 decay effect, thereby maintaining the focus on T1 contrast.

7. **Image contrast:** In T1-weighted spin echo images, tissues with short T1 relaxation times (like fat) appear brighter, while those with longer T1 times (like water) appear darker.

By adjusting the TR and TE appropriately, a spin echo sequence can be used to produce images that emphasize the T1 relaxation properties of different tissues. This flexibility is part of what makes spin echo sequences widely used in MRI. $T1$ weighting in MRI is achieved by choosing specific repetition times (TR) and echo times (TE) in the imaging sequence (as is shown in Figure 2) with short TR and short TE enhancing $T1$ contrast:

- **Repetition time (TR):** This is the time interval between successive RF pulse applications. A shorter TR allows for less complete recovery of longitudinal magnetization, emphasizing the differences in T1 relaxation times between tissues.

- **Echo time (TE):** This is the time interval between the application of the RF pulse and the acquisition of the MRI signal (echo). A shorter TE minimizes the decay of transverse magnetization and the influence of T2-weighting effects, resulting in images mainly influenced by T1 relaxation properties.

In T1-weighted imaging, a short repetition time and short echo time are used to create images with high signal-to-noise ratio (SNR) and high spatial resolution. T1-weighted images typically have bright signal intensity for fat, and intermediate signal intensity for gray matter, while white matter has a relatively low signal intensity. Fluids, such as cerebrospinal fluid (CSF), are dark on T1-weighted images. The imaging characteristics of different brain tumor regions in T1-weighted images will be discussed in section 3.4.

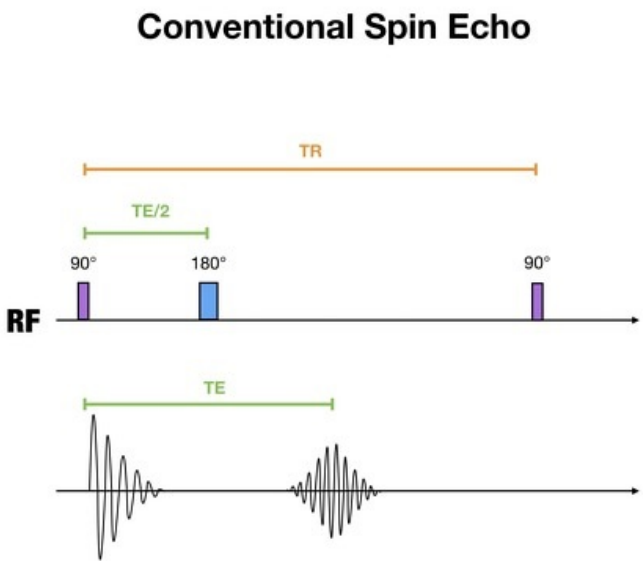

Figure 2: The two variables of interest in spin echo sequences are the repetition time (TR) and the echo time (TE). Spin echo sequences include a 90 degree pulse followed by one (or more) 180 degree refocusing pulses as shown in the diagrams. Illustration from Radiopedia [76].

3.2.2 T2-Weighted Imaging

On the other hand, $T2$ or transverse relaxation time is the time it takes for the transverse magnetization to decay to about 37% of its initial value. This decay is described by the equation:

$$M_{xy}(t) = M_{xy}(0)e^{-\frac{t}{T2}} \tag{6}$$

where

- $M_{xy}(t)$ represents the transverse magnetization at time t,
- $T2$ is the transverse relaxation time constant.

T2-weighted MRI sequences are designed to emphasize differences in T2 relaxation times between tissues. In a T2-weighted sequence, the contrast between tissues is primarily determined by their respective T2 relaxation times. Tissues with longer T2 relaxation times will appear brighter on T2-weighted images, while those with shorter T2 relaxation times will appear darker.

Similar to T1-weighted imaging, in T2-weighted spin echo sequences, the same parameters echo time (TE) and repetition time (TR) are used to be adjusted. The main steps of spin echo sequence with emphasis on T2 behaviour are listed below and illustrated in Figure 2:

1. **Initial magnetization:** In MRI, the body is placed in a strong magnetic field, causing the protons (mainly in water molecules) in the body to align with the field.

2. **90-degree RF pulse:** A radiofrequency (RF) pulse is applied at a certain angle, commonly 90 degrees. This pulse disturbs the alignment of the protons, causing them to tip away from the direction of the main magnetic field.

3. **Precession and phase coherence:** After the RF pulse is turned off, the protons start to precess (spin) and initially have phase coherence, meaning they spin in synchrony with each other.

4. **De-phasing:** Over time the protons begin to lose this coherence and start to precess at slightly different rates, leading to de-phasing. This de-phasing causes a decrease in the signal that can be detected. The de-phasing of spins is primarily caused by **spin-spin interactions**, which are generally interactions between the magnetic moments of different nuclei. In particular, this is caused by **dipole-dipole interactions** of hydrogen nuclei, which is a specific type of spin-spin interactions. These interactions lead to de-phasing of the spins because each nucleus experiences a slightly different local magnetic field due to the presence of neighboring nuclei. Another reason for de-phasing are inhomogeneities in the main magnetic field. They affect the precession of neighboring protons, leading to a loss of phase coherence.

5. **180-degree RF pulse:** Before the signal decreases too much, a second RF pulse, typically 180 degrees, is applied. This pulse flips the direction of the spinning protons.

6. **Re-phasing:** After the 180-degree pulse, the protons that were de-phasing now start to re-phase. This means they gradually come back into sync with each other.

7. **Echo signal:** As the protons re-phase, they emit the spin echo signal. This signal reaches its peak when the protons are most in phase, and this peak is detected and used to create the image.

8. **T2 relaxation:** The time it takes for the protons to lose phase coherence after the 180-degree pulse is associated with the T2 relaxation time, a property that varies between different tissues and is used to generate contrast in the images.

This sequence is particularly sensitive to differences in the T2 relaxation times of tissues, making it useful for detecting pathological changes in various organs and structures within the body. T2-weighted sequences are useful for visualizing pathologies that result in increased water content, such as edema, inflammation, and tumors. These tissues tend to have longer T2 relaxation times and, therefore, appear hyperintense on T2-weighted images.

3.2.3 T1-Weighted Contrast-Enhanced T1ce

T1ce MRI sequences are T1-weighted sequences performed after the administration of an intravenous contrast agent, typically a gadolinium-based compound. Gadolinium is a paramagnetic substance that shortens the T1 relaxation time of tissues it flows through. Consequently, regions that take up the contrast medium will have a shortened T1 relaxation time and will appear hyperintense on the T1-weighted images. This increased contrast between normal and abnormal tissues can help to identify and characterize various pathologies.

Some common clinical applications of T1ce MRI sequences are:

- **Inflammation and infection:** T1ce sequences can help detect and evaluate areas of inflammation or infection, as these regions often exhibit increased vascular permeability and contrast uptake. Examples include abscesses, osteomyelitis, and inflammatory conditions such as multiple sclerosis.
- **Vascular lesions:** T1ce sequences can be used to examine vascular abnormalities, such as arterial malformations, aneurysms, or areas of ischemia, as these lesions often show contrast enhancement due to disrupted blood-brain barrier or increased blood flow.
- **Post-surgical evaluation:** T1ce sequences can help evaluate the extent of surgical resection, detect residual tumor tissue, or identify postoperative complications such as hemorrhage or infection.
- **Tumor detection and characterization:** T1ce sequences are very valuable for visualizing and examining the extent of tumors in the brain, spine, and other organs. The contrast enhancement can help to delineate the tumor borders, differentiate between solid and necrotic components, and identify areas of tumor infiltration or invasion into neighbour tissues.

3.2.4 Fluid Attenuated Inversion Recovery FLAIR

Fluid attenuated inversion recovery (FLAIR) is a refined magnetic resonance imaging sequence that is widely used in clinical practice for its ability to suppress the signal from cerebrospinal fluid (CSF) and enhance the visualization of certain structures, particularly in the brain. The FLAIR images can provide valuable information about various neurological disorders, white matter diseases, and other brain abnormalities.

To understand FLAIR MRI sequences, it is important to recall the principles of T2-weighted MRI sequences, which emphasize differences in T2 relaxation times between tissues. In a T2-weighted sequence, tissues with longer T2 relaxation times appear brighter, while those with shorter T2 relaxation times appear darker. However, T2-weighted sequences can sometimes be negatively affected by the bright signal from CSF, which may overlay or hide the appearance of certain other tissues or

structures.

FLAIR sequences are thus designed to overcome this limitation by selectively suppressing the signal from CSF. This is achieved by incorporating an inversion recovery (IR) preparation pulse before the typical T2-weighted sequence.

In addition to suppressing the CSF signal, FLAIR sequences keep the T2-weighted contrast between different tissues, allowing to show improved visualization of various problems. Some clinical applications of FLAIR sequences are [90, 15]:

- **Detection of white matter lesions:** FLAIR sequences are highly sensitive for identifying white matter lesions, such as those associated with multiple sclerosis, small vessel ischemic disease, or vasculitis. These lesions often appear as hyperintense areas on FLAIR images due to their longer T2 relaxation times compared to normal white matter.
- **Assessment of intracranial hemorrhage:** FLAIR sequences can help differentiate between acute, subacute, and chronic intracranial hemorrhages by suppressing the CSF signal and highlighting the different T2 relaxation times of blood products at various stages of clot evolution.
- **Identification of meningitis and encephalitis:** FLAIR sequences can reveal meningeal enhancement and cortical abnormalities associated with infectious or inflammatory processes, such as bacterial or viral meningitis and encephalitis.
- **Evaluation of brain tumors:** FLAIR sequences can provide valuable information about the extent and infiltration of brain tumors, particularly when the tumor is located near the brain's surface, where the bright CSF signal on conventional T2-weighted images might cover the tumor's boundaries.

The key to suppressing the CSF signal in FLAIR sequences lies in the selection of an appropriate inversion time (TI). TI is the time interval between the application of the 180° inversion recovery preparation pulse and the subsequent excitation pulse, as is shown in Figure 3. The inversion time is chosen to coincide with the null point of the CSF signal, which is the time at which the longitudinal magnetization of CSF is zero. At this time point, the excitation pulse has minimal effect on the CSF signal, as there is no net magnetization to tip into the transverse plane. Consequently, the CSF signal is effectively suppressed in the resulting image.

Meanwhile, other tissues with different relaxation times will have non-zero longitudinal magnetization at the time of the excitation pulse, producing a detectable signal on the final image. This allows FLAIR sequences to maintain the T2-weighted contrast between different tissues while selectively suppressing the CSF signal, resulting in improved visualization of various brain parts and tissues.

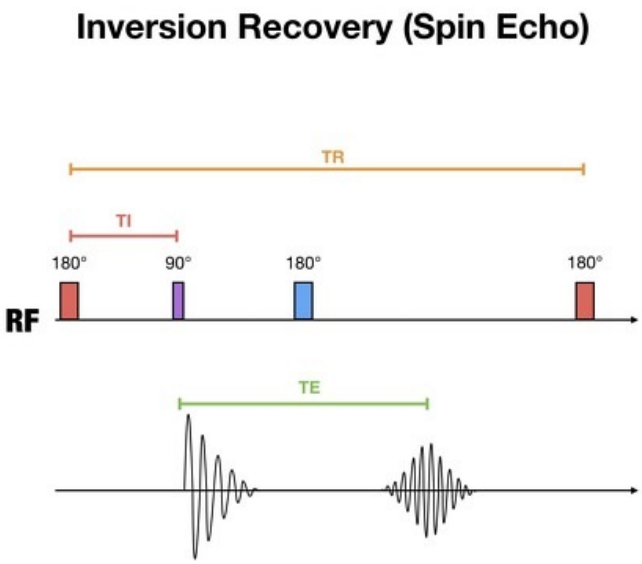

Figure 3: The spin echo 90° readout pulse is applied at the exact time when longitudinal magnetization reaches the null point for the tissue which is to suppress. The time elapsed between the preparatory 180° IR pulse and the 90° excitation RF pulse is called time to inversion (TI). Illustration from Radiopedia [76].

3.3 Advanced MRI Sequences

3.3.1 Diffusion Weighted Imaging DWI

Diffusion weighted imaging (DWI) is different type of magnetic resonance imaging technique, which uses the diffusion of water molecules to generate contrast in the images. This technique allows the imaging of cellular structures and provides a unique insight into the microscopic characteristics of tissues [9, 42]. The DWI imaging is based on the random, microscopic motion of water molecules, which is known as Brownian motion. This motion is influenced by the cellular environment, mainly the cell density and the integrity of cell membranes. When the diffusion is unrestricted (as for example in cerebrospinal fluid) it is called **isotropic**. But within certain structures the diffusion can be more restricted in some directions than in others, then it is called **anisotropic**.

In clinical practice, DWI is particularly useful for detecting early strokes, as restricted diffusion is one of the earliest signs of cerebral ischemia. However, it is also useful for analysing infection or inflammation or in the characterization of tumors [9].

The physics behind DWI is based on the concept of diffusion weighting. In DWI, the strength of the diffusion weighting is expressed as a so called *b-value*. The *b-value* determines the sensitivity of the DWI sequence to diffusion. Higher *b-values* lead to greater diffusion weighting, but they also lead to a lower signal-to-noise ratio. The most common approach used in DWI is the Stejskal-Tanner pulse sequence, which is

characterized by a pair of diffusion-sensitizing magnetic field gradients [109]. These gradients are typically added to a spin-echo or echo-planar imaging sequence. The pulse sequence uses a specific row of gradient pulses to encode the diffusion of water into the phase of the MRI signal (see Figure 4). The equation that describes the signal attenuation $S(b)$ due to diffusion in the presence of a gradient pulse is given by the Stejskal-Tanner equation:

$$S(b) = S_0 e^{-bD} \tag{7}$$

where

- $S(b)$ is the signal intensity with diffusion weighting.
- S_0 is the signal intensity without diffusion weighting (i.e., when $b = 0$),
- D is the diffusion coefficient of the tissue, which here is a measure of the diffusion rate of water molecules,
- b is the diffusion weighting factor or *b-value*, which is given as:

$$b = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}) \tag{8}$$

In the above equation for the *b-value*

- γ is the gyromagnetic ratio (see Table 1),
- G is the amplitude of the gradient pulses,
- δ is the duration of each gradient pulse,
- Δ is the time between the application of the gradient pulses.

The Stejskal-Tanner equation (7) is used to determine the diffusion coefficient by measuring the signal attenuation at different *b-values*. It is assumed that the diffusion process is Gaussian, which is true for free and unrestricted diffusion. But in biological tissues, the diffusion may not be free caused by barriers such as cell membranes, and thus, the received signal attenuation curve may deviate from the simple exponential behavior predicted by the Stejskal-Tanner equation. This helps in quantifying the diffusion of molecules (usually water) within tissue. By applying magnetic field gradients during an MRI scan, it is possible to detect variations in the diffusion of water molecules, which can indicate various abnormal conditions. For example, in areas where diffusion is restricted (like in some types of stroke or tumor), the actual signal intensity will be higher compared to the one predicted by the Stejskal-Tanner equation (free diffusion).

An even more enhanced technique is available in Diffusion Tensor Imaging (DTI), where diffusion effects are analysed towards their direction using the diffusion tensor.

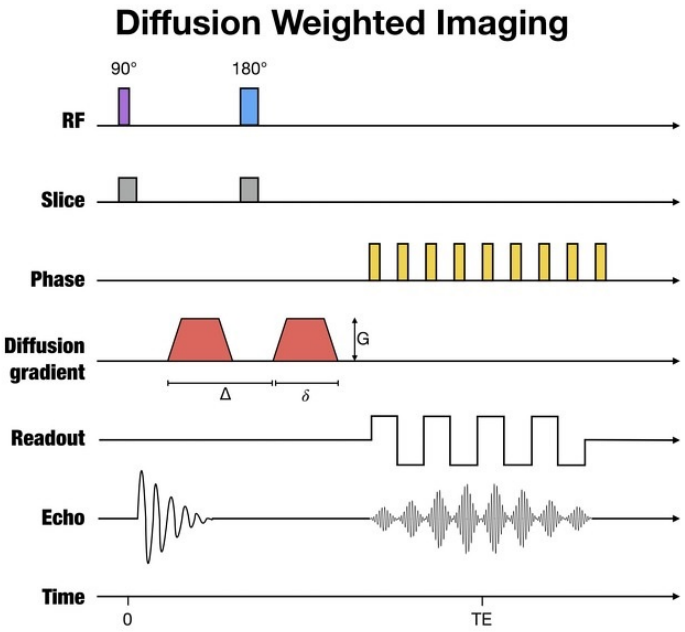

Figure 4: Schematic representation of a Diffusion Weighted Imaging sequence. Diffusion gradients are applied with amplitude G, pulse duration δ and repetition time Δ. Illustration from Radiopedia [76].

3.3.2 Diffusion Tensor Imaging DTI

Diffusion Tensor Imaging (DTI) is an even specialized form of Diffusion Weighted Imaging that allows the visualization and quantification of water diffusion directions within tissues. Similar to DWI it is also based on the water molecules' Brownian motion. In many tissues, this motion is isotropic. However, in structured tissues this motion is restricted by cell membranes etc. making it anisotropic (directionally dependent). This is the basic principle of DWI imaging as discussed above. But where DWI only captures the degree of anisotropic diffusion, DTI also measures the diffusion gradient **direction**.

Similar to the Stejskal-Tanner equation (7), the signal attenuation with a gradient direction $\mathbf{g}$ is:

$$S(\mathbf{g}, b) = S_0 e^{-b\mathbf{g}^T \mathbf{D}\mathbf{g}} \tag{9}$$

Where $\mathbf{D}$ is the diffusion tensor and $\mathbf{g}^T$ is the transpose of the gradient direction vector. The diffusion tensor $\mathbf{D}$ is a mathematical construct which is used here to represent and calculate the diffusion of water molecules in all three spatial dimensions. This provides a more comprehensive description of diffusion than a single diffusion coefficient. $\mathbf{D}$ can be given as a 3x3 symmetric matrix with the diagonal elements

representing diffusion along the axes of the coordinate system, and the off-diagonal elements representing diffusion occurring off-axis. The calculation involves several steps:

1. **Acquisition of Diffusion-Weighted Images (DWIs):** Multiple images are acquired using different b-values and using the Stejskal-Tanner pulse sequence with gradient pulses applied in at least six non-collinear directions. At least two b-values are needed, including one very low or zero b-value (b_0), which doesn't sense the image to diffusion, and at least one higher b-value (b_i).

2. **Signal intensity measurements:** For each b-value, the MRI measures the signal intensity of the tissue. The diffusion gradient causes signal loss in the tissue proportional to the amount of diffusion. The signal attenuation in each voxel for each gradient direction can then be expressed using the generalized Stejskal-Tanner equation (9) which can be rearranged as shown in the next step.

3. **Linear equation system:** The natural logarithm of the signal attenuation ratio between a diffusion-weighted signal S_i (using b_i) and a non-diffusion-weighted signal S_0 (using b_0) is linearly related to the diffusion tensor and the b-value b_i, as shown in equation (10).

4. **Solving for the tensor:** Measurements from multiple directions result in a system of equations that can be solved for the six unique elements of the diffusion tensor (three diagonal elements and three off-diagonal elements due to its symmetry).

$$\ln\left(\frac{S_0}{S_i}\right) = b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i \tag{10}$$

Here, S_0 is the signal intensity in the absence of diffusion weighting (using low b_0 value or zero b-value), S_i is the signal intensity with diffusion weighting in the i-th direction, b_i is the corresponding b-value, and $\mathbf{g}_i$ is the gradient direction vector for the i-th measurement.

The diffusion tensor $\mathbf{D}$, is represented as:

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} \tag{11}$$

Where the diagonal elements D_{xx}, D_{yy}, and D_{zz} are principal diffusivities along the x, y, and z axes. The off-diagonal elements represent correlated diffusivities between

the axes.

From the eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$, quantitative measures like the Fractional Anisotropy (FA) and the Mean Diffusivity (MD) can be calculated. Fractional Anisotropy (FA) is a scalar value between 0 and 1 that describes the degree of anisotropy of a diffusion process. A value of 0 means that diffusion is isotropic, whereas a value of 1 indicates that diffusion occurs only along one axis and is restricted along the other axes. MD gives the average diffusion across all directions:

1. The Mean Diffusivity (MD):

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \tag{12}$$

2. The Fractional Anisotropy (FA):

$$FA = \sqrt{\frac{3}{2}} \cdot \frac{\sqrt{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{13}$$

Where λ_1, λ_2, and λ_3 are the eigenvalues of the tensor $\mathbf{D}$. The eigenvalues give the magnitude of diffusion in the directions defined by their corresponding eigenvectors. The principal eigenvector provides the direction of maximum diffusion, which is often aligned with the orientation of fibrous tissue.

The diffusion tensor provides much richer information about tissue structure than just the diffusion coefficient. Diffusion Tensor Imaging data is often visualized as color maps as shown in Figure 5, where the color encodes the primary direction of diffusion, providing a visual assessment of the tissue structure.

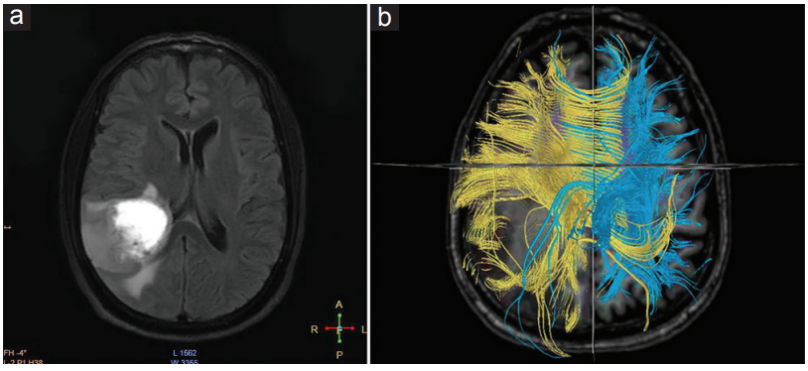

Figure 5: Diffusion Tensor Image example from "Role of Diffusion Tensor Imaging in Brain Tumor Surgery" by Dubey et al. [32]. The colored lines show the primary directions of diffusion. Displacement of the direction tracts is observable in the tumor region.

3.4 Imaging Characteristics of Brain Tissues and Brain Tumor Regions

Magnetic Resonance Imaging is a key diagnostic tool for identifying brain tumors and brain tumor regions, providing detailed images of the brain's anatomy and pathology. The different conventional MRI sequences presented above (T1, T1ce, T2, FLAIR) highlight different aspects of the tissue and different parts of a tumor. This section will describe how these regions typically appear in each of the mentioned sequences.

The interpretation of these imaging characteristics is very nuanced and can be influenced by numerous factors such as the specific type of brain tumor or the settings of the MRI scan itself. Radiologists and clinicians combine information from the different sequences to form an image of the brain tumor regions. The same is expected to be achieved automatically by the BraTS challenge's segmentation algorithms presented in the Sections 5.2.

Cerebrospinal Fluid CSF

Cerebrospinal fluid (CSF) is a clear, colorless fluid that surrounds the brain and spinal cord, providing mechanical and immunological protection. In magnetic resonance imaging, the signal from CSF can be particularly prominent due to its unique physical properties, including its high water content and long T1 and T2 relaxation times.

In MRI, the contrast between tissues is primarily determined by the differences in their relaxation times (T1, T2) and proton density. CSF is mainly composed of water, which is rich in hydrogen atom cores, thus resulting in a high proton density. The large number of protons in CSF leads to a strong signal in MRI, as the signal intensity increases with proton density.

In T1-weighted MRI sequences, tissues with shorter T1 relaxation times appear bright (hyperintense), while those with longer T1 relaxation times appear darker (hypointense). Conversely, in T2-weighted sequences, tissues with longer T2 relaxation times appear bright, while those with shorter T2 relaxation times appear dark. Due to the long T1 and T2 relaxation times of CSF, it appears hypointense on T1-weighted images and hyperintense on T2-weighted images, see Table 2. Since CSF is not affected by any contrast enhancing agent it will also appear hypointense on T1ce modalities. Due to the active suppression in Flair records, it will also appear hypointense there.

White Matter WM

White matter appears hyperintense on T1 compared to gray matter. The higher signal intensity of white matter in T1 images is due to the shorter T1 relaxation time

of a substance named myelin, which is abundant in white matter. Myelin, being rich in fats, has a shorter T1 relaxation time compared to the surrounding brain tissues, making it appear brighter.

On T2 weighted sequences white matter appears hypointense compared to grey matter. In T2 sequences, the longer T2 relaxation time of water-rich tissues makes them appear brighter. Since white matter contains less water compared to grey matter, it appears darker. The T2 relaxation time is influenced by the water content and the structural integrity of tissues. White matter, with its dense myelin sheaths, has a longer T2 relaxation time, leading to lower signal intensity.

For white matter, the appearance on T1ce depends mostly on the presence of pathological changes. In normal conditions, white matter does not significantly accumulate gadolinium contrast, and therefore its appearance is similar to that in standard T1-weighted images – hyperintense compared to grey matter. However, in pathological conditions, such as in areas of inflammation, infection, or tumors, white matter may accumulate the contrast agent. This leads, in such cases, to increased signal intensity, making these areas appear brighter than the surrounding tissues.

For FLAIR images white matter appears darker compared to gray matter, similar to T2 images. FLAIR is a special inversion recovery sequence designed to clean the signal from fluids like CSF. Brain tissue on FLAIR images appears similar to T2 images, with grey matter being brighter than white matter. The long inversion time used in FLAIR sequences suppresses the signal from free water, making CSF appear very dark. Since white matter has less free water content compared to gray matter, it appears darker than gray matter in FLAIR images as well, but still brighter than CSF.

Gray Matter GM

Gray matter has an intermediate signal intensity on T1 images. The intermediate signal intensity has it's reason in the balance of water and fat content in gray matter. The T1 relaxation time, which is influenced by the molecular environment within the tissue, results in an intermediate appearance.

Similarly to T1 images, gray matter also appears with intermediate signal intensity on T2 images. In T2 imaging, tissues with longer T2 relaxation times, such as fluids, appear brighter. Gray matter, having an intermediate T2 relaxation time due to its composition, appears as intermediate in intensity.

In normal conditions, gray matter does not significantly accumulate gadolinium-based contrast agents, so its appearance in contrast enhanced T1ce remains similar or slightly hyperintense to non-contrast T1 images. However, in the presence of pathological conditions, areas of gray matter that accumulate the contrast agent will appear brighter.

On FLAIR images, gray matter appears brighter than white matter, since FLAIR is a variation of T2 imaging with suppression of the signal from fluids like CSF. The contrast between gray and white matter is similar to that in T2 images, with gray matter appearing isointense or slightly brighter due to its higher water content compared to white matter.

Component	**T1**	**T2**	**T1ce**	**FLAIR**
Cerebrospinal Fluid CSF	Hypointense	Hyperintense	Hypointense	Hypointense
White Matter WM	Hyperintense	Hypointense	Hyperintense	Hypointense
Gray Matter GM	Isointense	Isointense	Hyperintense	Isointense
Edema ED	Variable	Iso/Hyperintense	Hypointense	Hyperintense
Necrosis NCR	Hypointense	Hyperintense	Hypointense	Hyperintense
Enhancing Tumor ET	Iso/Hypointense	Iso/Hyperintense	Hyperintense	Iso/Hyperintense
Non-Enhancing NET	Iso/Hyperintense	Iso/Hyperintense	Variable	Iso/Hyperintense

Table 2: Intensities of various brain and tumor components on MRI modalities. Variable intensity means the signal can vary depending on specific conditions.

Edema ED

Edema generally appears hypointense on T1 due to it's increased water content. Edema tissue has a longer T1 relaxation time than normal brain tissue, resulting in lower signal intensity on T1-weighted images [81]. However, the appearance of edema on T1-weighted images can vary depending on the underlying cause and the stage of the edema. For example, in some cases, such as with certain types of inflammation or hemorrhage, the edematous area might appear slightly brighter or isointense (having similar signal intensity as surrounding tissues) compared to normal tissue. On contrast enhanced T1ce, edema usually appears hypointense since contrast mainly enhances solid tumor components.

On T2 Edema appears isointense or hyperintense because of the increased water content, which shortens the T2 relaxation time, leading to higher signal intensity on T2-weighted images. FLAIR imaging, which suppresses the signal from cerebrospinal fluid, makes the edematous tissue appear even more obvious [45]. It suppresses the signal from cerebrospinal fluid and makes the edema tissue be highlighted. The signal intensities for edema regions are listed in Table 2.

Necrosis NCR

Necrotic tissue appears hypointense on T1 due to the lack of viable tissue and increased water content [22]. On T2-weighted images, it can appear bright due to the presence of fluid in the necrotic tissue, while the surrounding tumor may be isointense or less hyperintense.

On contrast-enhanced T1ce, the necrotic parts of a brain tumor typically appear as a central area of hypointense tissue, surrounded by a rim of enhancement after the administration of contrast material. This is due to the breakdown of the blood-brain barrier in the viable tumor tissue surrounding the necrosis. The necrotic core generally appears hypointense, but there can be an area of enhancement surrounding the necrosis, reflecting viable tumor tissue or also, in some cases, treatment-related changes [94]. Advanced imaging techniques, such as the discussed Diffusion Weighted Imaging, can provide additional information to help distinguish tumor necrosis from other entities like treatment-related changes.

In FLAIR, similar to T2, the necrotic core may appear hyperintense reflecting the increased water content. The surrounding rim of viable tumor tissue may appear slightly hyperintense or isointense [25].

Enhancing Tumor ET

Before contrast administration, an enhancing tumor can be isointense or hypointense on T1, making it difficult to distinguish from surrounding tissue. Also in T2, the intensity is not clearly shown, the enhancing tumor regions may appear hyperintense or isointense.

After administration of a contrast agent, the enhancing tumor appears very hyperintense due to the breakdown of the blood-brain barrier, allowing the contrasting medium to accumulate in the tumor tissue [62]. This means, in contrast-enhanced T1ce, the enhancing tumor parts ET are highlighted, which is crucial for identifying active tumor tissue.

In FLAIR, again the enhancing tumor regions only appear hyperintense or isointense and the brightness is not as pronounced as on post-contrast T1-weighted images.

Non-Enhancing Tumor NET

The detectability of non-enhancing tumor segments on MRI depends on various factors such as the MRI sequences used and the specific characteristics of the tumor as well as the surrounding brain tissue. Mainly the presence of surrounding edema can affect the visibility of non-enhancing tumor segments. While the non-enhancing tumor segments can be very challenging to detect, the use of multiple standard MRI sequences (T1, T2, T1ce, Flair) and the additional use of advanced imaging techniques (such as DWI and DTI) can improve their visibility and contribute to a more accurate analysis of the tumor.

The standard MRI sequences offer imaging characteristics pointing towards non-enhancing parts, but often only give subtle expressions of the same making it hard for radiologists to reliably detect the NET parts [76, 65, 34]. In T1-weighted images, non-enhancing tumor segments may appear isointense or hypointense and can

be difficult to distinguish from surrounding normal brain tissue. That makes pre-contrast T1-weighted images typically no good choice for detecting non-enhancing tumor segments since it may not be clearly visible unless there is significant mass effect or distortion of normal brain anatomy.

T2-weighted images can provide good contrast between tumor tissue and normal brain tissue, helping to delineate the extent of the overall tumor. Also on T2-weighted images non-enhancing tumor segments are often more visible, appearing slightly hyperintense compared to surrounding brain tissue, which is due to increased water content, cellularity, and changes in the tissue's micro structure, all of which affect the T2 relaxation time.

T1-weighted records with contrast enhancement (T1ce) can sometimes make non-enhancing regions more apparent by contrast, but only if there is a clear border between enhancing and non-enhancing regions [76]. The contrast results from non-enhancing regions absorbing less or no gadolinium-based contrast agents. For non-enhancing tumor parts lying apart from enhancing tissue, this effect is not helpful. This means T1ce can highlight enhancing tumor regions to some degree, depending on their localization.

FLAIR sequences are similarly useful for detecting non-enhancing tumor segments as T2, as they provide high contrast between tumor tissue and normal brain tissue. FLAIR images are sensitive to changes in water content and can highlight areas of edema or infiltrative tumor that might not be apparent on other sequences. As it additionally suppresses the signal from CSF, it allows the tumor tissue to stand out more clearly and gives together with T2 and within the standard MRI sequences, a good resource for identifying NET.

The non-enhancing tumor parts are generally located within regions of edema or surrounded by edema and are thus hard to distinguish from this domain. In their paper "Non-Contrast-Enhancing Tumor: A New Frontier in Glioblastoma Research" [65] Lasocki et al. describe possible criteria for differentiating between edema and non-enhancing tumor parts:

> *The white matter involvement of edema is typically relatively concentric around the enhancing lesions, other than where hindered by the relative barrier afforded by the gray matter. In contrast, eccentric extension of FLAIR hyperintensity, not accounted for by anatomic constraints, is highly suggestive of nCET [NET]. [...]*
> *Another useful differentiating feature is that edema is typically associated with more marked T2 and FLAIR hyperintensity than nCET [NET] often fading somewhat toward the periphery of the edema. In contrast, FLAIR hyperintensity related to nCET [NET] is usually more subtle. The relatively mild FLAIR hyperintensity of nCET [NET] is well-demonstrated in the gray matter where the distinction is simpler but can also be seen*

> *in the white matter, for example with an eccentric tongue of nCET [NET] extending beyond an area that has the typical appearance of edema. This finding has recently been supported by a study correlating radiographically localized biopsies with multiparametric MR imaging, which found that T2 FLAIR was inversely correlated with cell density. These results support mild FLAIR hyperintensity being a feature of nCET [NET], while greater hyperintensity indicates that edema dominates.* [65]

In summary, NET comes with slightly less intensity in the T2 and Flair modalities compared to the actual edema parts. Also non-enhancing parts are often present in an eccentric extension adjacent the enhancing tumor core. This phenomenon will be visible in the 3D modeled representations shown in Section 7.5.
Advanced MRI sequence images such as Diffusion Weighted Imaging (DWI) and Diffusion Tensor Imaging (DTI) can provide valuable additional information about the cellular density of the tumor, which can aid in detecting non-enhancing tumor segments. High cellular density can lead to restricted diffusion, which then appears hyperintense on DWI and DTI [20]. Lasocki et. al [65] state the following about the use of DWI and DTI for differentiating NET from edema, which is challenging in the conventional sequences:

> *The potential use of DWI for identifying nCET [NET] is based on the correlation between ADC [diffusion coefficient D] values and tumor cellularity. This association suggests that relative diffusion restriction (lower ADC values) is a marker of nCET [NET], in contrast to the facilitated diffusion occurring with edema. Published results have been conflicting, but reassuring results have been obtained with newer techniques. For example, Price et al. have suggested value in DTI, with infiltrating tumor suggested by the presence of an increase in the isotropic component of the diffusion tensor and a marginal increase in the anisotropic component. Stadlbauer et al. have suggested that fractional anisotropy correlates better with histopathologic parameters than mean diffusivity. Functional diffusion maps and high-b-value DWI have also been suggested as useful tools for identifying nCET [NET].* [65]

Other advanced imaging techniques such as MR Spectroscopy and Perfusion-Weighted Imaging (PWI) can provide additional information that may aid in the detection and characterization of non-enhancing tumor segments [82, 35].

4 Machine Learning

This section will provide the machine learning concepts needed for the presented application of non-enhancing tumor extraction in the BraTS datasets of the recent years. Therefore the general concept of *semantic image segmentation* will be introduced in Section 4.1. Section 4.2 will show different types of *convolutional neural nets for image segmentation* and will detail the U-Net architecture, which will make up the basis for further elaborations in the later chapters. Section 4.3 will describe the common metrics in image segmentation applications and Section 4.4 will give the needed loss functions respectively. An essential image analysis and enhancement tool in this work will be *morphological filtering*, which will be used to process binary brain tumor region masks and hence will be introduced in Section 4.5.

4.1 Semantic Image Segmentation

Image segmentation using neural networks is a powerful technique in computer vision that involves dividing an image into segments to isolate specific objects or regions. The goal is to understand and interpret the image's content at a more detailed level by identifying objects and their boundaries, making it a key component in many computer vision applications such as autonomous driving, robotics, medical imaging, and scene understanding [71, 86].

Image segmentation applications occur in two forms, **semantic image segmentation** and **instance segmentation**. While both, semantic image segmentation and instance segmentation aim to classify pixels in an image, they differ in what they classify the pixels as.

Semantic image segmentation focuses on assigning a class label to each pixel in an image. However, it does not differentiate between different instances of the same class. For example, if there are two cars in an image, semantic segmentation would label all the pixels belonging to both cars as "car" without distinguishing between the two individual cars.

Instance segmentation, on the other hand, not only classifies each pixel in an image but also differentiates between different instances of the same class. Using the previous example, instance segmentation would label the pixels belonging to one car as "car1" and the pixels belonging to the other car as "car2."

As already stated in the previous chapters, one goal of medical screening and diagnostic assistance applications is the automatic semantic segmentation of a variety of different regions in medical imaging data. As described by a comprehensive survey on the medical use of semantic image segmentation by Risheng Wang [119] it plays a vital role in medical applications, enabling the automatic analysis and interpretation of medical images. It has the potential to significantly improve diagnosis, treatment planning, and monitoring of various diseases and medical conditions. Some key

medical applications of semantic image segmentation are:

- **Organ and tissue segmentation**: Semantic segmentation can be used to identify and delineate organs, tissues, and other anatomical structures in medical images, such as CT, MRI, or ultrasound scans. This helps physicians to assess organ size, shape and position, which is crucial for diagnosis, surgical planning, and treatment monitoring [99].
- **Lesion segmentation**: In various medical conditions, such as multiple sclerosis, diabetic retinopathy, and skin diseases, detecting and quantifying lesions is crucial for diagnosis and monitoring. Semantic segmentation can assist in identifying and measuring lesions, providing valuable information about their size, shape, and distribution [60].
- **Vascular segmentation**: Analyzing blood vessels in medical images is important for diagnosing and treating vascular diseases, such as atherosclerosis or aneurysms. Semantic segmentation can help identify and measure blood vessel structures, aiding in the assessment of vascular abnormalities and planning of interventions [121].
- **Cardiac segmentation**: In cardiac imaging, semantic segmentation can help delineate the heart's chambers, valves, and blood vessels. This enables accurate assessment of cardiac function, diagnosis of heart diseases, and planning of interventions such as valve replacement or coronary artery bypass surgery [26].
- **Brain segmentation**: In neuroimaging, semantic segmentation can be used to differentiate various brain structures, such as the cortex, white matter, and subcortical regions. This information is valuable for studying brain development, understanding neurological disorders, and planning neurosurgical procedures [64].
- **Tumor segmentation**: Accurate identification and localization of tumors in medical images are essential for cancer diagnosis, staging, and treatment planning. Semantic segmentation can help delineate tumor boundaries and differentiate tumors from healthy tissues, enabling precise measurements and more informed clinical decisions. The current state in breast cancer detection and segmenation, for example, is shown in [84].

Accurate segmentation of anatomical structures and regions of interest is crucial for planning and performing minimally invasive procedures. Semantic segmentation can help create detailed 3D models, which can be used for surgical navigation and real-time guidance during image-guided surgery interventions. These are just a few examples of the many potential applications of semantic image segmentation in the medical domain. By automating the analysis of medical images, semantic segmentation can reduce the workload of clinicians, increase diagnostic accuracy, and improve patient care.

4.2 Convolutional Neural Networks for Image Segmentation

4.2.1 Fully Convolutional Networks

Fully Convolutional Networks (FCNs), historically mark an innovation in the field of semantic segmentation. Introduced by Long et al. in 2015 [71], FCNs radically deviated from the conventional architectural paradigm of Convolutional Neural Networks (CNNs), particularly by enabling spatially dense predictions without necessitating input of a fixed size or deploying fully connected layers. Before Long et al.'s work the concept and architecture evolved from a rich history and prior works in neural networks and image processing.

Introduced by Kunihiko Fukushima in 1980 [36], the Neocognitron, an early Convolutional Neural Network, set a milestone in neural networks capability of hierarchical, multi-layered pattern recognition. This model was influencing later models, including the CNNs foundational for FCNs. Later LeCun et al. developed LeNet-5 in 1998 [66], a significant development in CNNs that recognized handwritten characters. Although not directly linked to FCNs, LeNet-5 laid down the foundational architectures of convolutional networks.

Before FCNs, semantic segmentation often involved patch classification, wherein image patches were independently classified into categories, often disregarding spatial and contextual information among neighboring patches. Classical machine learning techniques like TextonForest [104] and Random Forests [16] were popular for such applications. Some methods implemented superpixel segmentation [67], where images were over-segmented into superpixels, followed by labeling via classifiers. These, however, lacked end-to-end training and dense prediction capabilities.

The leading work of Long et al. in 2015 [71] introduced FCNs, employing CNNs in an end-to-end trainable architecture, efficiently performing semantic segmentation by replacing fully connected layers with convolutional layers and utilizing upsampling to generate dense pixel-wise outputs.

Looking into the architectural details, a FCN is divided into an encoder, which is primarily responsible for downsampling and semantic understanding of the input image, and a decoder that concerns itself with upsampling and spatially dense prediction. The encoder employs a sequence of convolutional and pooling layers, successively reducing spatial dimensions while incrementally increasing the depth of feature channels. The principle scheme of this architecture is shown in Figure 6. Often in practice, well established pre-trained models like VGG16 or AlexNet are employed as the encoder, exploiting the powerful principle of transfer learning.

In contrast, the decoder is used to upscale the low-resolution feature maps derived from the encoder, generating a high-resolution output. This is achieved via the use of transpose convolutional layers, which effectively manage to upscale feature maps.

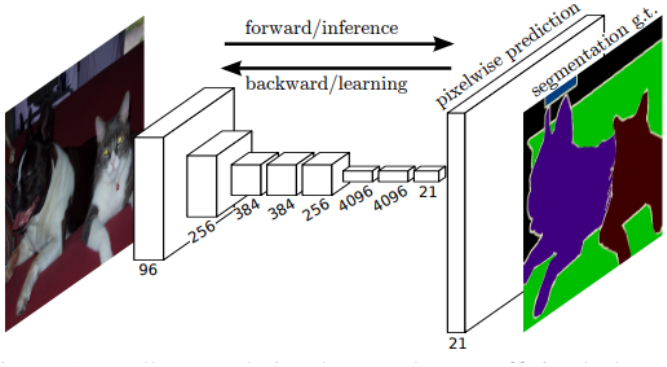

Figure 6: Encoder-Decoder architecture as proposed by Long et al. in his paper "Fully Convolutional Networks for Semantic Segmentation" [71].

Within this architecture, the traditional fully connected layers of conventional CNNs, typically employed for classification, are replaced with 1x1 convolutional layers in FCNs. This strategic modification enables the network to accept input of arbitrary size, thereby enhancing its versatility and application potential. The concept of transposed convolutions, or fractionally-strided convolutions, further enables this architecture in the precise upsampling of lower-resolution feature maps to higher resolutions.

FCNs are typically trained using pixel-wise cross-entropy loss, which compares the predicted pixel-wise class probabilities against the one-hot encoded ground truth, computing the loss for each pixel and averaging over the entire image or batch. Alternative loss functions, such as the Sørensen-Dice loss, are also employed, especially in situations where class imbalance might be prevalent, gauging the overlap between the predicted segmentation and the ground truth. Fully Convolutional Networks have been deployed for semantic segmentation, object detection, and instance segmentation in a broad spectrum of applications, with capability across different tasks in computer vision.

4.2.2 Mask R-CNN

Mask R-CNN is a deep learning model for instance segmentation, a task that combines object detection and semantic segmentation. It was proposed by Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick in a 2017 paper titled "Mask R-CNN" [48]. This model extends Faster R-CNN [96], a model for object detection, by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition.

The major elements in a Mask R-CNN architecture are a *Backbone Network* followed by a *Region Proposal Network (RPN)* using *RoIAlign* and incorporating *Segmentation Masks*. A Convolutional Neural Network (CNN) acts as the backbone of the

Mask R-CNN. The backbone is responsible for the initial extraction of features from the input image. The original Mask R-CNN paper [48] used a ResNet-101 model with a Feature Pyramid Network (FPN) extension as the backbone. The Region Proposal Network uses the output of the backbone to propose regions of the image that may contain an object (that means regions of interests or RoIs). These proposals are then used by the following layers to identify objects and their locations.

RoIAlign is a crucial part of the Mask R-CNN architecture that was not present in Faster R-CNN. While Faster R-CNN used RoIPool [96], which could cause misalignment between the RoIs and the extracted features, RoIAlign fixes this by using bilinear interpolation to maintain spatial precision [48]. Similar to Faster R-CNN, Mask R-CNN also includes a layer for object classification and bounding box regression. Given the RoIs from the RPN, this layer outputs class labels for each RoI along with refinements to the bounding box coordinates.

In parallel to the classification and bounding box regression layer, Mask R-CNN includes a branch for predicting a binary mask for each RoI. This mask indicates the pixels in the bounding box that belong to the object. This is the component that makes Mask R-CNN suitable for *instance segmentation* tasks. The key innovation in Mask R-CNN was the addition of the segmentation mask branch to the existing Faster R-CNN model, along with the use of RoIAlign to maintain spatial precision. A schematic representation of the proposed Mask R-CNN model structure from He et al.'s paper is shown in Figure 7.

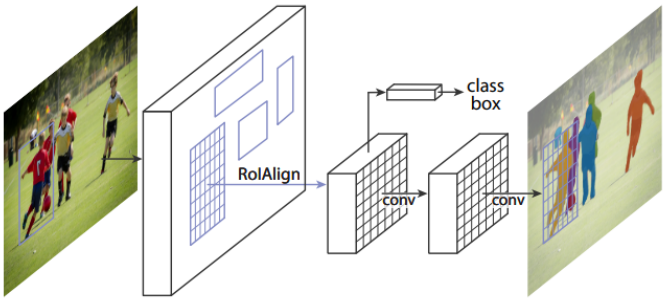

Figure 7: Mask R-CNN architecture, segmentation mask branch is added in parallel to Faster R-CNN's bounding box and classification path. Image from He et al. [48].

Besides medical image processing, Mask R-CNN instance segmentation finds it's application in other popular domains like defect detection in manufacturing, robotics, archeology, wildlife monitoring and more [30, 70, 14, 120].

4.2.3 U-Net Architecture

The U-Net architecture is a deep convolutional neural network that was first introduced in 2015 by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in a paper

titled ”U-Net: Convolutional Networks for Biomedical Image Segmentation”, which was published in the Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) in 2015 [97]. It has quickly become a popular choice for image segmentation tasks due to its effectiveness and adaptability. Despite it's original scope of use was biomedical image analysis, it was later increasingly used in different fields of segmentation applications.

The team around Olaf Ronneberger from the University of Freiburg was working on a medical image segmentation project and found that existing segmentation models were not performing well enough. This led to the development a new model specifically for medical image segmentation tasks in their original work from 2015 [97] and was extended to the use of 3D medical image data in their paper ”3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation” from 2016 [24].

The U-Net architecture was inspired by the encoder-decoder architecture [110, 72] as was shown in the previous section 4.2.1 *Fully Convolutional Networks*, which was first proposed in the context of autoencoders. Autoencoders are neural networks that are trained to reconstruct their input data, a leading work in this was done by G. E. Hinton [51] regarding the utility of autoencoders for dimensionality reduction and feature learning. The encoder-decoder architecture consists of an encoder that compresses the input data into a low-dimensional representation and a decoder that reconstructs the original data from the low-dimensional representation, see Figure 6 from [71].

The U-Net architecture extends the encoder-decoder architecture by adding skip connections between the encoder and decoder path, see Figure 8. These skip connections allow the decoder to access high-resolution feature maps from the encoder, which improves the accuracy of the segmentation. The skip connections were a key contribution of the U-Net architecture and have since been used in other segmentation models.

The effectiveness of the U-Net architecture in segmenting biomedical images, including CT and MRI scans was greatly proved since then, for example in brain tumor segmentation, lumbar spine segmentation and various other MRI and CT segmentation applications [117, 116, 100, 89, 43, 103]. Since its introduction, the U-Net architecture has been widely adopted researchers and has been applied to a variety of segmentation tasks. Also there were several different modifications and extensions proposed to the U-Net architecture to improve its performance in different contexts.

One such modification is the Attention UNet, which was introduced in a paper titled ”Attention U-Net: Learning Where to Look for the Pancreas” by researchers from the University of Pennsylvania [92]. An attention mechanism is used there to help the model focus on more informative features and ignore irrelevant ones. This can be particularly useful in processing data where certain areas of the images might be more relevant for a given task than others. This modification has been applied suc-

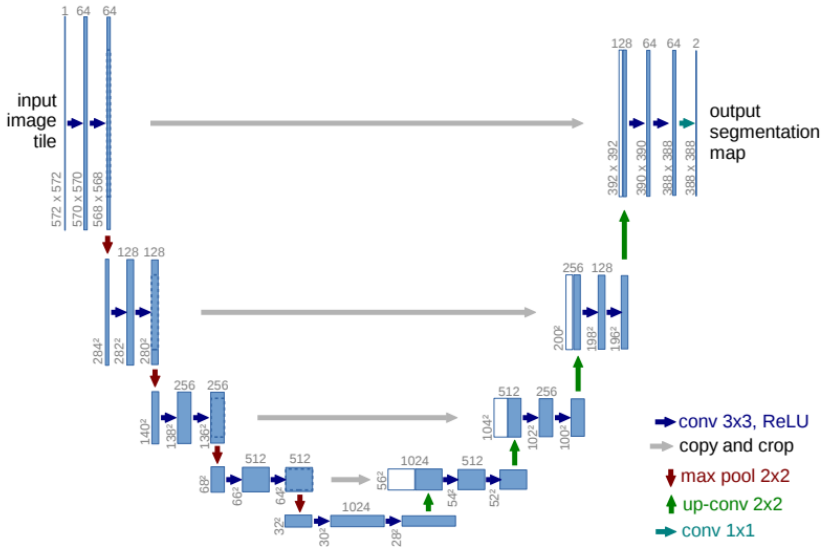

Figure 8: Original U-Net architecture design of Ronneberger et al. from their paper [97]. Skip connections are added from the encoding to the decoding path.

cessfully to pancreas segmentation in CT scans.

Another modification is the ResUNet, which was introduced in a paper titled "ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data" by researchers from the Indian Institute of Technology (IIT) Delhi. The ResUNet adds residual connections to the U-Net architecture, which help to mitigate the vanishing gradient problem and to facilitate the training of very deep networks by providing an alternate pathway for the gradient to flow through. Other successful enhancements of the original U-Net which were developed in the context of brain tumor segmentation will be presented in section 5.2.

U-Net models often outperform Fully Convolutional Networks where precise localization is crucial. Mask R-CNN based models are generally more capable when the focus is on instance segmentation where the R-CNN's object detection roots are beneficial. But especially in biomedical image analysis where detailed boundaries and small details in a semantic segmentation context are required, the U-Net architectures are widely superior to R-CNN models. This is largely due to U-Net's extensive use of skip connections, which help combining coarse, semantic information with fine, spatial information [5, 69].

4.3 Metrics

4.3.1 Pixel-Wise Accuracy

Pixel-wise accuracy is one of the basic and straightforward evaluation metrics used in image segmentation, where each pixel in an image is classified into a particular class. Pixel-wise accuracy computes the fraction of pixels that are correctly classified over the entire image or dataset [71, 97]:

$$A = \frac{1}{N} \sum_{i=1}^{N} \delta(p_i, g_i) \tag{14}$$

Where N is the total number of pixels and

$$\delta(p_i, g_i) = \begin{cases} 1 & \text{if } p_i = g_i \\ 0 & \text{if } p_i \neq g_i \end{cases} \tag{15}$$

is an indicator function that outputs 1 if the predicted label matches the ground truth label and 0 otherwise. The accuracy is then the average of these values across all pixels.

Another representation of pixel-wise accuracy, in binary classifier terms is:

$$ACC(X,Y) = \frac{TP+TN}{TP+FP+TN+FN} \tag{16}$$

where TP and FP are true and false positives, TN and FN are true and false negatives.

The obvious advantage of pixel-wise accuracy is that it is straightforward to understand and to compute. The major downside is that it can be widely misleading, especially when the class distributions are imbalanced. If one label is dominant, the network will tend to predict the dominant label, since this will give high pixel-wise accuracy anyway. For instance, in medical images where a pathology might be present in only a small part of the image, achieving high pixel-wise accuracy can be trivial by simply classifying every pixel as the dominant (healthy) class.

In [38] a comprehensive review of commonly used metrics for semantic segmentation, including pixel-wise accuracy, to evaluate the performance of segmentation algorithms is discussed.

4.3.2 Sørensen–Dice Coefficient

Another common metric in semantic segmentation, especially in medical applications, is the so called Sørensen–Dice coefficient (or only Dice-coefficient), also referred to as F1-score. It is based on the independent works of Thorvald Sørensen [108] and Lee Raymond Dice [31] and is used as a measure for comparing two discrete sets. In the case of segmentation it compares the predicted segmentation area to the ground-truth area.

The Sørensen–Dice coefficient serves as a measure of similarity between two sets of data and is commonly used in image processing and computational biology to evaluate the performance of segmentation algorithms. It is defined as two times the size of the intersection of the two sets divided by the sum of the sizes of the two sets. Therefore the coefficient ranges from 0 to 1, with a value of 1 indicating perfect similarity and a value of 0 indicating no similarity:

$$Dice(X,Y) = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FP + FN} \tag{17}$$

In image processing, the Dice-coefficient is often used to evaluate the performance of image segmentation algorithms, which aim to identify and separate objects or regions of interest within an image. Another application example is in computational biology, where the Dice-coefficient can be used to evaluate the similarity of predicted and observed protein structures. Protein structures are often represented as 3D models, and the Dice coefficient can be used to compare the similarity of two such models. This is important in the field of protein structure prediction, where the goal is to predict the 3D structure of a protein based on its amino acid sequence.

In the context of this work or in medical semantic image segmentation in general, a segmentation model may be used to identify and isolate the region of a MRI or CT scan that represents a tumor. The Dice-coefficient can then be used to evaluate the accuracy of the segmentation by comparing the algorithm's finding to a ground truth segmentation [122].

The Dice-coefficient has the advantage of being symmetric, meaning that it does not matter which set is considered the predicted set and which is considered the ground truth set. Additionally, it is less sensitive to the size of the sets than other measures of similarity.

However, it has some limitations as well, for instance, it can be over-sensitive to the size of the structures being compared. It tends to give excessively low similarity scores for small structures, which can be a significant drawback in medical imaging where small anatomical structures are important. Another limitation is that it treats all errors the same, whether they are false positives, false negatives, or total misses.

This lack of distinction can be problematic in cases where different types of errors have different consequences.

4.3.3 Hausdorff Distance

The Hausdorff distance is a mathematical concept that is used to quantify the similarity or dissimilarity between two sets of points, typically used in image processing and computer vision to evaluate the performance of image registration and object recognition algorithms [47, 55, 111]. It is defined as the greatest distance from any point in one set to the closest point in the other set:

$$d_H(X,Y) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x,y), \sup_{y \in Y} \inf_{x \in X} d(x,y) \right\}$$

This means the Hausdorff distance is determined in the following steps:

1. For each point in set X, find the distance to the nearest point in set Y.
2. For each point in set Y, find the distance to the nearest point in set X.
3. The Hausdorff distance is the maximum of the maximum distance from step 1 and the maximum distance from step 2.

The given definition is the typically used *symmetric* version of the Hausdorff distance. The so called *directed Hausdorff distances* would only follow step 1 or 2 of the above, without selecting the maximum of the both.

In image processing, the Hausdorff distance is often used to evaluate the performance of image registration algorithms, which aim to align two or more images of the same scene taken at different times or from different viewpoints. The Hausdorff distance can be used to measure the degree of alignment between the images by comparing the locations of corresponding features in the images. In computer vision, the Hausdorff distance is also used to evaluate the performance of object detection algorithms. It measures the dissimilarity between the predicted bounding boxes of an object and the ground truth bounding boxes.

In the context of semantic image segmentation, set X could represent the pixels of the ground truth segmentation of an object, while set Y represents the pixels of the predicted segmentation. The Hausdorff distance then provides a measure of the largest local discrepancy between the two segmentations.

However, it also has some limitations as well. It is sensitive to the density of points in the sets, making it more appropriate for sets with a similar number of points. Additionally, it is not invariant to translation and rotation, meaning that the results may vary depending on the position and orientation of the sets. It can also be sensitive to outliers, as a single incorrect pixel in the predicted segmentation can lead to a large

Hausdorff distance. As a result, in medical image analysis and other fields, the modified 95th percentile Hausdorff distance (i.e., the distance below which 95% of the pairwise distances lie) is often used to reduce sensitivity to outliers [33].

4.4 Loss functions

4.4.1 Cross-Entropy Loss

Cross-entropy is a measure that quantifies the difference between two probability distributions. The concept of cross entropy is widely used in machine learning, particularly in classification problems, as a loss function and was first established in the work of Claude Shannon on information theory, particularly his definition of entropy as a measure of information [102].

The cross-entropy H between the true distribution P and the predicted distribution Q generally is defined as:

$$H(P,Q) = -\sum_{c=1}^{C} P(c)\log Q(c) \tag{18}$$

with the *true distribution P* being the actual distribution of the data, the *predicted distribution Q* being the distribution predicted by the model and an index c over all classes. The logarithm is typically taken to be the natural logarithm, but the base can be changed depending on the context.

In the context of machine learning, particularly in classification tasks $P(c)$ is usually taken to be the ground-truth label, i.e. it is 1 for the correct class and 0 for all others. $Q(c)$ is the predicted probability of class c, often obtained from the softmax function when resulting from neural networks.

For a **binary classification** task, the cross-entropy loss (or then often called the *logistic loss*) simplifies to:

$$L(y,p) = -(y\log(p) + (1-y)\log(1-p)) \tag{19}$$

where y is the true label (0 or 1) and p is the predicted probability of the positive class.

For **multi-class classification**, with C classes the definition is therefore

$$L(y,p) = -\sum_{c=1}^{C} y_c \log(p_c) \tag{20}$$

where

- C is the total number of classes
- y_c is 1 if the true class is c and 0 otherwise
- p_c is the predicted probability for class c

Cross-entropy penalizes confident and wrong predictions more heavily than non-confident wrong predictions. Moreover its gradient works well with optimization algorithms, such as gradient descent. The gradient of cross-entropy can be shown to be beneficial for neural nets and deep learning applications, it's derivation can for example be found in Ian Goodfellow's *"Deep Learning"* book [41].

Given a vector of raw scores or logits z for each class, the **softmax** function $\sigma(z)$ is used to convert these scores into normalized probabilities. The softmax vector's component $\sigma(z)_c$ for class c is defined as:

$$\sigma(z)_c = \frac{e^{z_c}}{\sum_{j=1}^{C} e^{z_j}} \tag{21}$$

where

- C is the total number of classes
- z_c is the logit for class c
- $\sigma(z)_c$ is the predicted probability for class c after applying the softmax function.

The cross-entropy L in a multi-class application using softmax is then given as

$$L(y, \sigma(z)) = -\sum_{c=1}^{C} y_c \log(\sigma(z)_c) \tag{22}$$

having the gradient

$$\frac{\partial L}{\partial z_k} = \sigma(z)_k - y_k \tag{23}$$

which is an favorable equation since $\sigma(z)_k$ is the predicted probability of class k and y_k is the true label for class k, which is either 1 (for correct class) or 0 (otherwise). The gradient of the loss with respect to the logits is generally crucial, as it is used

to backpropagate the error and update the weights in neural networks. Here the gradient results in a very simple form and essentially measures the difference between the predicted probability and the true label for each class. It is then used in optimization algorithms like gradient descent to adjust the parameters in the direction that decreases the loss. If the model's prediction is correct (i.e. high probability for the correct class), the gradient will be small, resulting in a minor update to the weights. If the model's prediction is incorrect, the gradient will be large, pushing the weights to adjust more significantly. Also the gradient will then be used with the chain rule for derivatives of composed functions to further backpropagate through earlier layers and update weights. The chain rule is foundational in computing gradients for deep networks and the efficiency and simplicity of the given gradient equation is one of the reasons why cross-entropy loss combined with the softmax activation is very popular in classification problems in deep learning.

4.4.2 Weighted Cross-Entropy Loss

In Olaf Ronneberger's influential work on the U-Net architecture for biomedical image segmentation [97], he used a pixel-wise soft-max over the final feature map combined with a **weighted cross-entropy loss** function for training. This loss function L is defined as:

$$L(w) = \sum_{x \in \Omega} w(x) \log(p_{c(x)}(x)) \tag{24}$$

with

- Ω is the set of all pixels in the training samples
- $p_{c(x)}(x)$ is the predicted probability of the true class label $c(x)$ at pixel x
- $w(x)$ is a weight map introduced to give some pixels more importance during training. In the original U-Net paper [97], this weight map was used to give more weight to the border between segmented cells to improve separation

The weights w which are used here are generalizing and replacing the binary factors $y \in \{0,1\}$ in the basic cross-entropy definition. Instead, application or data adapted weights are assigned to each pixel, for example this can also be used to handle class imbalances present in a segmentation tasks.

4.4.3 Soft Dice Loss

The *Soft-Dice-Loss* (or Dice-Coefficient Loss) is derived from the Dice-coefficient, which was described in the previous section. For two sets X and Y, the Dice-coefficient $Dice(X,Y)$ is defined as:

$$\text{Dice}(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{25}$$

Since this formula is not differentiable, it can not be directly used as loss function for training deep learning models, so a soft version of this Dice-coefficient function is used instead. The Soft-Dice-Loss function is therefore defined as:

$$\text{SoftDice}(p,g) = 1 - \frac{2pg + \varepsilon}{p^2 + g^2 + \varepsilon} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \varepsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \varepsilon} \tag{26}$$

Where:

- N is the total number of pixels in the images.
- p the prediction vector, where p_i is the predicted probability of pixel i belonging to the class in question.
- g the ground truth vector, with g_i being the ground truth label 0 or 1 for pixel i.
- ε is a small constant to avoid division by zero.

For multi-class segmentation, as used in this work, the Soft-Dice-Loss will be computed for each class separately and will be then averaged, a code example is given in Appendix 9.1.

The Soft-Dice-Loss, especially in it's application to medical imaging, is often associated with the V-Net architecture proposed by Milletari et al in [85]. In this paper, the authors introduce the V-Net architecture and employ a differentiable loss function based on the Dice-coefficient, which is effectively the described Soft-Dice-Loss. The use of this loss allowed the network to produce probabilistic segmentations, which could then be thresholded to achieve binary segmentation results.

4.5 Morphological Filtering

Morphological filtering is an essential component of image processing, derived from the principles of mathematical morphology to process digital images based on their shape and structure. It can be used in various tasks, from noise reduction and image enhancement to feature extraction. The book "Morphological Image Analysis: Principles and Applications" by Soille and Serra [106] as well as the work "Image analysis using mathematical morphology" by Haralick, Sternberg & Zhuang [46], provide useful insights.

Traditionally and also in this work, these operations are applied to binary images, which only have two pixel values, typically indicating an 'object' and 'background'. In the scope of this work, they will be applied to segmentations masks, however, they

are generally also applicable to grayscale images. The operations use a so called structuring element or kernel, which is essentially a small binary or grayscale image, often in the form of a matrix, that determines the neighborhood to be considered during the operation. Common structuring elements include shapes like squares, circles, and lines. In application to 3D data, the structuring elements are given as 3D arrays, as shown in Figure 9.

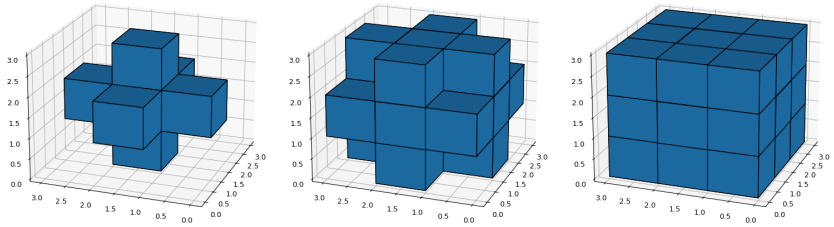

Figure 9: The figure shows typical binary 3D structures, which can be used for morphological operations. Left: rank 3, connectivity 1. Center: rank 3, connectivity 2. Right: rank 3, connectivity 3. The image source is SciPy v1.11.4 Manual [29].

The basic morphological operations are **erosion** and **dilation**. Erosion works by shrinking the image objects. For a binary image, if every pixel under the structuring element is 1, the output pixel value is set to 1, otherwise it becomes 0. Dilation, on the other hand, expands image objects. Here, the output pixel turns to 1 if any pixel under the structuring element is 1. Further morphological operations are opening and closing. Opening is an erosion followed by a dilation, which is particularly effective in removing noise. Closing, on the other hand is a dilation followed by an erosion and can therefore close small holes in objects.

Taking these principles into the context of grayscale images, grayscale morphology becomes an intuitive extension of binary morphology. For grayscale images, dilation and erosion consider the maximum and minimum pixel values under the structuring element, respectively [101, 40].

The applications of morphological operations are broad, for example in the field of noise reduction, where they can efficiently minimize or eliminate noise without blurring distinct edges in the image. Feature extraction is another significant application area, where morphological operations can emphasize or extract particular shapes or structures. They also find great use in image segmentation tasks as done in the presented application, where they will be used to separate previously merged brain tumor areas and therefore reconstruct training data.

Dilation

Dilation is one of the basic morphological operations, and its mathematical foundation will be given in the following. Consider a given image A and a structuring element B. In the context of binary images, dilation can be defined using set operations. The dilation of image A by the structuring element B is denoted as $A \oplus B$ and is defined as:

$$A \oplus B = \bigcup_{b \in B} A_b = \bigcup_{a \in A} B_a \tag{27}$$

where A_b is the translation of A by b and B_a vice versa. The dilation operation is commutative and the structuring element B is thereby moved over the image A with the dilation $A \oplus B$ being the union of all translated elements B_a.

In the case of grayscale images, the dilation at a pixel (x, y) is given by:

$$(A \oplus B)(x, y) = \max_{(s,t) \in B} \{A(x - s, y - t) + B(s, t)\} \tag{28}$$

This means essentially sliding the structuring element B over the image A and, for each position, computing the local maximum value of the image A where the structuring element B overlaps.

To understand this operation visually the structuring element can be imagined as a small neighborhood or a shape. When binary image is dilated, the structuring element probes the image at each position. If the structuring element overlaps anywhere in the object in the image, it sets the corresponding central pixel in the output image to 1 (or the maximum pixel value in the case of grayscale images). This results in expanding or growing the white regions in a binary image or bright regions in a grayscale image.

Erosion

Given an image A and a structuring element B, the erosion of A by B is typically denoted as $A \ominus B$. In the context of binary images, erosion can be understood through set operations and the erosion of A by B is defined as:

$$A \ominus B = \{z \in E \mid B_z \subseteq A\} \tag{29}$$

where E is the domain containing A and B_z is the translation of B by z. In the application of grayscale images, the erosion at a pixel (x,y) is given as

$$(A \ominus B)(x,y) = \min_{(s,t)\in B} \{A(x+s,y+t) - B(s,t)\} \tag{30}$$

where the structuring element B is moved across the image A and for each position and the local minimum value of A under B is calculated.

To visualize erosion, the structuring element can be considered as a small neighborhood or shape. When eroding a binary image using this element, the structuring element attempts to "fit" within the image's objects. If the structuring element fits completely inside an object in the image, the corresponding central pixel in the output image is set to 1, otherwise to 0. For grayscale images, the idea is similar, but instead of a binary outcome, the local minimum value where the structuring element fits is computed. Consequently, erosion typically shrinks the masks in binary images or reduces bright areas in grayscale images.

Opening

The morphological operation of *opening* is a sequence of two operations, erosion followed by dilation. It is mainly used to remove noise, especially small objects or details from an image, while largely preserving the shape and size of other objects.

Given an image A and a structuring element B, the opening of A by B is denoted as $A \circ B$ and is defined as the composition:

$$A \circ B = (A \ominus B) \oplus B \tag{31}$$

- **Erosion $A \ominus B$:** First, the image A is eroded by B. This will generally shrink the white or bright regions in the image, removing small objects or details smaller than the structuring element.
- **Dilation $(A \ominus B) \oplus B$:** Then, the eroded image is dilated using the same structuring element B. This process restores the size of the objects that were not entirely removed by erosion, ensuring that larger objects return to a size close to their original.

The outcome of this combined operation is that small objects, noise, or details, especially those that are smaller than the structuring element, get eliminated. At the same time, larger objects, which survive the erosion, get restored to their approximate original size after the dilation. The operation can be applied in multiple iterations to

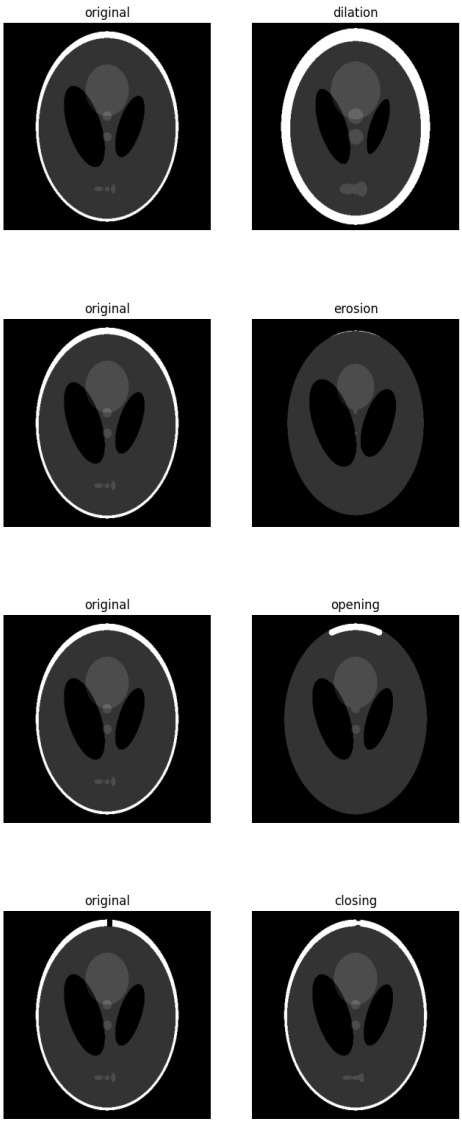

Figure 10: Examples of morphological operations. Dilation: The white border region is extended by the dilation operation, the small center spots are enlarged. Erosion: The white border region is shrinked by the erosion operation, the size of the center spots is reduced, some are deleted. Opening: Consecutive application of erosion and dilation. Closing: Consecutive application of dilation and erosion, note the small gap in the white border region, which is getting filled by the closing operation. The image source is Scikit-Image documentation [114].

increase the effect. Figure 10 illustrates a schematic example of this operation.

Closing

Closing is another fundamental morphological operation, essentially the reverse of opening. It involves performing dilation followed by erosion. Closing is useful for tasks like closing small holes within objects or bridging gaps between objects.

Given an image A and a structuring element B, the closing of A by B is denoted as $A \bullet B$. It is defined as:

$$A \bullet B = (A \oplus B) \ominus B \tag{32}$$

with the two operations being performed in composition:

- **Dilation** $A \oplus B$**:** First, the image A is dilated by B. This operation generally expands the white or bright regions in the image, filling small gaps or holes and connecting nearby objects.
- **Erosion** $(A \oplus B) \ominus B$**:** Subsequently, the dilated image undergoes erosion using the same structuring element B. This step shrinks the white or bright regions back down. In the process, any structures or details that became connected or filled during dilation and are larger than the structuring element remain, while smaller extraneous details introduced by the dilation are removed.

The effect of closing is to bridge narrow breaks and fill small holes and gaps without causing significant changes to the larger structures in the image. An example of 3D closing is shown in Figure 11. Closing is often employed in image post-processing to heal artifacts in binary segmentations, close up holes in detected objects, or to bridge gaps in lines or object boundaries.

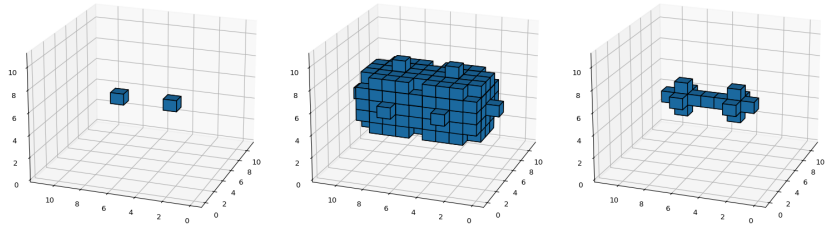

Figure 11: Example of binary 3D dilation followed by erosion. Left: Original. Center: Binary dilation of original using a 3D ball structure with radius 3. Right: Binary erosion of the previous using a 3D ball structure with radius 2. For having a simple and demonstrative example, different structuring elements are used here for dilation and erosion. The image source is SciPy v1.11.4 Manual [29].

5 MICCAI BraTS Challenge

5.1 Datasets and Labels

The BraTS (Brain Tumor Segmentation) datasets are collections of MRI brain scans used for the evaluation of brain tumor segmentation algorithms. They are made available for the Multimodal Brain Tumor Segmentation (BraTS) challenge, which is organized annually by the International MICCAI Society [23].

The BraTS datasets contain large number of MRI scans of patients with brain tumors, containing the four different MRI modalities: T1-weighted, T2-weighted, contrast-enhanced T1-weighted (T1ce) and Fluid Attenuated Inversion Recovery (Flair). Each scan also includes a manually-annotated ground truth segmentation of different tumor regions, which is used to evaluate the performance of segmentation algorithms.

The BraTS datasets have become an important benchmark for the evaluation of brain tumor segmentation algorithms, as they provide a large and diverse set of data that can be used to test the performance of different models and algorithms. They have played a significant role in the development of new prediction algorithms in this field and to help improving existing ones. The data made available by the BraTS dataset is a very valuable resource for the research work on brain tumor segmentation. The large and diverse set of data can be used to evaluate the performance of different algorithms and plays a significant role in the development of new methods. The BraTS challenge provides an important platform for pushing the development of new and more accurate brain tumor segmentation algorithms. A main challenge in the BraTS dataset is the high variability in the appearance of brain tumors. Tumors can vary in size, shape, and intensity, making it difficult for models to accurately segment them. Additionally, the presence of healthy tissue, necrosis, and edema within the tumor region can also add to the complexity of the segmentation task. To address these challenges a variety of segmentation algorithms have been developed and have demonstrated promising results on the BraTS dataset. In the coming paragraphs, some of the most successful models from the competitions of the last years will be discussed.

In addition to the standard BraTS MRI data, the challenge also includes a track for the segmentation of low-grade gliomas which is a subtype of brain tumors that are usually less aggressive and have a better prognosis and a track which includes survival data for a smaller group of patients [53, 2, 1]. Both of these will not be used in the presented work.

The BraTS datasets changed and evolved over the years in terms of volume, meaning that there were more records added over the years or in some years the whole previous dataset was left behind and a new one was started with new segmentation criteria. In the different years, the definition and number of brain tumor regions was thus subjected to variation. Hence, in this chapter, the datasets of different years,

namely BraTS 2015, BraTS 2018, BraTS 2021 and BraTS 2022 will be presented.

5.1.1 BraTS 2015

The BraTS 2015 [83] dataset consists of 220 high grade glioma (HGG) and 54 low grade glioma (LGG) MRIs. The data is co-registered, skull-stripped, annotated and it extends the data from the preceding years. All data was annotated manually by experts and went through a review and approval process. The segmentation labels given in the BraTS 2015 dataset are shown in Table 3.

Label	Data Name	Description
0	*BG*	Background
1	*NCR*	Necrosis
2	*ED*	Edema
3	*NET*	Non-Enhancing Tumor
4	*ET*	Enhancing Tumor

Table 3: BraTS 2015 dataset labels.

For evaluation, the BraTS challenges do not use the labels given in the training data, instead hierarchical combinations of the labels given in the dataset are used [83, 8, 6]. This leads to composed evaluation segments *AT*, *TC* and *WT* as given in Table 4.

Evaluation segment	Description
AT	Active tumor
TC	Tumor core
WT	Whole tumor

Table 4: BraTS evaluation segments.

The listed evaluation segments are the same in all discussed BraTS challenges (2015, 2018, 2021, 2022). But the provided data labels as well as the relation between data labels and evaluation segments varies between the BraTS datasets of the different years. In the 2015 dataset, the data labels *NCR*, *ED*, *NET*, *ET* and the evaluation segments *ET*, *TC*, *WT* are linked as described in Table 5.

Figure 12 shows an example as provided by the BraTS 2015 announcement paper [83]. In the image on the right side the data labels are shown as colored regions where the necrotic core *NCR* is colored green, the enhancing tumor *ET* is blue, the non-enhancing tumor *NET* is red and the edema *ED* is yellow. The three image pairs on the left side are showing the evaluation segments, the bottom images show the plain MRI slices, the top images showing the segmentation masks. From left to right there is the whole tumor region *WT* (shown all yellow), the tumor core *TC* (shown

Evaluation Segment	BraTS 2015 Data	Brats 2015 Labels	Description
AT	*NCR + ET*	1, 4	Active tumor
TC	*NCR + ET + NET*	1, 4, 3	Tumor core
WT	*NCR + ET + NET + ED*	1, 4, 3, 2	Whole tumor

Table 5: BraTS 2015 evaluation segments as combinations of data labels.

red) and the active tumor (shown green and blue).

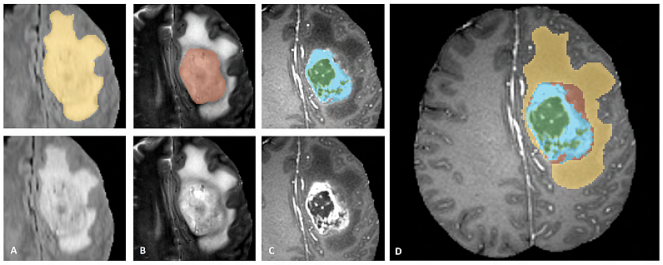

Figure 12: From BraTS 2015 [83]. Left: Three image pairs showing the evaluation segments, whole tumor *WT* (yellow), tumor core *TC* (red), active tumor (green and blue). Right: Colors showing data labels, necrotic core *NCR* (green), enhancing tumor *ET* (blue), non-enhancing tumor *NET* (red), Edema *ED* (yellow)

It should be noted, that compared to the datasets of BraTS 2018 and 2021 this is the only one, which provides a dedicated label for the Non-Enhancing *NET* region. But despite having the label, it does not provide very consistent *NET* data with respect to the definition in the latest literature as discussed in section 2. For this reason, as well as for the minor data quantity and partially very low MRI records quality, the BraTS 2015 data was not used as a reliable source for the identification of non-enhancing tumor parts in this work.

5.1.2 BraTS 2018

The BraTS 2018 dataset [8] provides a set of 285 multimodal MRI records of human brains, each containing scans of a brain with tumor (glioblastoma/high-grade-glioma GBM/HGG or lower grade glioma LGG) in four modalities T1, T1ce, T2 and Flair similar to the Brats2015 dataset. The segmentation masks were also created and proved by human experts, similar to the one of BraTS 2015. The data provided in BraTS 2018 does not extend the BraTS 2015 data, but instead is a completely new dataset and therefore closing the previous track of BraTS 2015. The records have a spatial resolution of 1x1x1mm and the dataset also contains overall survival data for each of the records, which will be not considered by this work.

It also contains segmentation masks for different tumor regions which are comparable to those of Brats2015, but with a notable difference. The non-enhancing regions

are now incorporated into the necrosis parts. The list of data labels present in the BraTS 2018 dataset is shown in Table 6.

Label	Data Name	Description
0	*BG*	Background
1	*NCR + NET*	Necrosis + Non-Enhancing Tumor
2	*ED*	Edema
3	-	not used
4	*ET*	Enhancing Tumor

Table 6: BraTS 2018 dataset labels.

The announcement paper and documentations for Brats2018 [8] do not state any reason for this change, also any obvious medical or technical reason is missing. Figure 13 shows sample images of the BraTS 2018 dataset. The combined *NCR + NET* (label 1) region is shown in red. It can be observed, that this red region captures the inner necrotic area as well as the non-enhancing parts lying around the enhancing tumor core. Typically the non-enhancing *NET* region is found as a border region between the green edema part and the yellow enhancing part, while the necrotic *NCR* area is a scattered or uniform area within the yellow enhancing region. Nevertheless, in BraTS 2018 the two distinct regions *NCR* and *NET* are combined to a single segmentation label.

The final regions which are used for evaluation are similar but not all the same as in BraTS 2015, since they are combined using the pre-merged *(NCR + NET)* subregion. The combinations of evaluation segments from data labels is shown in Table 7.

Evaluation Seg.	BraTS 2018 Data	BraTS 2018 Labels	Description
AT	*ET*	4	Active tumor
TC	*ET* + (*NCR* + *NET*)	4, 1	Tumor core
WT	*ET* + (*NCR* + *NET*) + *ED*	4, 1, 2	Whole tumor

Table 7: BraTS 2018 evaluation segments as combinations of data labels.

It should already be sent ahead here, that in BraTS 2021 there will even be another variant of combining the non-enhancing tumor part.

5.1.3 BraTS 2021 and 2022

The BraTS 2021 challenge [6] provides a total of 2040 records where 1251 of them come as training data records together with labels. After BraTS 2015 and BraTS

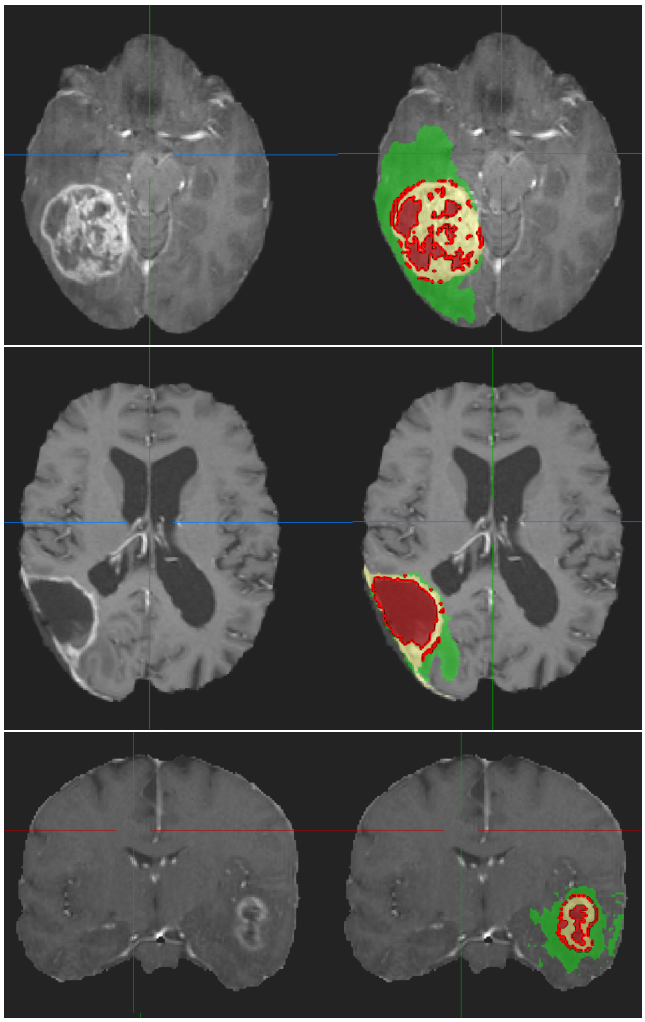

Figure 13: Brats 2018 sample images. Red: NCR + NET (label 1). Green: ED (label 2). Yellow: ET (label 4)

Label	**Data Name**	**Description**
0	*BG*	Background
1	*NCR*	Necrosis
2	*ED + NET*	Edema + Non-Enhancing Tumor
3	-	not used
4	*ET*	Enhancing Tumor

Table 8: BraTS 2021 dataset labels.

2018 the 2021 dataset comes with a third variation of data labels as listed in Table 8.

The evaluation regions are varied again compared to the 2015 and 2018 datasets. Now the name *active tumor AT* is dropped and only *enhancing tumor ET* is used as name for the smallest of the hierarchical parts, having the same name now as for the corresponding data label (image A in Figure 14). The tumor core *TC* region is now a combination of the center necrosis part and the surrounding enhancing tumor region (image B in Figure 14), but *without* the non-enhancing part, as it was in both previous datasets. Instead the non-enhancing parts are now combined to edema and are therefore only counted towards the total whole tumor *WT* domain (image C in Figure 14). A list showing the compositions of the evaluation segments is given in Table 9.

Evaluation Segment	**BraTS 2021 Data**	**BraTS 2021 Labels**	**Description**
ET	*ET*	4	Enhancing tumor
TC	*ET* + *NCR*	4, 1	Tumor core
WT	*ET* + *NCR* + (*ED* + *NET*)	4, 1, 2	Whole tumor

Table 9: BraTS 2021 evaluation segments as combinations of data labels.

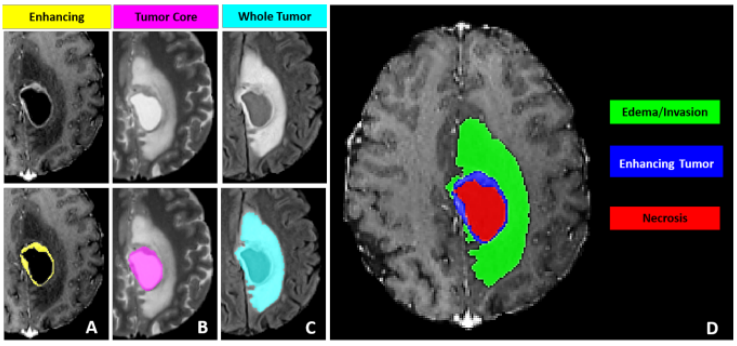

Figure 14: From Brats 2021 paper [6]. Left side showing the evaluation regions ET, TC and WT. Right side showing the data labels NCR, ET and ED+NET.

Figure 15 shows some sample slices from the BraTS 2021 dataset with the three given labels. Compared to Figure 13 from BraTS 2018 the non-enhancing regions are not distinctly observable. They are usually found in the border region between edema (green in Figures 13 and 15) and enhancing tumor (yellow in both figures). Since BraTS 2021 counts the non-enhancing regions towards edema, the small non-enhancing parts visually disappear within the large outer Edema region.

The training and validation data of the BraTS 2022 dataset are the same as in BraTS 2021. Only the testing dataset has been updated with more MRI scans, according to announcement [12].

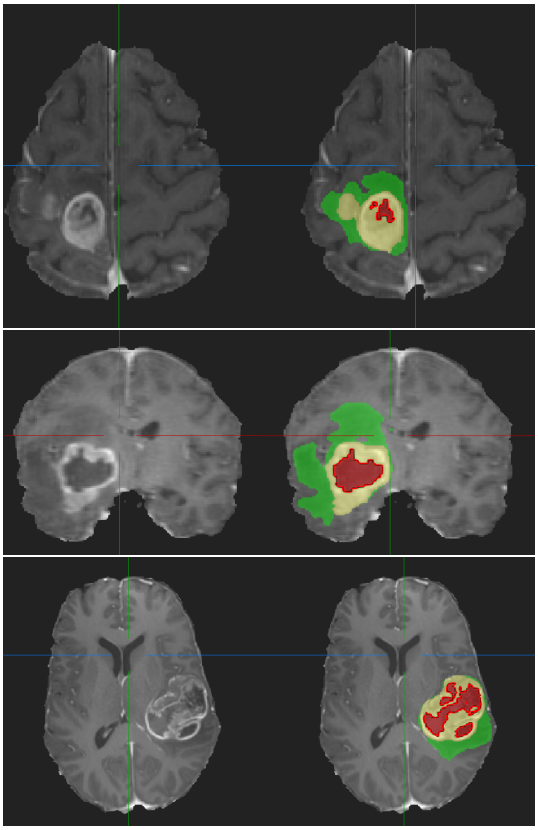

Figure 15: Brats 2021 sample images. Red: NCR (label 1). Green: ED + NET (label 2). Yellow: ET (label 4)

5.2 BraTS Winning Models

This section will present the winning deep learning models of the MICCAI BraTS brain tumor segmentation challenge of the years 2018, 2019, 2020 and 2021.

5.2.1 Residual U-Net with Autoencoder Regularization (SegResNetVAE)

The SegResNetVAE network was introduced by Andriy Myronenko [87] and won 1st place in the BraTS 2018 brain segmentation challenge. It combines three different concepts in deep learning: Segmentation, ResNet, and Variational Autoencoders (VAEs). The segmentation part was already introduced extensively in the previous chapters and is implemented in form of a basic U-Net architecture in [87].

The **ResNet** (short for Residual Network) concept, which is also used in the SegResNetVAE model, is a type of convolutional neural network architecture introduced by Kaiming He et al. in 2015 in a paper titled "Deep Residual Learning for Image Recognition" [49]. ResNet is characterized by its use of residual connections or skip connections, which allow the network to learn residual functions or shortcuts between layers. This design helps to mitigate the vanishing gradient problem, allowing for the training of much deeper networks than was previously possible. In general ResNets have been widely adopted in various computer vision tasks, including image classification, object detection, and also image segmentation. In this special application context, Myronenko uses residual connections within the horizontal filtering parts of his U-Net design to support backpropagation and avoid vanishing gradients. Skip connections like these will also be used in the model presented later in this work.

The third major concept of SegResNetVAE are **Variational Autoencoders (VAEs)**, which are a type of generative model introduced by Kingma and Welling in 2013 in the paper "Auto-Encoding Variational Bayes" [61]. VAEs combine concepts from deep learning and Bayesian inference to learn a latent representation of the input data in an unsupervised manner. VAEs consist of an encoder that maps input data to a latent space and a decoder that reconstructs the input data from the latent space. A key aspect of VAEs is the incorporation of a variational inference technique that allows the model to learn a probabilistic distribution over the latent space, enabling it to generate new samples similar to the input data.

In the SegResNetVAE model of A. Myronenko, VAE is implemented in a parallel decoding branch and, in the original version, is used as a regularization method, see Figure 16. In later SegResNet architectures which were presented in 2019 and 2021 [88, 105, 37], no VAE regularization was used anymore. The validation results of the different iterations of this model on the BraTS 2018, BraTS 2019 and BraTS 2021 data is shown in Table 10.

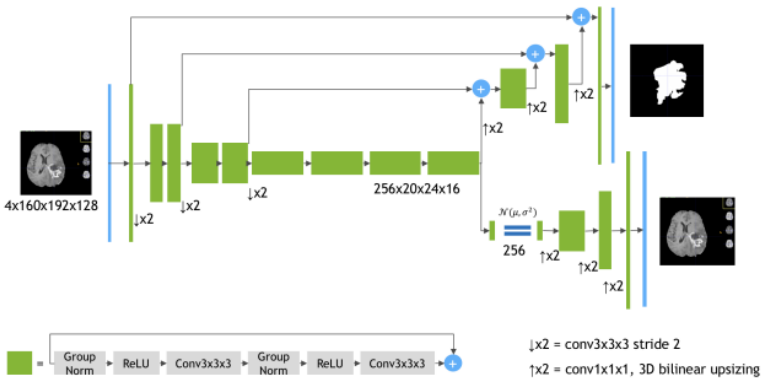

Figure 16: SegResNetVAE Network for brain tumor segmentation as used by A. Myronenko in [87]

	SegResNetVAE 2018 [87] (Single model)	**SegResNetVAE 2018 [87] (Ensemble of 10 models)**	**SegResNet 2019 [88] (Single model)**	**SegResNet 2021 [105] (Single model)**
ET	0.8145	0.8233	0.800	0.8600
TC	0.8596	0.8668	0.834	0.8868
WT	0.9042	0.9100	0.894	0.9265
Mean	**0.8594**	**0.8667**	**0.843**	**0.8911**

Table 10: Validation results for SegResNetVAE 2018, SegResNet 2019 and SegResNet 2021 given by A. Myronenko et al. [87, 88, 105]. The table shows the Sørensen–Dice-coefficients for the enhancing tumor (ET), tumor core (TC) and whole tumor (WT) regions as well as their mean, evaluated on the corresponding validation sets.

5.2.2 Two-Stage Cascaded U-Net (TSC U-Net)

The Two-Stage Cascaded U-Net, developed by Zeyu Jiang et al. [59], is a state-of-the-art model developed mainly for brain tumor segmentation. The main publication detailing the Two-Stage Cascaded U-Net is titled "Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task" and was published online on Springer in May 2020. The authors presented their model in the context of the BraTS Challenge 2019 segmentation challenge [7], showcasing how the novel two-stage cascaded U-Net could effectively segment the substructures of brain tumors from coarse to fine. Their *Two-Stage Cascaded U-Net model (TSC U-Net)* scored 1st place in the BraTS 2019 challenge.

The overall design of the used model is a two-step structure for accurately outlining the different regions of the brain tumor. The first stage of the model starts on a broader segmentation task, finding a coarse outline of the tumor regions. This preliminary segmentation serves as a basis for the second stage, which incorporates a more refined segmentation task, building out finer details of the tumor substructures with an enhanced level of precision. The particular design of the second stage, constructed with a higher number of channels and dual decoders, amplifies the model's performance and enables a precise delineation of the tumor regions.

The two-stage design is essential for boosting the performance of the segmentation task. By initially focusing on broader segmentation and subsequently refining the details, the model can potentially achieve high accuracy and robustness in segmenting brain tumor substructures.

The architecture is shown in Figure 17 and operates in two distinct stages to improve the segmentation accuracy progressively. In general it is based on the architecture of the U-Net but incorporates two U-Net like stages which follow each other. The first stage aims to provide a coarse segmentation of the tumor regions and follows a standard U-Net architecture for an initial coarse segmentation of the tumor substructures. This stage is crucial as it sets the groundwork for the finer segmentation carried out in the next stage. Building upon the coarse segmentation from the first stage, the second stage focuses on refining the segmentation to delineate finer details of the tumor substructures.

The second stage of the model is designed with a higher number of channels, which allows the model to capture more complex features of the tumor substructures. Additionally, the use of two decoders in this stage is a significant design choice. Dual decoders may aid in handling the hierarchical nature of the segmentation task more effectively and allow the model to learn and represent different levels of details in the segmentation. The second stage has a higher number of channels, which supports for capturing more complex features. The utilization of two decoders is a novel addition which likely helps in handling the hierarchical nature of the segmentation task, allowing for more refined segmentation of different substructures. The key innovation

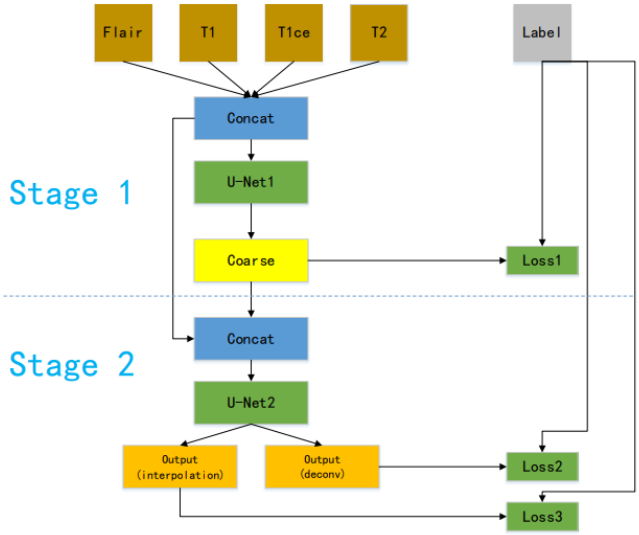

Figure 17: TSC U-Net architecture from original paper [59], showing the two cascaded U-Nets, the second stage using a dual decoder.

is the cascading of the two U-Nets. The coarse segmentation from the first stage is used as an input to the second stage, enabling a more refined segmentation. This cascading approach allows for a more hierarchical understanding and delineation of the tumor substructures.

The entire network was trained end-to-end, meaning that the gradients are back-propagated through both stages of the network during training, optimizing the model to better learn the hierarchical representation of the data for improved segmentation accuracy. This also ensures that the learning process is optimized across both stages of the model, enabling the network to learn a hierarchical representation of the data for improved segmentation accuracy.

The BraTS 2019 validation results of Z. Jiang et al.'s TSC U-Net model from their original paper are listed in Table 11 for reference.

5.2.3 No New U-Net (nnU-Net)

The nnU-Net architecture, which stands for "no-new-Net", was developed by a research group led by Fabian Isensee and Klaus H. Maier-Hein [57, 58]. The model has gained significant attention for its excellent performance in various medical imaging challenges and was not exclusively intended for the use on brain tumor segmentation. The core idea behind nnU-Net is to provide a flexible and efficient pipeline that can be easily adapted to different medical imaging segmentation tasks without the need to develop new architectures. The framework is built upon the U-Net architecture and claims to use the original U-Net in a very plain way.

	TSC U-Net 2019 (Ensemble of 5-fold)	**TSC U-Net 2019 (Best single model out of 12)**	**TSC U-Net 2019 (Ensemble of 12 models)**
ET	0.7967	0.8020	0.8021
TC	0.8589	0.8632	0.8647
WT	0.9080	0.9082	0.9094
Mean	**0.8545**	**0.8578**	**0.8587**

Table 11: Validation results of Z. Jiang et al.'s Two-Stage Cascaded U-Net (TSC U-Net) for BraTS 2019 [59]. The table shows the Sørensen–Dice-coefficients for the enhancing tumor (ET), tumor core (TC) and whole tumor (WT) regions as well as their mean, evaluated on the corresponding validation sets.

Although named "No New U-Net" (nnU-Net) to indicate it doesn't introduce new architectural variations, the network optimizes the existing U-Net structure to extract its full potential and is also embedded in a framework which allows the automatic adaption of the nnU-Net to different medical segmentation tasks. Therefore it can automatically configure various aspects like preprocessing, network architecture, training, and post-processing. Being such adaptable, it demonstrated superior performance across 23 public datasets used in international biomedical segmentation competitions without any manual adjustments, thus making state-of-the-art segmentation accessible to a broader audience without necessity of expert knowledge or additional computing resources beyond standard network training [58, 98].

The essence of nnU-Net lies in its ability to self-adapt to new tasks, eliminating the need for manual intervention in configuring the network, thus accelerating the deployment of deep learning solutions in biomedical imaging. By applying nnU-Net to the *BraTS 2020* segmentation challenge, even the unmodified baseline configuration of the architecture yielded notable results. With the incorporation of BraTS-specific modifications regarding post-processing, region-based training, and a more aggressive data augmentation strategy the segmentation performance could even be improved. This resulted in winning the first place in the *BraTS 2020* challenge, the corresponding validation results from the paper are given in Table 12.

The winning model of the following BraTS 2021 challenge, was an extension of the previous nnU-Net by Huan Minh Luu and Sung-Hong Park. In their paper titled "Extending nn-UNet for Brain Tumor Segmentation" [78] they use the nn-UNet, as the basis for their methods. They explored several modifications to nn-UNet to improve its performance, including enlarging the network (by using a higher number of filters in the encoding path, as done in Myronenko's work previously [87]), substituting batch normalization with group normalization, and incorporating axial attention in the decoder. Using 5-fold cross-validation they demonstrated the effectiveness of their approach, achieving a slight improvement in quantitative metrics compared to the baseline model, see Table 13. Their proposed models achieved first place in the final ranking on unseen test data in the BraTS challenge of 2021.

	nnU-Net 2020 (BL)	**nnU-Net 2020 (BL, R+DA)**	**nnU-Net 2020 (BL, R+DA, BN+BD)**
ET	0.7664	0.7867	0.7945
TC	0.8370	0.8461	0.8524
WT	0.9093	0.9090	0.9119
Mean	**0.8376**	**0.8473**	**0.8529**

Table 12: Validation results of F. Isensee et al.'s No-New U-Net (nnU-Net) for BraTS 2020 [57]. BL: baseline nn-U-Net without modification, R+DA: region-based training and more aggressive data augmentation, BN+BD: batch normalization and batch dice. The table shows the Sørensen–Dice-coefficients for the enhancing tumor (ET), tumor core (TC) and whole tumor (WT) regions as well as their mean, evaluated on the corresponding validation sets.

	nnU-Net 2021 (BL)	**nnU-Net 2021 (BL, GN)**	**nnU-Net 2021 (BL, AA)**	**nnU-Net 2021 (BL, L+GN)**
ET	0.8837	0.8817	0.8723	0.8823
TC	0.9206	0.9211	0.9188	0.9235
WT	0.9378	0.9366	0.9321	0.9383
Mean	**0.9140**	**0.9130**	**0.9077**	**0.9147**

Table 13: Validation results of H. M. Luu et al.'s Extended nnU-Net for BraTS 2021 on 5-fold cross validation [78]. BL: baseline nnU-Net with batch normalization, GN: using group normalization, AA: using axial attention, L+GN: using larger U-Net and group normalization. The table shows the Sørensen–Dice-coefficients for the enhancing tumor (ET), tumor core (TC) and whole tumor (WT) regions as well as their mean, evaluated on the corresponding validation sets.

The axial attention extension which was used in their approach is based on a variant of self-attention mechanism designed to work efficiently with multi-dimensional data by aligning with the multiple dimensions of the tensors in both the encoding and decoding settings. It was initially introduced as criss-cross attention in 2018 by Z. Huang in a model named CCNet in the paper "CCNet: Criss-Cross Attention for Semantic Segmentation" [54]. The criss-cross attention gathers the contextual information of all the pixels along a criss-cross path. By employing a further recurrent operation, each pixel can therefore potentially capture the full-image dependencies. In the context of neural networks, particularly those used for segmentation tasks, axial attention can be adopted to ensure both global connection and efficient computation. For example, in the work "Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation" [118] axial attention was applied in a stand-alone self-attention setting. In this setup, an axial-attention layer is defined on the width-axis of an image as a one-dimensional position-sensitive self-attention, and a similar definition is used for the height axis. This is similar to how axial attention is used in the decoder of H. M. Luu's and S. H. Park's extended nnU-Net shown in Figure 18.

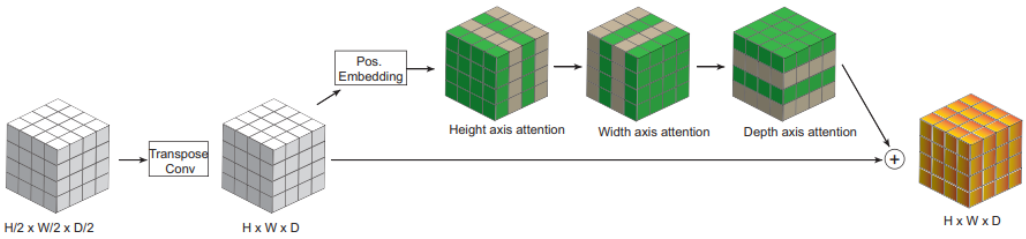

Figure 18: Axial attention applied to each axis in upsampling step as proposed by Luu et al. in [78]. The output is added back to the original input, which is then concatenated with the features from the U-Net's encoder path.

Furthermore, attention mechanisms enable the decoder to leverage the most relevant parts of the input sequence in a flexible manner through a weighted combination of all the encoded input vectors, with the most relevant vectors being attributed the highest weights [52]. This mechanism enhances the capability of neural networks in segmentation tasks by capturing long-range dependencies in the input data which is crucial for tasks like brain tumor segmentation, as mentioned in the paper by Luu and Park [78].

6 Segmentation of Non-Enhancing Regions

The main focus of this section is to provide an adapted segmentation model for the identification of the different tumor regions and especially focusing on the specific and often small-scale non-enhancing tumor areas. This model will be based on the previously presented BraTS winning models nnU-Net, TSC U-Net and mainly SegResNetVAE together with a newly introduced resolution enhancing extension on the decoder side. The basic model architecture will be defined and fixed in Section 6.1, but will include general filter blocks, which can be varied are evaluated against each other in Sections 6.1.2 and 6.1.3. Section 6.2 will introduce the upscaling decoder extension which will allow higher resolution outputs for the segmentation masks. Subsequently, Section 6.3 will define how the non-enhancing tumor area can be extracted from the BraTS 2018 data and will unify the two large BraTS datasets from 2018 and 2021. Also, it will provide all formal aspects to set up further training and evaluation runs, whereas Section 7 will evaluate the according results.

6.1 Basic Segmentation Model

6.1.1 3D U-Net with General Filter Blocks

Following the idea of nnU-Net, the overall construction of the used model will be very plain and near to the original U-Net as shown in Figure 19. The principal architecture is an encoding path followed by a decoding path, each going through four levels of spatial resolution and each of the three spatially related top levels having a skip connection from encoder to decoder side. The structure within the single layers is repeated with each level having a different number of filters and spatial size. The following paragraph will go from input to output and will give a detailed description of the single parts of the underlying architecture.

As a pre-processing step, the multi-channel input records will be center-cropped to size $80 \times 160 \times 128 \times 4$ (axial, coronal, sagittal, channel). Working with the original full record size of $155 \times 240 \times 240 \times 4$ will not be possible due to GPU memory limitations. Cropping the records to the given size will have nearly no effect regarding image data loss, since a wide outer area of the records is empty. The dataset labels NCR, ET and ED will be pre-processed to the evaluation segments ET, TC and WT as described in Section 5 and will be cropped to the same size, a code example is given in Appendix 9.2.

The input layer is followed by an initial 3D-convolutional layer with 32 filters of kernel size $3 \times 3 \times 3$ and a spatial dropout layer with dropout rate of 20%. In Ronneberger's original work [97] he states, that the dropout layer in his architecture is placed at the end of the contracting path but today the placing of a regularizing dropout layer right after the initial convolution is more common in many deep learning and image processing architectures [21].

The result of the concatenated filter block will be spatially downsampled to half size using $3 \times 3 \times 3$ convolutions with stride 2. In the original U-Net design the downsampling was done using 2×2 max-pooling, but stride 2 convolutions are capable of carrying parameters for the downsampling operation, whereas max-pooling is a fixed operation. By learning the downsampling operation, stride 2 convolutions can potentially learn more complex and task-specific downsampling patterns, which may lead to better performance. Also in some cases, stride 2 convolutions can be more computationally efficient than max-pooling combined with spatially invariant convolutions, since it combines both convolution and downsampling into a single operation. This can result in fewer overall operations to perform during both, the forward and backward passes [4]. In the continuing encoder path downsampling together with **two** filter block structures per resolution layer will be repeatedly applied until reaching a spatial resolution of $10 \times 20 \times 16$ with a feature depth of 256.

In the following decoding path, the upsampling is performed by applying transposed 3D convolutions of stride 2. After each of the upsampling steps, the result is summed with the skip connection path connecting from the encoding side. After reaching back to the original spatial shape another two residual filter blocks are applied, followed by a $1 \times 1 \times 1$ convolution. The final convolution uses three filters, yielding three output masks with sigmoid activation, which will be used as the three prediction probability masks for enhancing tumor, tumor core and whole tumor (ET, TC, WT). This completes a typical U-Net architecture with input shapes, channels and outputs being adapted for the application on the BraTS data introduced in Section 5.

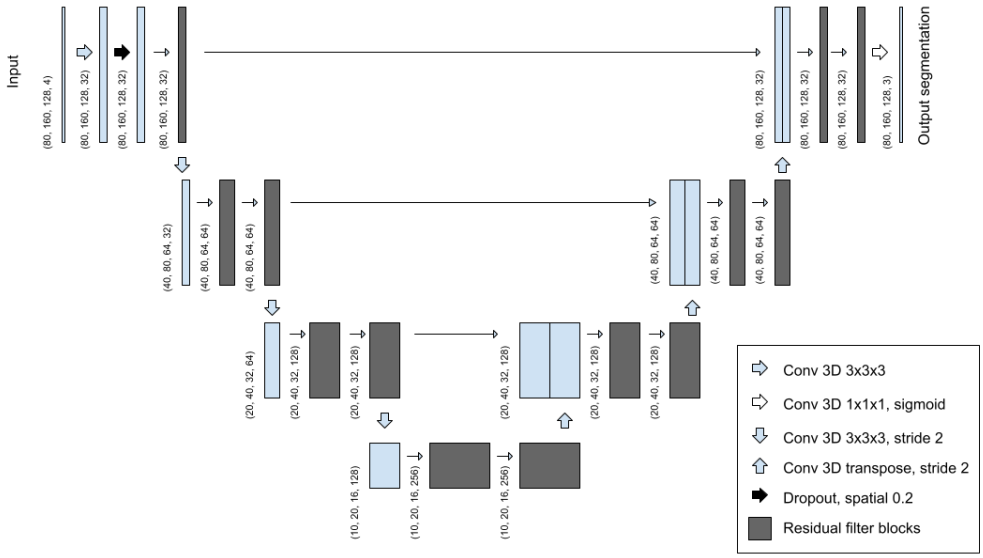

Figure 19: U-Net architecture with four spatial layers and with general residual filter blocks (gray). The skip connections are combined additively into the upsampling path.

6.1.2 Filter Block Structures

After giving the outer model architecture in the previous section, this part will introduce different possible implementations for the filter blocks in use. Historically, there are basically three main iterations of filter blocks which have been used. In the original U-Net paper [97], Ronneberger et al. used plain convolutions with ReLU activation as shown in the top row of Figure 20. Using these plain U-Net filter blocks gives a very original U-Net architecture as a reference.

The second model will use ResNet like filter blocks, similar to those introduced in the original ResNet paper [49]. The structure is the characteristic combination of two subsequent convolutions with an intermediate ReLU activation and an additive residual connection bypassing the operations. The schematic representation of this is shown in the second place of Figure 20.

In the third and fourth model, normalization blocks will be added. The common way to do this is to add normalization and activation as successor layers of the convolutional layer. The skip connection may end before or after the final activation. In this work the version with activation after the skip will be used. This is shown in the third row of Figure 20. Usually, batch normalization is a typical choice here, but since the model has high memory demand it will have to be trained with very low batch size (even batch size 1 in the training runs below) and thus limits the use of batch normalization. Instead, group normalization will be used, which is independent of batch size. A variation of this filter block is shown in the bottom row of Figure 20. The normalization and activation layers are placed *before* the convolution layers instead of following after them, therefore it is called *pre-activation filtering*. This structure is less common than the previous one, it is used by the BraTS winning model of Myronenko et al. [87] and will be added to the comparison. Appendix 9.3 shows the structure of the described filter blocks in Tensorflow code.

In the pre-activation structure, the activation functions (like ReLU) are applied before the convolution layers. This arrangement is used to improve the flow of gradients during backpropagation, making it easier to train deeper networks. This is because it reduces the vanishing gradient problem, a common issue in deep networks. Moreover, by applying normalization before the convolution, pre-activation blocks can reduce internal covariate shift, which is a change in the distribution of network activations due to the parameters update during training. This leads to faster training and requires less careful initialization. In their papers ”Deep Residual Learning for Image Recognition” and ”Identity Mappings in Deep Residual Networks”, He at al. introduced and developed the concept of using pre-activation in ResNet blocks [50, 49]. They demonstrated improved performance and easier training of very deep networks using this approach.

This results in four different models, which will be trained and evaluated on the BraTS data for comparing their performance. The best performing model will be

used for the further extraction process of the non-enhancing segment.

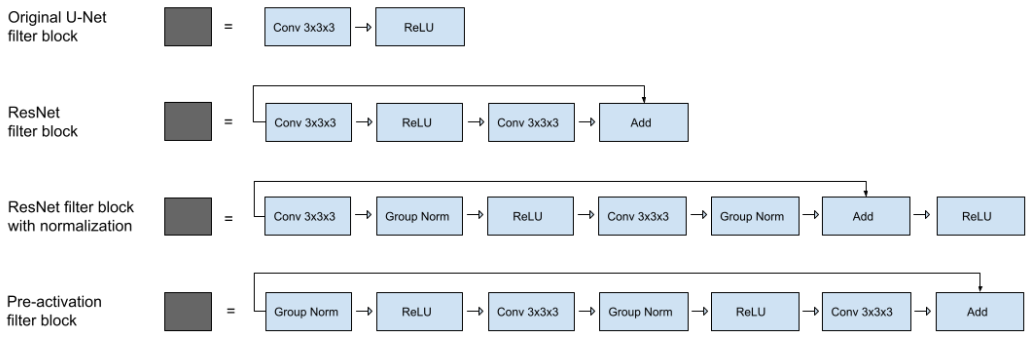

Figure 20: Different types of residual filter blocks, each type can be used in the above U-Net architecture

6.1.3 Training and Evaluation on BraTS 2018/21 Data

Each of the previously shown filter blocks results in a network with a different complexity, gradient behaviour and a varying number of parameters. To get a quantified insight into the prediction capabilities of the models, the four presented filter block types shown in Figure 20 will be used in the previously given basic U-Net model (Figure 19) and will be compared against each other by training and evaluating the derived models on the BraTS 2018 as well as on the BraTS 2021 data. The following Table 14 summarizes the presented filter blocks together with the resulting number of parameters of each model integrating them.

Model name	*Filter blocks*	*Parameters*	*Input dim.*
Plain-UNet	Original U-Net filter blocks	7.04M	(4, 80, 160, 128)
ResNet-UNet	ResNet filter blocks	12.88M	(4, 80, 160, 128)
ResNetN-UNet	ResNet filter blocks with norm.	12.88M	(4, 80, 160, 128)
PreAct-UNet	Pre-activation filter blocks	12.88M	(4, 80, 160, 128)

Table 14: U-Net variations for different filter block structures together with number of parameters.

The training of each of the models was run for 20 epochs on the BraTS 2018 and for 15 epochs on the BraTS 2021 data with a training/validation split of 90:10. The total number of records is 285 for the BraTS 2018 dataset and 1251 for BraTS 2021. The training results are shown in Table 15, Table 16 and Figure 21. A training speed comparison is presented in Table 17.

In the BraTS 2018 training data diagram of Figure 21, a rather clear ranking is observable, with the following model types having highest to lowest performance:

	BraTS 2018 train				**BraTS 2018 validation**			
	Plain	ResNet	ResNetN	PreAct	Plain	ResNet	ResNetN	PreAct
ET	0.6342	0.6767	0.6741	0.6745	0.6842	0.7352	0.7126	0.7239
TC	0.7708	0.7660	0.8038	0.7895	0.7349	0.7791	0.7859	0.7573
WT	0.8679	0.8712	0.8878	0.8833	0.8061	0.8294	0.8210	0.8266
Mean	**0.7576**	**0.7713**	**0.7885**	**0.7824**	**0.7417**	**0.7813**	**0.7732**	**0.7693**

Table 15: Accuracy comparison results for BraTS 2018 and filter block types *Plain*, *ResNet*, *ResNetN* and *PreAct*, 20 epochs

	BraTS 2021 train				**BraTS 2021 validation**			
	Plain	ResNet	ResNetN	PreAct	Plain	ResNet	ResNetN	PreAct
ET	0.8155	0.8282	0.8282	0.8245	0.8131	0.8215	0.8108	0.8138
TC	0.8655	0.8802	0.8748	0.8755	0.8725	0.8857	0.8939	0.8660
WT	0.8927	0.9006	0.9073	0.9040	0.8766	0.9166	0.9163	0.9014
Mean	**0.8579**	**0.8697**	**0.8701**	**0.8680**	**0.8541**	**0.8746**	**0.8737**	**0.8604**

Table 16: Accuracy comparison results for BraTS 2021 and filter block types *Plain*, *ResNet*, *ResNetN* and *PreAct*, 15 epochs

Model name	*Training speed [ms/step]*	*Training speed [s/epoch]*
Plain-UNet	356	89
ResNet-UNet	623	156
ResNetN-UNet	700	175
PreAct-UNet	700	175

Table 17: Training speed comparison for the given U-Net variations. Speed measured on a single Tesla V100-SXM2-16GB GPU.

ResNetN U-Net, pre-activation U-Net, ResNet U-Net, Plain U-Net. In the BraTS 2018 validation data diagram of Figure 21, the results seems similar with ResNetN and pre-activation performing very comparative, ResNet has lower performance throughout the training process and the Plain U-Net having lowest performance. Taking into account the similar results of ResNetN and PreAct on the validation data and PreAct having lower performance on the training data, this indicates a slightly stronger regularization effect for the pre-activation model. The ResNetN model has already higher training data performance but resulting in similar validation data results, which can indicate an earlier trend towards overfitting.

The BraTS 2021 curves in Figure 21, on the other hand, show very similar behaviour of ResNet, ResNetN and PreAct on the training data. The Plain U-Net performing clearly lower here. The results on the BraTS 2021 validation data however is very mixed with the Plain U-Net being below the others again. The pre-activation model performs good in the first half but is surpassed by ResNet and ResNetN in the end.

In total, the training diagrams in Figure 21 show, that there is a distinct gap between the lower performing *Plain U-Net* and the three remaining ones. But based on the given performance comparison results, there is not an obvious winner. Since in recent work on segmentation using deep learning methods [49, 50], the pre-activation filter block has shown considerable advantages, it will therefore be used in the following models of this work.

6.2 Proposed Upscaling Extension for the Pre-Activation U-Net

Figure 22 shows two versions of the proposed upscaling U-Net model, which will be used train on the BraTS 2018 and 2021 datasets for segmentation. The new model inherits from the base model given in Section 6.1 by the addition of a further upscaling branch on the decoder side, which exceeds the spatial input resolution by a factor of 2. This increases the output segmentation mask resolution and therefore benefits the detection of small scale structures, such as the non-enhancing tumor regions. Additionally there is a filtered skip path connecting going from the first of the original skip connections the new high resolution level. This is a novel design choice since U-Nets and their variations are found in a very symmetric structure throughout the present research literature. In some variations, e.g. in A. Myronenko's adapted U-Nets [87, 88] the decoder path comes with different filter block structures than the encoding path which is in contrast to the very original design by Ronneberger [97]. But the rigorous concept of the output shape exceeding the input shape in terms of spatial resolution by adding an additional stage in decoding, thus allowing a resolution upscaling behaviour, contributes a unique non-symmetric design idea to U-Nets.

The ground-truth masks have to adapted by upscaling them to double size. The upscaling method of the binary masks was by repeating the given binary values, no interpolation was used. The Sørensen-Dice coefficient metric as well as the Soft-Dice loss function were accordingly adapted to the increased size of the prediction

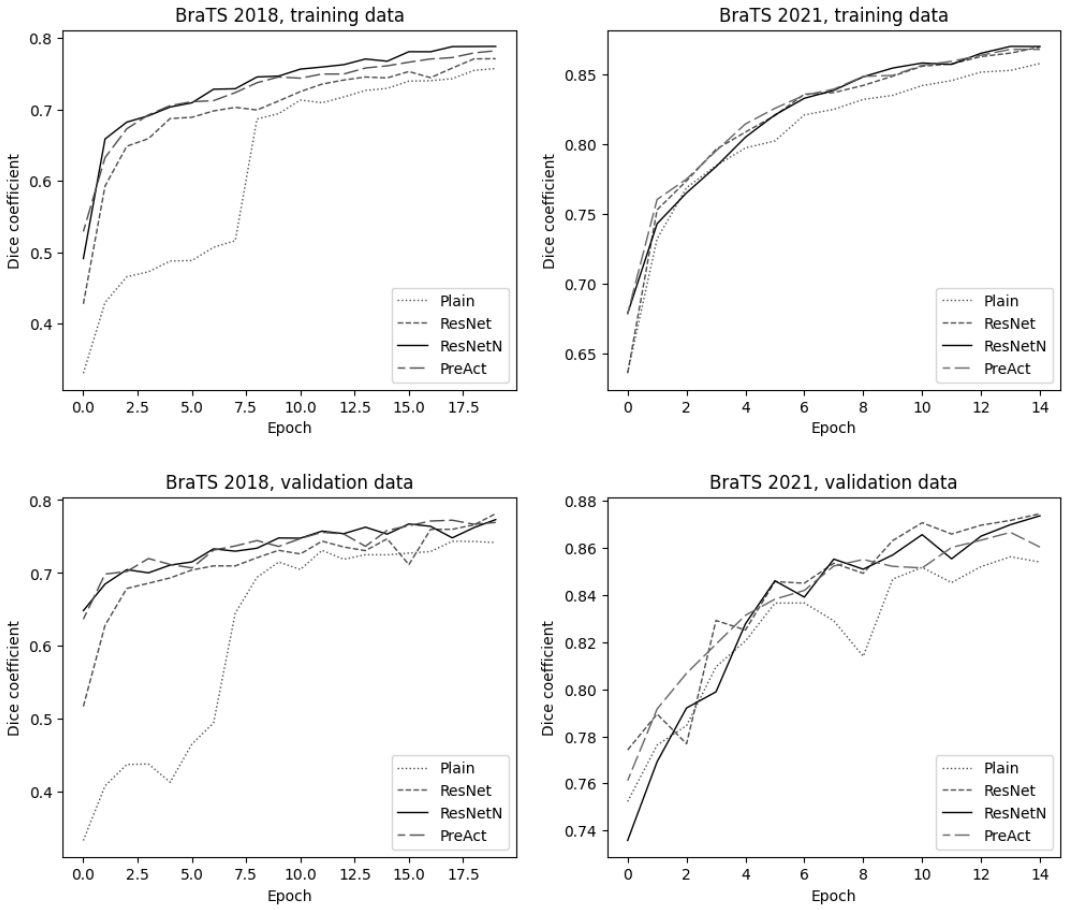

Figure 21: Training and validation results of given basic U-Net model using different types of filter blocks: Plain, ResNet, ResNetN (with normalization), Pre-Activation.

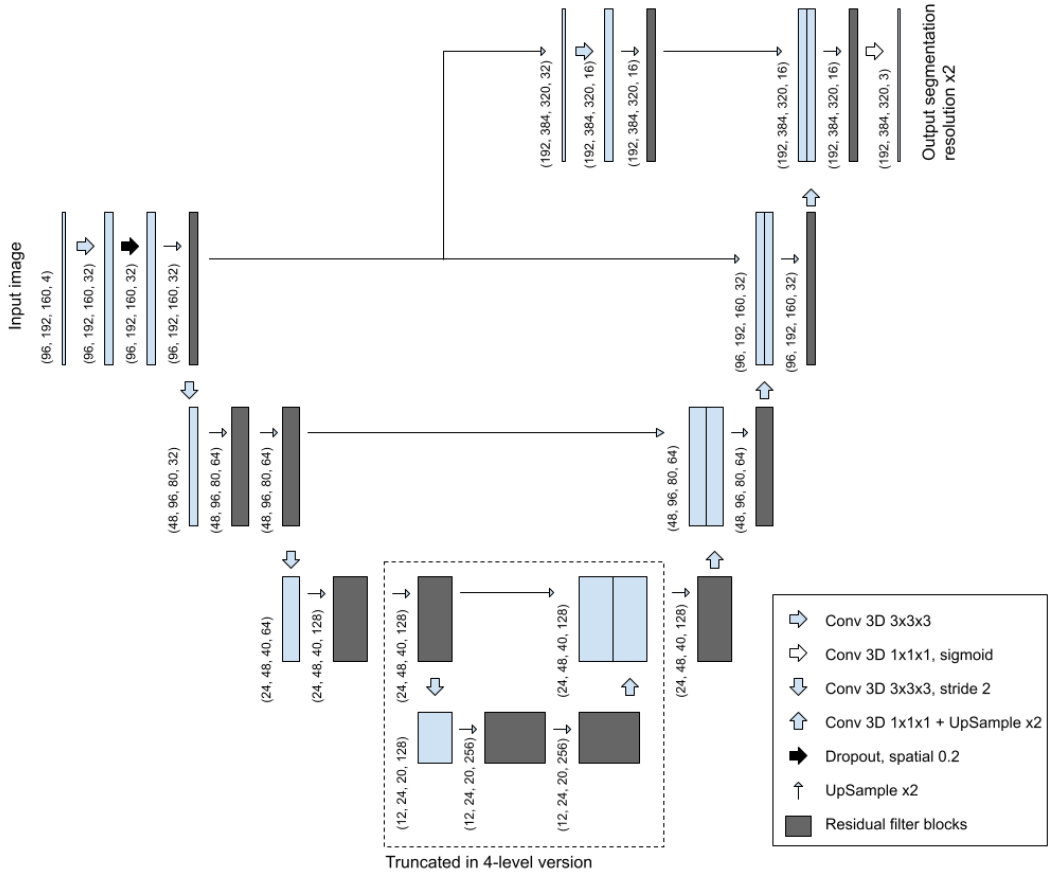

Figure 22: U-Net architecture with residual filter blocks and upscaling extension in the decoding path. The bottom part can be truncated for a 4-level version.

and ground truth maps.

The model was trained in two different architectural versions. The first one extending the previously presented pre-activation U-Net with the described resolution enhancing extension. This means the U-Net from the previous chapter, which is originally designed with four levels of spatial resolution will be extended by another fifth resolution level on the decoder side. In the second version, the bottom layer, carrying a very high number of parameters will be eliminated and the model is brought back to a 4-level design, the Tensorflow code for the 4-layer version is given in Appendix 9.4. This lower level design will support regularization and allow larger sized inputs within limited GPU memory constraints. The results of this model depth comparison regarding performance costs will be presented in the following.

Table 18 shows comprehensive results comparing the 4-level model variant against the 5-level model variant. The 4-level version achieves competitive results for the BraTS 2018 validation data and even higher results on the BraTS 2021 validation data. The validation performance difference of the 4-level and 5-level models on BraTS 2018 in classes TC and WT is very small, while the 5-level version is dominating on ET slightly. On the Brats 2021 validation data, the 4-level model's performance is clearly superior. In total, on the **validation data**, the 4-level version shows better performance than the 5-level variant.

Interestingly, the 5-level model reaches higher values when regarding the performance results for the BraTS 2018 and BraTS 2021 **training data**, which is not necessarily a benefit. On the BraTS 2018 training data, the 5-level model shows higher values on TC and WT and is only very minor behind the 4-level model in ET. On the BraTS 2021 training data, both models show very comparative results. In summary this shows that the 4-level model is outbeating the 5-level model on the validation data, while still reaching lower training data performance. This is a strong indicator for the 5-level model showing overfitting behaviour, while the 4-level model is still generalizing well.

The validation result plots for the training runs of the 4-level version of the model on the BraTS 2018 and the BraTS 2021 datasets is shown in Figure 23.

A reduced model architecture depth leads to a reduced number of parameters and model complexity which, as discussed in various deep learning literature [41, 13] generally makes the it less sensitive to noise and enforces the regularization effects and the stability of the model. This approach will be used here with respect to the improved model behaviour in it's 4-level form. The reduced architecture depth further allows larger input size without exceeding the GPU memory of the available hardware. Table 19 shows the number of parameters of both model versions, as well es the training speed and a maximum possible input data size for the given computation hardware.

	BraTS 2018 train		**BraTS 2018 val.**		**BraTS 2021 train**		**BraTS 2021 val.**	
	5-level	4-level	5-level	4-level	5-level	4-level	5-level	4-level
ET	0.7162	**0.7179**	**0.7842**	0.7802	0.8524	**0.8566**	0.8357	**0.8464**
TC	**0.8803**	0.8646	0.8125	**0.8230**	**0.9140**	0.9121	0.8798	**0.8841**
WT	**0.9117**	0.9010	**0.8913**	0.8912	**0.9282**	0.9232	0.9157	**0.9186**
Mean	**0.8361**	0.8278	0.8293	**0.8315**	**0.8982**	0.8973	0.8771	**0.8830**

Table 18: Upscaling pre-activation U-Net results, dice coefficients for enhancing tumor (ET), tumor core (TC) and whole tumor (WT). Comparison for 4-level and 5-level model. 60 epochs training on BraTS 2018 data. 30 epochs training on BraTS 2021 data. Bold numbers indicating the higher numbers in 5-level/4-level comparison.

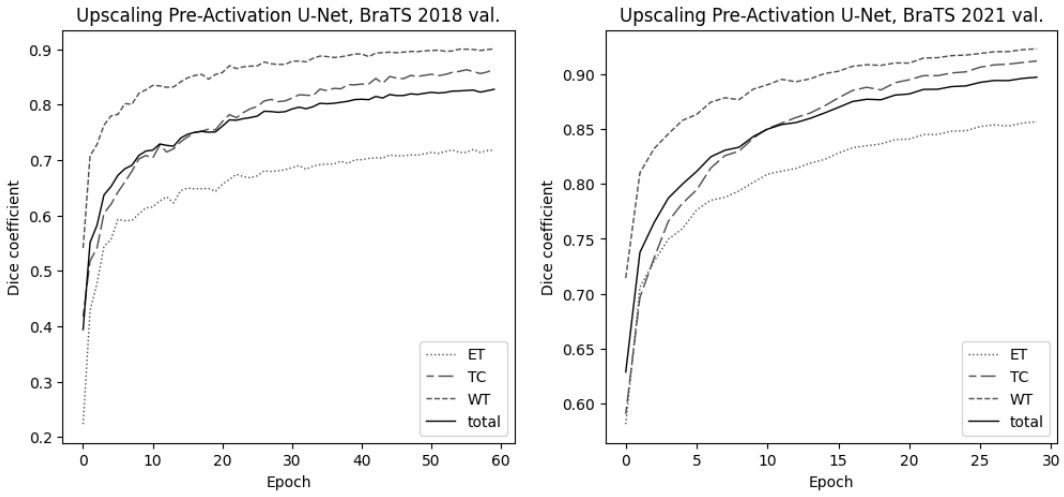

Figure 23: Upscaling Pre-Activation U-Net 4-level architecture, training results for BraTS 2018 and BraTS 2021 validation data

Model	*Parameters*	*Input size*	*Training speed*
Upscaling PreAct-UNet 5-level	13.15M	80 x 160 x 128 (1.64M)	696 ms/step
Upscaling PreAct-UNet 4-level	2.87M	96 x 192 x 160 (2.95M)	1096 ms/step

Table 19: Model comparison regarding number of parameters, input size and training speed (measured on a Tesla A100-SXM4-40GB GPU)

The two center images of Figure 24 show an example of the original resolution segmentation masks given in the BraTS data. Due to the repeating pixel upscaling method for carrying the ground truth data to the new resolution level, this structure is basically inherited to binary masks which are used for training. The right column of Figure 24 shows a segmentation output of the already trained model for the same record. Remarkably, the prediction mask which is inferred by the upscaling model does not reflect any of the lower resolution edges or steps, but fully constructs a binary mask with resolution adequate details. It may be expected, that with a high model complexity and extensive training runs, the edged shape could be approximated to minimize loss. The given model does not show any of this behaviour and can thus be supposed to be generalizing very well.

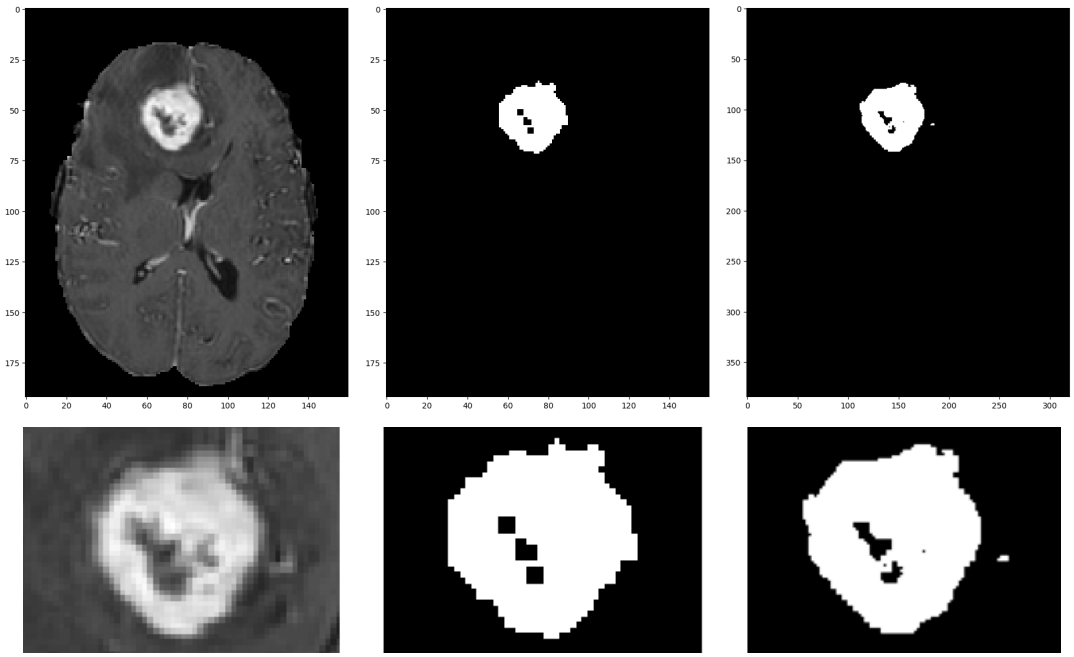

Figure 24: Left: MRI T1ce sample record from BraTS 2018 dataset. Center: BraTS 2018 ground-truth segmentation mask for the Enhancing Tumor (ET) region. Right: Prediction using the presented resolution enhancing model

Figures 25 and 26 also show predictions done by this presented upscaling U-Net model. They were done after training the presented model on a combined BraTS 2018/21 dataset in section 7.4 but are already shown here to demonstrate the resolution enhancing capabilities of the presented architecture. The shapes of the generated segmentation masks (right) are forming a detailed and natural structure of the brain tumor tissue, despite exceeding the underlying resolution of the MRI record record, which come in the same pixel resolution as the original segmentation masks (left). Additionally the newly generated segmentations show the introduced prediction of the non-enhancing region NET (green). Similar slices illustrating the resolution enhancing aspect and showing NET predictions are given in Appendix 9.5.

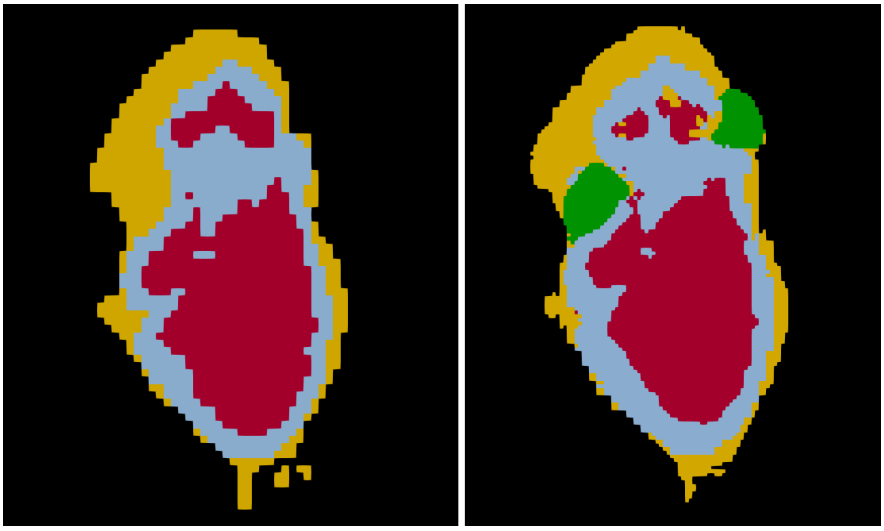

Figure 25: Sample using resolution enhancing model including NET prediction, BraTS 2021 record 00203. Left: BraTS 2021 ground truth segmentation, original resolution, edema (yellow), enhancing tumor (blue) and necrosis (red). Right: Prediction using the presented resolution enhancing model, includes non-enhancing tumor NET prediction (green).

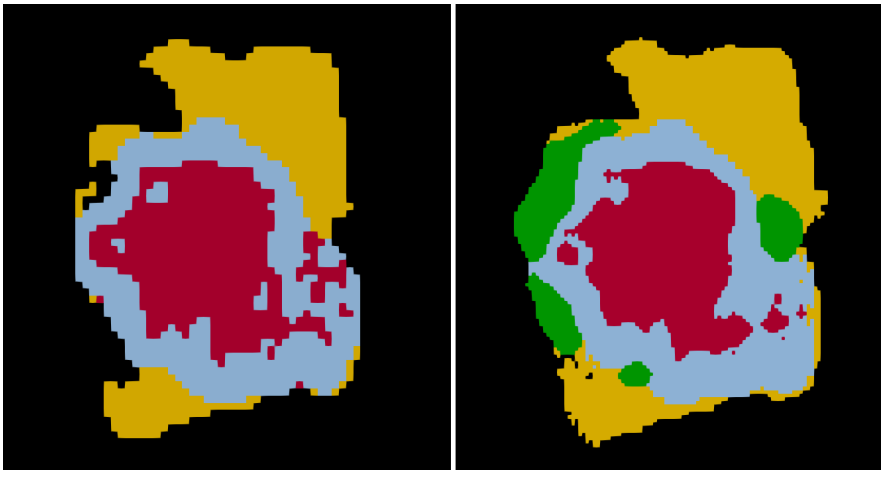

Figure 26: Sample using resolution enhancing model including NET prediction, BraTS 2021 record 00209. Left: BraTS 2021 ground truth segmentation. Right: Prediction using the presented resolution enhancing model with NET.

6.3 Reconstruction of a Subset of Isolated NET from BraTS 2018

As described in Section 5, the BraTS datasets of the recent years did not come with dedicated segmentation masks for the non-enhancing *NET* region [8, 6]. More specifically, 2015 was the last year where the published BraTS data provided specific NET masks [83], yet not throughout consistent with recent definitions. After that, in the datasets of the years 2018 and 2021 the labels were always combined to higher level hierarchical labels *ET* (enhancing tumor), *TC* (tumor core) and *WT* (whole tumor).

Since there was a variation in these combinations of labels in the datasets between the years 2018 and 2021, and using the previously trained BraTS 2021 model, it will be possible to break up some fused *NCR* masks from BraTS 2018 containing *NET* region information. Together with the application of morphological filters this will provide a subset of records with isolated NET masks which will serve for further model training.

Figure 27 shows a sample slice from the BraTS 2021 dataset, the first segmentation mask shows the data provided by BraTS 2021. The second segmentation is extended by a *NET* label, which will be predicted by the presented method. More sample slices with results for the extraction of NET are shown in Appendix 9.6. The relevant tumor regions are listed in Table 20.

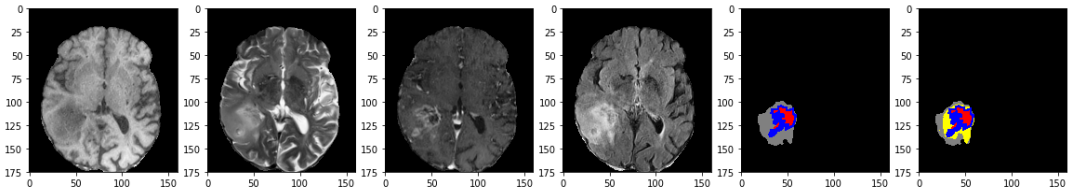

Figure 27: Sample record from BraTS 2021, from left to right: T1, T2, T1ce, Flair, BraTS 2021 Segmentation data, extended Segmentation data with Non-Enhancing region (yellow)

Color	*Region*	*Short*	*Description*
gray	Edema	*ED*	existing, from Brats 2021
red	Necrosis	*NCR*	existing, from Brats 2021
blue	Enhancing	*ET*	existing, from Brats 2021
yellow	Non-Enhancing	*NET*	new, predicted using model trained on NET isolated BraTS 2018 subset

Table 20: Original segmentation labels in BraTS 2021 with new Non-Enhancing label added.

The BraTS 2018 challenge paper [8] states that the necrosis segment *NCR* is labeled together with the non-enhancing segment *NET*. Therefore the *NET* segment is implicitly included in the BraTS 2018 *NCR* labels. The reason why these regions

were merged to one label are not explain in the official announcements. Nevertheless, necrosis and non-enhancing parts are typically found in spatially well separable regions. Necrosis is typically found in center tumor parts, surrounded by the enhancing tumor region and forming the tumor core *TC* together with it. Outside if this *TC* region, in it's bordering area and reaching into the surrounding edema *ED* domain the non-enhancing *NET* part can be found. In many cases necrosis *NCR* and non-enhancing tumor areas *NET* have only few or no touching or intersecting points.

By a process of predicting, filtering and substracting the plain necrosis regions using the trained BraTS 2021 model, it is widely possible to separate the two regions. The basic prediction and classification steps performed are the following:

- Train model on BraTS 2021 data (model will predict ET_{21}^{pred}, TC_{21}^{pred}, WT_{21}^{pred} as stated in datasets section) with $TC_{21}^{pred} = NCR_{21}^{pred} \cup ET_{21}^{pred}$.
- Calculate NCR_{21}^{pred} region using the previously trained BraTS 2021 model by using the equation $NCR_{21}^{pred} = TC_{21}^{pred} \setminus ET_{21}^{pred}$.
- For a given BraTS 2018 record, read the provided NCR_{18}^{gt}, where $NCR_{18}^{gt} = NCR_{21}^{gt} \cup NET$ (note that by BraTS data definition $NCR_{18}^{gt} \neq NCR_{21}^{gt}$). Calculate NCR_{21}^{pred} using the trained BraTS 2021 model prediction for TC_{21}^{pred} as described above.
- The non-enhancing region can then be extracted as $NET = NCR_{18}^{gt} \setminus NCR_{21}^{pred}$ or even as $NET = NCR_{18}^{gt} \setminus TC_{21}^{pred}$ given that the non-enhancing tumor part does not reach into the enhancing tumor region, which should hold, since characterization and ground truth data of the enhancing region is supposed to be very profound.
- Reassign pixel labels to NET if pixel is classified as NCR_{18}^{gt} but not TC_{21}^{pred}. Also edema ED, enhancing tumor ET and background pixels will be changed to NET if classified as such.

For improving the results of the above process, morphological filters are applied to clean the masks and to reduce noise effects. A detailed list of the operation steps and the applied filters is shown in Table 21.

The main advantages of using morphological filters in this application are the elimination of small (single pixel sized) border noise, which is present in some of the given BraTS 2018 ground truth masks, the reduction of artifacts resulting from substraction of masks and the possibility of filling small gaps in border areas. The small pixel sized noise appearing in some of the BraTS 2018 NCR data (see Figure 28 for examples) may be a result of labeling suggestions which were automatically generated and confirmed in the original labeling process of the BraTS dataset.

Set	*Operation*	*Iterations*	*Structuring el. B*
TC_{21}^{pred}	$= TC_{21}^{pred} \ominus B$	1	rank 3, connectivity 1
NET	$= TC_{18}^{gt} \setminus (ET_{18}^{gt} \cup TC_{21}^{pred})$	-	-
NCR	$= TC_{18}^{gt} \setminus (ET_{18}^{gt} \cup NET)$	-	-
NET	$= NET \circ B$	1	rank 3, connectivity 1
NET	$= NET \oplus B$	1	rank 3, connectivity 2
NET	$= NET \setminus ET_{18}^{gt}$	-	-
NCR	$= NCR \circ B$	1	rank 3, connectivity 1
ET	$= TC_{18}^{gt} \setminus (NET \cup NCR)$	-	-
ED	$= WT_{18}^{gt} \setminus (TC_{18}^{gt} \cup NET \cup NCR)$	-	-

Table 21: Morphological filters applied for improving NET extraction results in BraTS 2018 data. See section 4.5 for a detailed description of the binary morphological filters $\ominus$ (erosion), $\oplus$ (dilation) and $\circ$ (opening). The structuring element B for the binary filters is given in the last column.

The sample images in Figure 28 show slices of BraTS 2018 records which have been processed by the described operations.

Having the isolated *NET* masks resulting from this process, the BraTS 2018 dataset was analysed towards presence and extent of contained *NET* data in each of the records. Therefore the total volume of *NET* mask was measured in each record and the data was classified to three groups using Gaussian Mixture Model Clustering (see Figure 29).

The first group (cluster 1 in Figure 29) contains no or low amount of NET pixels, which may have it's reason in the *NET* segment presumably not being labeled in this set of BraTS 2018 records. The second group (cluster 2) having a very high amount of *NET*, possibly reflecting an alternative definition of *NET* or some labeling error. And the third and intermediate group (cluster 3) having a moderate amount of *NET* pixels, this group will be used as training data.

The results from dividing the data using this method are shown in Table 22. The separation values between clusters, in the context of Gaussian Mixture Models (GMM), can be interpreted as the values where the probability density functions (PDFs) of the Gaussian components intersect with each other. These values serve as natural boundaries between the clusters. To determine the separation values, the PDFs of the Gaussian components are calculated and solved for the points of intersection. Using the selected subset of medium sized *NET* region data, the previously presented upscaling pre-activation U-Net was trained to classify the four segments ET, ED, plain NCR and the isolated NET.

Due to having a reduced number of training records together with some degree of inconsistency in the *NET* data, a high-fold cross validation approach would be needed

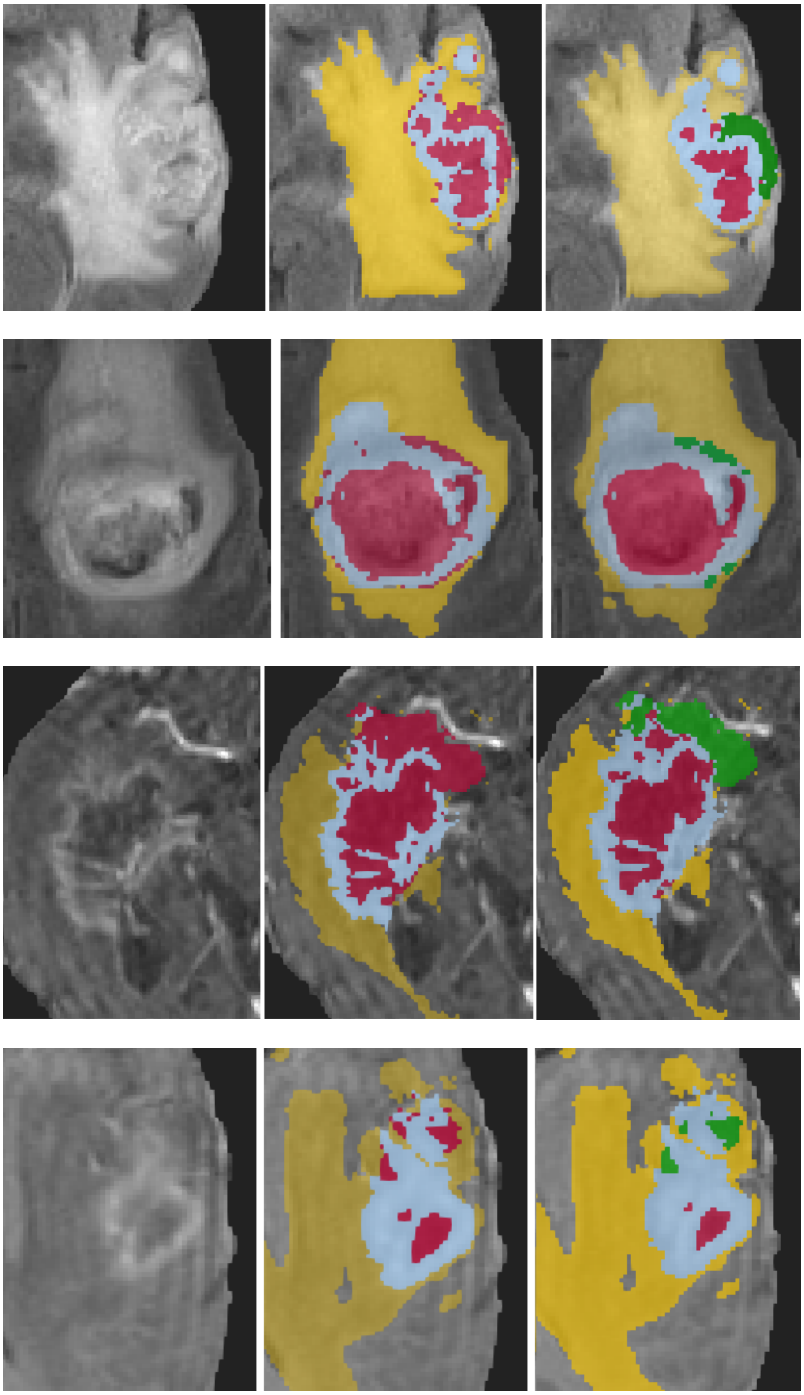

Figure 28: BraTS 2018 records CBICA-ANI-1, CBICA-AXO-1, CBICA-AQQ-1, TCIA03-419 (top to bottom). Left: Flair (CBICA-ANI-1, CBICA-AXO-1) or T1ce (CBICA-AQQ-1, TCIA03-419) slice, center: original BraTS 2018 segmentation mask (ED yellow, NCR+NET red, ET blue), right: NCR (red) and NET (green) separated and filtered.

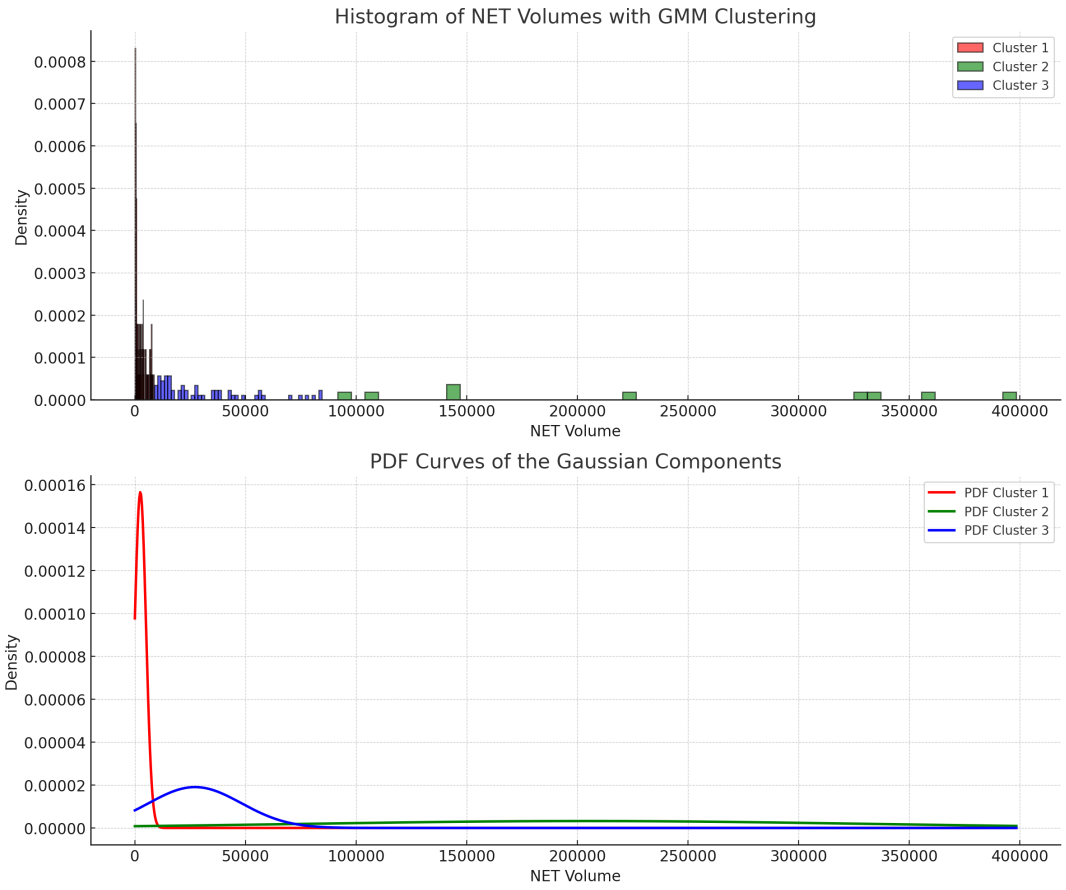

Figure 29: The top plot shows the histogram of NET volumes with the GMM clusters. The bottom plot presents the PDF curves of the Gaussian components inferred by the GMM.

Cluster	Selected boundaries	GMM PDF intersections	Min. value	Max. value	Size
1	0...10000	14823	0	9779	132
2	>100000	115560	107737	336035	10
3	10000...100000	14823; 115560	10525	84516	68

Table 22: Selected group boundaries and cluster boundary suggestions by Gaussian Mixture Models (GMM) using PDF intersection

to capture a representative validation accuracy, which was not performed due to high computational costs. Instead, the training was performed on the full available subset to a level of convergence without using validation data. The results are shown in Table 23 and Figure 30.

	BraTS 2018 subset
NCR	0.8197
ED	0.8053
NET	**0.4943**
ET	0.8122
Mean	0.7329

Table 23: Training results for presented pre-activation U-Net on BraTS 2018 subset with reconstructed and isolated NET segments. 100 epochs, 1044ms/step, learning rate 1e-4 with 5% decay per epoch, training data size 68, batch size 1, GPU Tesla T4

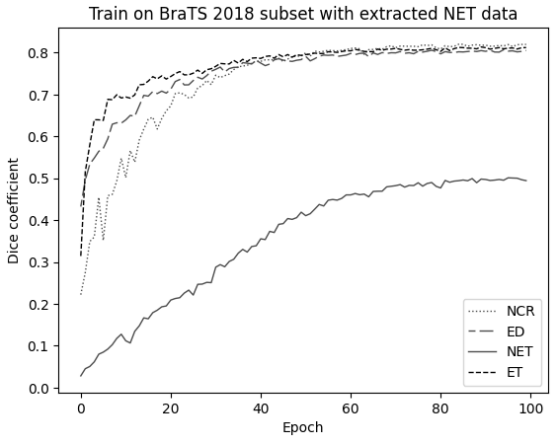

Figure 30: Training results for the presented pre-activation U-Net on BraTS 2018 subset with reconstructed and isolated NET segments.

Having a working model for *NET*, based on the computed subset of isolated *NET* records, the full BraTS 2018 data was again processed using it. This means, the records of group 1, which had few or no *NET* data were assigned new and additional *NET* predictions and the third group of data, which originally had a high amount of *NET* were reduced to smaller *NET* areas. Also the second group of records, which were used as training data, were re-processed to reach a higher level of consistency through the data. The volume distributions of the newly added NET segment and further analysis will be given in the next Section 7.

7 Results and Discussion

7.1 Discussion on Selected Samples

Identifying non-enhancing tumor parts on T1, T2, T1ce and Flair MRI records is a challenging and often ambiguous task. To judge the plausibility of the given model's predictions, some axial slices from the BraTS 2021 data records together with the original and the NET extended segmentation masks will be discussed in this section.

Figure 31 shows BraTS 2021 record 00062. The images show an extended area of edema which is very clearly indicated as a hyperintense domain in the T2 and Flair records. On the T1ce contrast-enhanced image, the typical tumor core, consisting of the very bright enhancing part together with the surrounded darker necrotic parts is visible. The dark inside parts in T1ce still are not labeled as necrosis, since in this example it was the ending or border part of necrosis. In the neighbouring of the axial cuts, necrosis was started to be labeled in the BraTS dataset. This covers the given original segmentations masks of enhancing tumor (blue) and edema (yellow). By observing the top left area next to the enhancing tumor, a slightly shaded area can be noted which is marked by an arrow on the T2, T1ce and Flair modalities. On the T2 record, this area extends the part of enhanced tumor and can be identified as less hyperintense as the homogeneously brighter edema showing. This is an indication of non-enhancing tumor tissue, which is predicted in green color in the bottom left segmentation, showing the prediction of the trained model. A similar phenomenon shows in the Flair modality, where the enhancing tumor domain has an adjacent, less hyperintense part on the top left side. The T1ce slice only shows a slight hyperintense effect in the border region between enhancing and non-enhancing part. NET typically shows very variable on T1 and T1ce modalities. All these observations match with the imaging characteristics of non-enhancing tumor parts summarized in Table 2 of Section 3.4.

In another example, BraTS 2021 00505 in Figure 32, the shape of the tumor without surrounding edema can well be determined in the T1 modality where the core part of the abnormal tumor tissue appears elliptically shaped. The same can be observed in the Flair image and, with a hypointense top and bottom region, also in the T2 slice. The neighbouring edema region lies at the bottom left and top left and is shown as a bright and rather homogeneous shade in the Flair image. The enhancing tumor tissue as well as the necrotic part are very clearly visible in the T1ce mode, namely as the very hyperintense and the very hypointense parts. The enhancing tumor part ET together with the necrosis part NCR form the classical tumor core part TC. This example shows clearly, that there is a residual part between this tumor core part TC showing in T1ce and the elliptical tumor region apparent in T1, T2 and Flair. This remainder area is identified as the non-enhancing tumor part, again being consistent with the intensity listing of Table 2 and shown as the green prediction mask in the extended segmentation on the bottom left image. In Appendix 9.7 more examples of this type are presented.

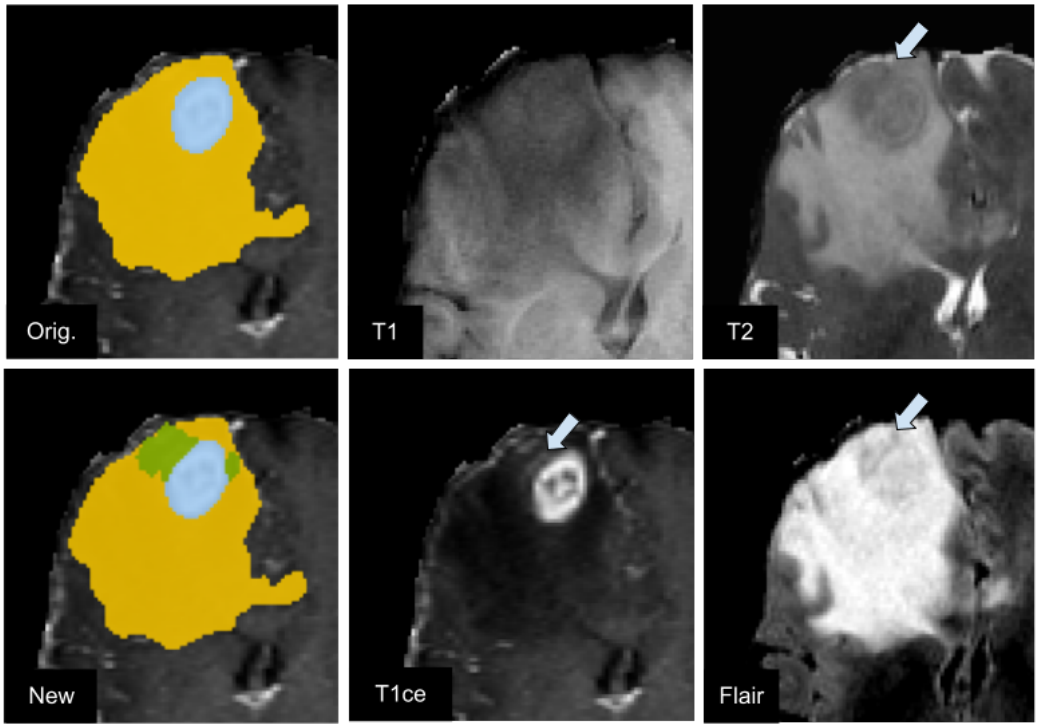

Figure 31: BraTS 2021, record 00062, axial slice, the left images show the original segmentation of enhancing tumor (blue) and edema (yellow) as well as the segmentation mask extended by the prediction for non-enhancing NET (green). The center and right images show the T1, T2, T1ce and Flair sequence images. The arrows are pointing onto shaded areas indicating NET.

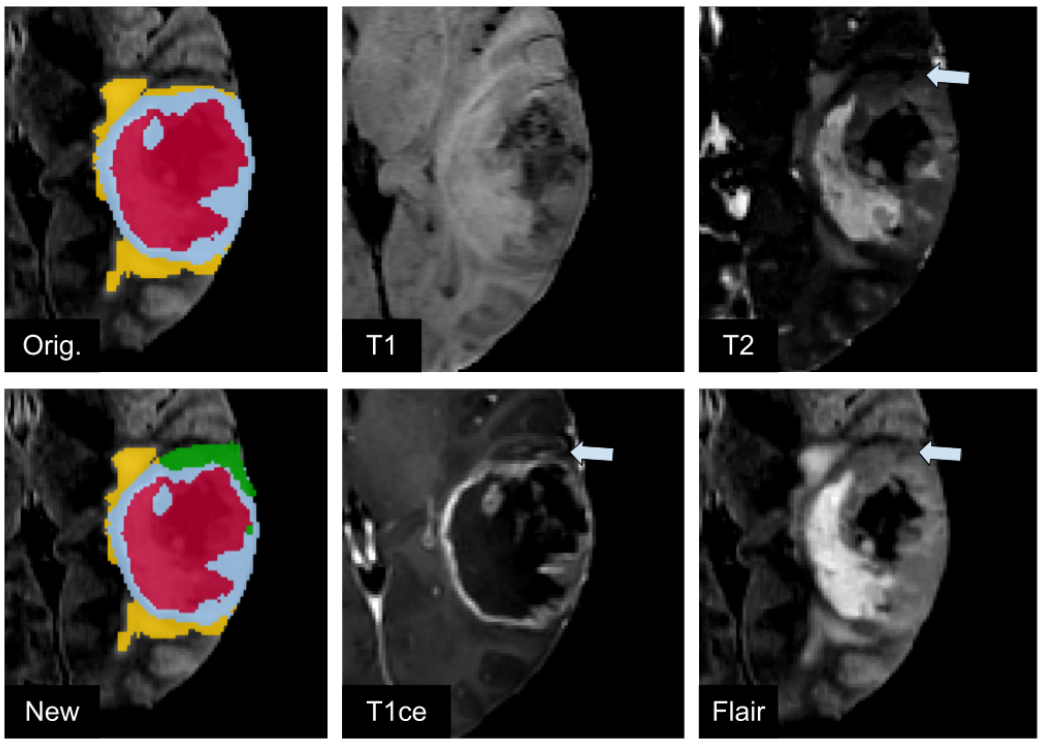

Figure 32: BraTS 2021, record 00505, axial slice, the left images show the original segmentation of enhancing tumor (blue), edema (yellow) and necrosis (red) as well as the segmentation mask extended by the prediction for non-enhancing NET (green). The center and right images show the T1, T2, T1ce and Flair sequence images. The arrows are pointing onto shaded areas indicating NET.

Figure 33 shows a summary of examples, where each of the sample records (BraTS 2021 00209, 00724, 01017, 01028, 01584) is given as a T1ce and a Flair slice (left) together with the original BraTS2021 segmentation mask and the extended 4-label segmentation. On the T1ce image, the actual tumor core (NCR + ET) can be identified, which is shown as blue (ET) and red (NCR) masks in the segmentations. The interpretation of non-enhancing parts is again very subtle in these examples. In the Flair image, the tumor center part comes with a bright surrounding edema area. If the edema is ignored, the tumor center part in the Flair image appears larger than the tumor core area on T1ce. This difference in extent of center areas, which is typically less hyperintense than the edema area on the Flair record, marks the non-enhancing part.

7.2 NET Volume Distributions

Another approach towards analyzing the segmentation outcome is by observing the total amount of classified non-enhancing volume and compare it towards the volumes of the remaining regions. This will explore the degree of covariance between the different tumor regions, since larger tumors will typically include a higher volume of

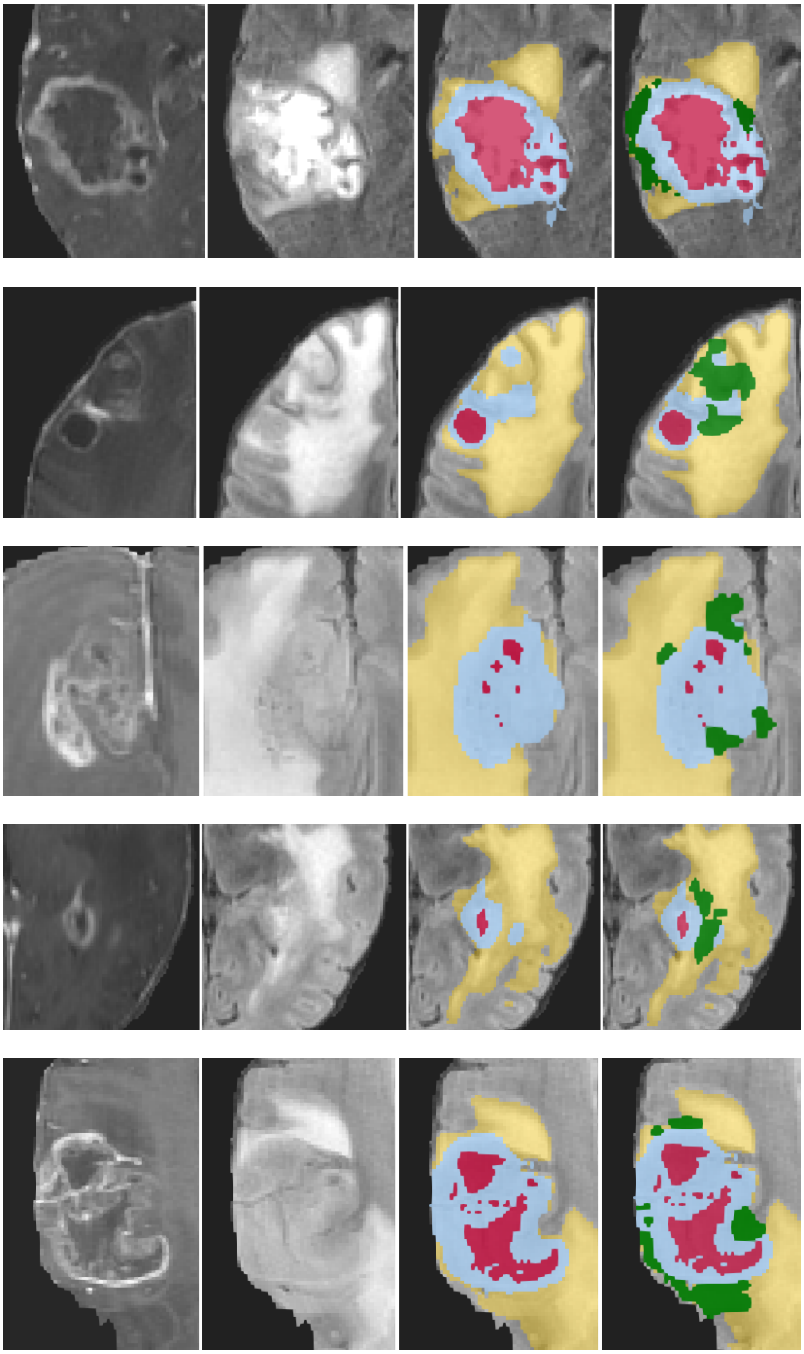

Figure 33: BraTS 2021 records 00209, 00724, 01017, 01028, 01584 (top to bottom). Left: T1ce slice, center left: Flair slice, center right: original BraTS 2021 segmentation mask with 3 labels (ED yellow, NCR red, ET blue), right: extended 4-label mask with newly added NET (green)

each of the segments, smaller tumors will enclose a smaller number of pixels for all of the classes. Following this assumption, it can be expected that for each of the segments *NCR*, *ED*, *NET* and *ET* volume distributions of similar type can be found when observing the entire dataset.

For this analysis, the number of pixels of each of the segments *NCR*, *ED*, *NET* and *ET* was counted throughout the newly extended BraTS 2021 4-label data. The cropped and upscaled record size in the 4-label dataset is 192 x 384 x 320. Each of the labels defines a subset of pixels, whose volume is counted and listed as indicated below:

```
ID,                 NCR,    ED,     NET,    ET
BraTS2021_00549,    3003,   87897,  264,    7284
BraTS2021_00565,    8102,   50585,  3057,   21808
BraTS2021_00378,    935,    30654,  2529,   3099
BraTS2021_01505,    63795,  25323,  2018,   1066
BraTS2021_01183,    19553,  78470,  4027,   21283
...
```

This gives a list containing all 1251 sample records in the BraTS 2021 dataset, which can be analyzed and visualized in various ways. Figure 34 shows the histograms of each of the regions volumes.

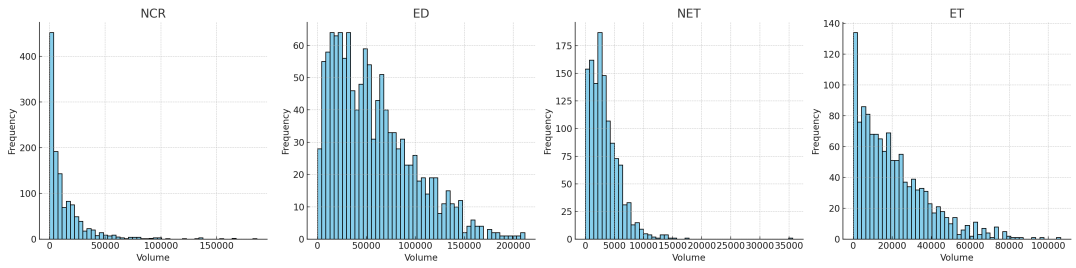

Figure 34: Histograms of NCR, ED, NET and ET volumes in the BraTS 2021s dataset. NET was newly added, the segmentation masks for NCR, ED, and ET are the original ones.

The plots of all four classes show positively skewed distributions of the tumor region volumes, that range from values close of zero to very large values, which is a common characteristic in biological data. After removing the very high and outlying values from NCR and NET, these volumes can closely be modeled by the gamma distribution, which can be used in medical and clinical data when describing disease progression, survival times or especially in oncology, the tumor growth distributions across patients as stated in [28]. Figure 35 shows the volume distributions with fitted gamma distribution curves. The newly generated NET volumes show a similar distribution behaviour across the patient data in the BraTS 2021 dataset than the remaining classes.

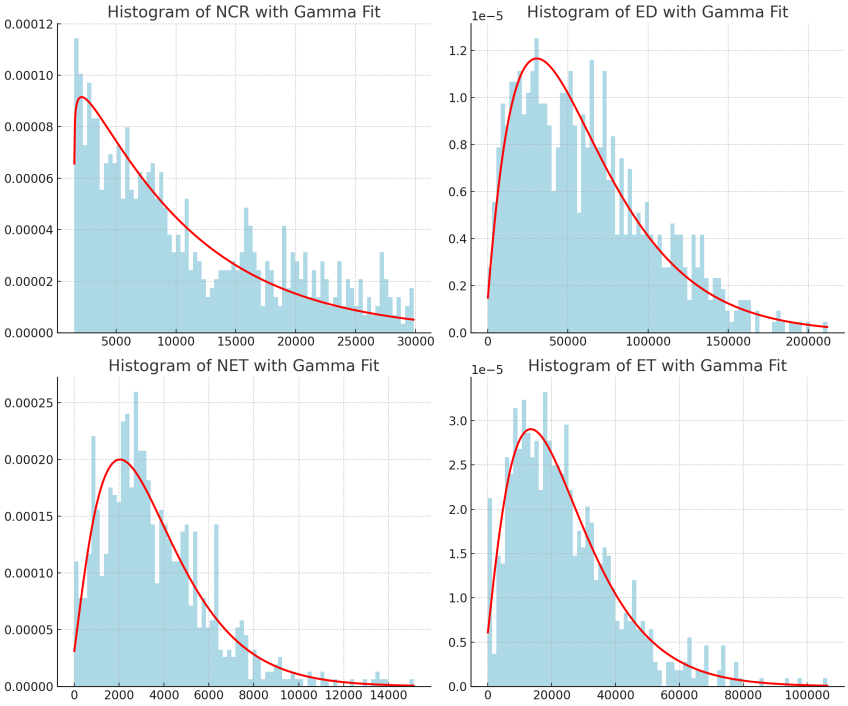

Figure 35: Gamma distribution fits for volume distributions of NCR, ED, NET and ET. Blue histograms showing the actual volume distributions, red graphs showing an appropriate gamma distribution fit.

Another common representation is a pairplot, which is shown in Figure 36. It shows the approximated density functions on the diagonal, which reflect the above given histograms of the single volume distributions for each of the labels. In the off-diagonal places, it shows pairwise scatterplots of the different segments, visualizing the correlations between their volumes.

	NCR	**ED**	**NET**	**ET**
NCR	1.000	0.105	0.203	0.201
ED	0.105	1.000	0.308	0.272
NET	0.203	0.308	1.000	0.338
ET	0.201	0.272	0.338	1.000

Table 24: Pearson correlation coefficients between NCR, ED, NET and ET volumes. Correlation coefficients were calculated based on extended BraTS 2021 4-label dataset.

The values in the correlation matrix shown in Table 24 represent the Pearson correlation coefficients between the volumes of NCR, ED, NET and ET segments. A value close to 1 implies a strong positive correlation, a value close to -1 would imply a strong negative correlation, and a value close to 0 implies little to no correlation. The pairwise values are listed in Table 24. The following list describes the Pearson correlation coefficients in increasing order:

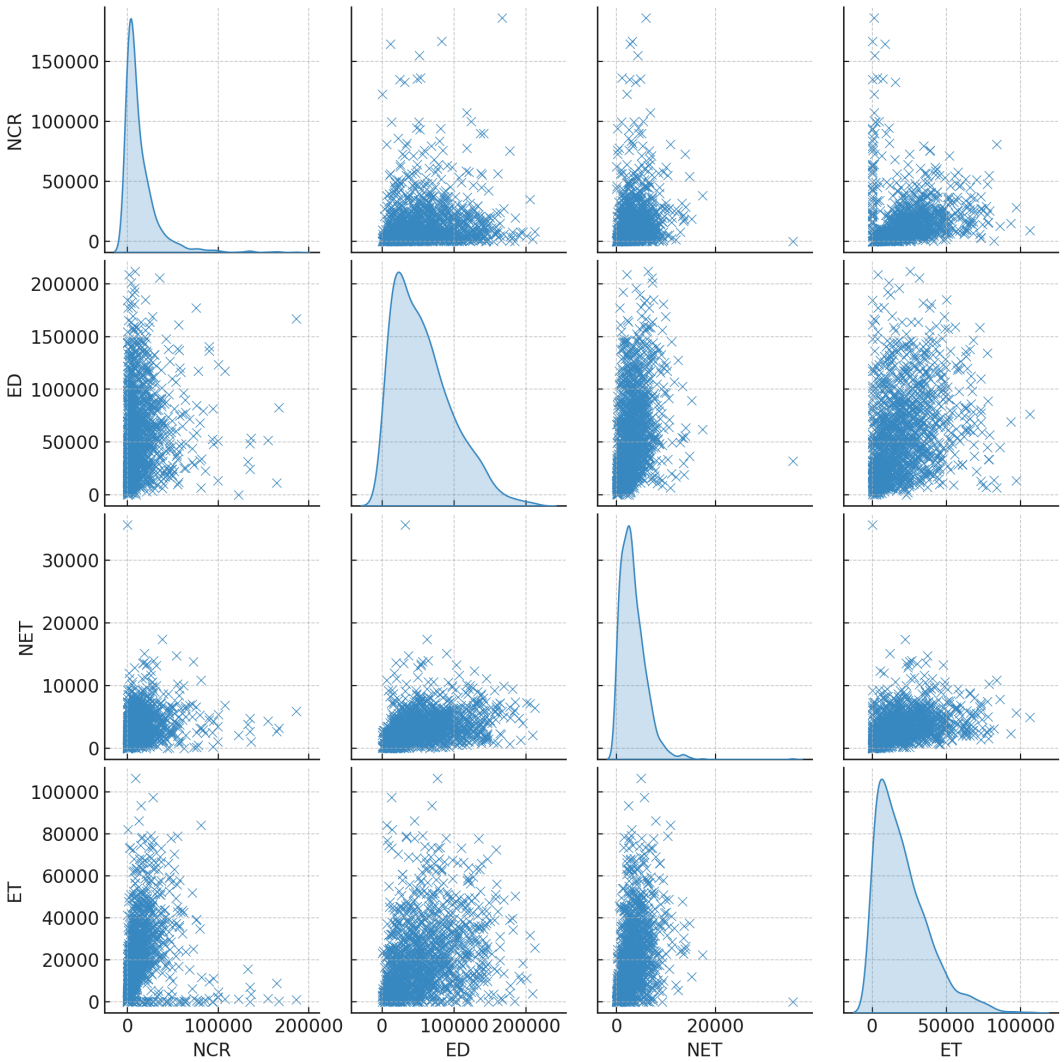

Figure 36: Pairplot of NCR, ED, NET and ET volumes. The diagonal shows the kernel density estimate (KDE) for each variable, which provides a smoothed representation of the data's distribution. The off-diagonal plots are scatter plots between pairs of the volumes.

- **NCR and ED (0.105)**: There is a slight positive correlation between NCR and ED. This suggests that as NCR values increase, ED values tend to increase slightly as well, but the relationship is weak.

- **NCR and ET (0.201)**: The correlation between NCR and ET is similar to that between NCR and NET, indicating a weak positive relationship.

- **NCR and NET (0.203)**: There is a weak positive correlation between NCR and NET. This suggests that as NCR values increase, NET values tend to increase to some extent.

- **ED and ET (0.272)**: There's a weak to moderate positive correlation between ED and ET. As ED values increase, ET values also tend to increase to some extent.

- **ED and NET (0.308)**: There's a moderate positive correlation between ED and NET. This suggests that as ED values increase, NET values also tend to increase moderately.

- **NET and ET (0.338)**: The correlation between NET and ET is moderate. As NET values increase, ET values tend to increase moderately.

Overall, there are no strong correlations between any pairs of the segments. The relationships that exist are mostly weak to moderate, but with all correlations being positive. That means they tend to move in the same direction to varying degrees. From the scatter plots in Figure 36, we can get a visual sense of these correlations. For instance, we see that *NET* and *ET* have a cloud of points that suggests a positive correlation, aligning with the correlation value which was computed.

Also, it can be noticed, that the Necrosis *NCR* segment volume seems more independent than the others. The three *NCR* volumes' correlation values *NCR to ED* (0.105), *NCR to ET* (0.201) and *NCR to NET* (0.203) are the lowest ones appearing, hinting towards a weaker medical or clinical coupling of necrosis volume and total tumor volume or intrinsic volume relations. The non-enhancing *NET* segment on the other hand is present in the two pairs of highest correlation, *ED to NET* (0.308) as well as *NET to ET* (0.338). This indicates a commonly increased presence of the three regions NET, ET to ED with increasing total tumor size.

7.3 Analysis of Variance for Intensity Masks

For evaluating the newly added *NET* segment and the overall BraTS 2021 4-label dataset towards intensity characteristics of the single regions *NCR*, *ED*, *NET* and *ET* on the different MRI modalities, the data is processed and prepared as follows:

- For a given modality (T1, T2, T1ce, Flair) the 3D MRI image data (x_i) is loaded and re-scaled to (z_i) in range [0, 1] by

$$z_i = \frac{x_i - \min_j(x_j)}{\max_j(x_j) - \min_j(x_j)}$$

- For one of the tumor regions *NCR*, *ED*, *NET* and *ET* the segmentation mask (m_i) with $m_i \in \{0,1\}$ is applied to the re-scaled image data by multiplication with the binary mask value $w_i = m_i \cdot z_i$
- The mean intensity for the given tumor region in the selected MRI modality is then calculated as $I = \sum_i w_i / \sum_i m_i$
- This is performed for all tumor regions *NCR*, *ED*, *NET* and *ET* in all MRI modalities T1, T2, T1ce and Flair

Figure 37 shows boxplots for the region intensities of all the combinations of *NCR*, *ED*, *NET* and *ET* on the different MRI modalities T1, T2, T1ce and Flair. Each boxplot shows the median inside the upper and lower quartile bounded box. The whiskers are set to the highest and lowest observations within the typical distance of 1.5 times interquartile range outside the upper and lower quartile.

Performing an analysis of variance (ANOVA) of the above region intensities gives the below results, where the extremely low p-value suggests that there is a statistically significant difference in the mean intensities across the four categories *NCR*, *ED*, *NET* and *ET*:

- F-statistic: 44.84
- p-value: 1.39×10^{-28}

Since the ANOVA Analysis only states, that there are at least two categories with different means, a post hoc group comparison test has to be performed. This is done by the Tukey Honestly Significant Difference (Tukey HSD) test [115], which compares all possible pairs of means and controls for the family-wise error rate, see Tables 25 to 28.

The non-enhancing region's intensity means are showing significant differences to most of the other regions in the different modalities:

- In T1: NET mean is significantly different to NCR and ED means but shows no significant difference to ET mean. The non-enhancing parts appear brighter than necrosis, darker than edema and in similar shading than the enhancing tumor.
- In T2: NET mean is significantly different to NCR and ET means but shows no significant difference to ED mean. The non-enhancing parts show distinctly darker than necrosis, slightly darker than enhancing tumor and with similar intensity as edema.

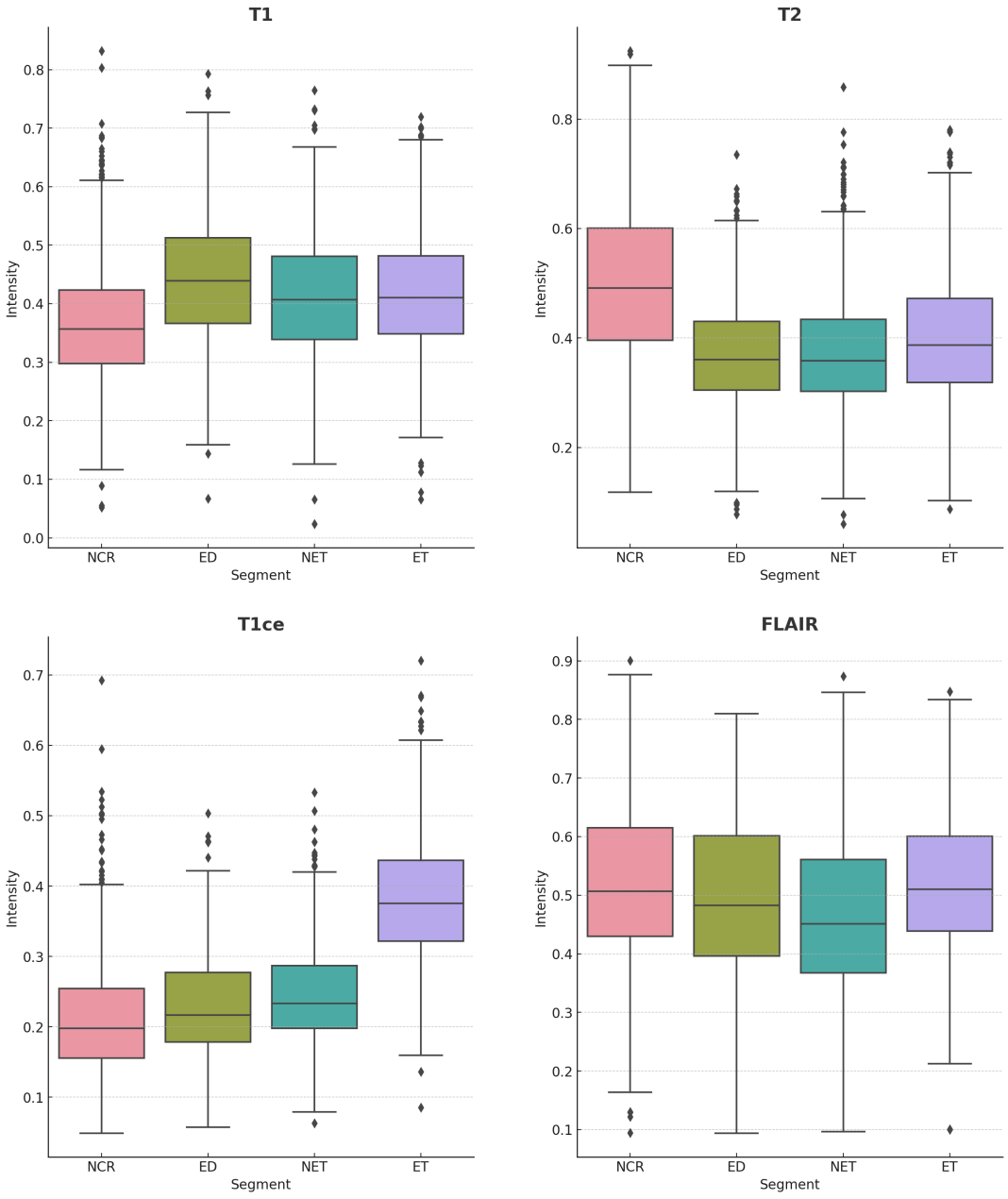

Figure 37: Boxplots for region intensities on different MRI modalities. The plots provide a visual representation of the distribution of intensity values, including the median, quartiles, and potential outliers for each segment (NCR, ED, NET, ET) in each modality (T1, T2, T1ce, FLAIR).

		meandiff	p-adj	lower	upper	reject
ED	ET	-0.0264	≤ 0.001	-0.0372	-0.0156	True
ED	NCR	-0.0755	≤ 0.001	-0.0863	-0.0648	True
ED	NET	-0.0298	≤ 0.001	-0.0406	-0.0190	True
ET	NCR	-0.0492	≤ 0.001	-0.0599	-0.0384	True
ET	NET	-0.0034	0.8263	-0.0142	0.0074	False
NCR	NET	0.0457	≤ 0.001	0.0350	0.0565	True

Table 25: Tukey HSD post-hoc comparison table for the **T1 modality**

		meandiff	p-adj	lower	upper	reject
ED	ET	0.0290	≤ 0.001	0.0165	0.0415	True
ED	NCR	0.1326	≤ 0.001	0.1201	0.1450	True
ED	NET	-0.0006	0.999	-0.0131	0.0119	False
ET	NCR	0.1036	≤ 0.001	0.0911	0.1161	True
ET	NET	-0.0295	≤ 0.001	-0.0420	-0.0171	True
NCR	NET	-0.1331	≤ 0.001	-0.1456	-0.1206	True

Table 26: Tukey HSD post-hoc comparison table for the **T2 modality**

		meandiff	p-adj	lower	upper	reject
ED	ET	0.1529	≤ 0.001	0.1449	0.1609	True
ED	NCR	-0.0173	≤ 0.001	-0.0253	-0.0094	True
ED	NET	0.0150	≤ 0.001	0.0070	0.0229	True
ET	NCR	-0.1702	≤ 0.001	-0.1782	-0.1623	True
ET	NET	-0.1379	≤ 0.001	-0.1459	-0.1300	True
NCR	NET	0.0323	≤ 0.001	0.0243	0.0403	True

Table 27: Tukey HSD post-hoc comparison table for the **T1ce modality**

		meandiff	p-adj	lower	upper	reject
ED	ET	0.0244	≤ 0.001	0.0107	0.0382	True
ED	NCR	0.0219	≤ 0.001	0.0082	0.0357	True
ED	NET	-0.0287	≤ 0.001	-0.0424	-0.0149	True
ET	NCR	-0.0025	0.966	-0.0162	0.0112	False
ET	NET	-0.0531	≤ 0.001	-0.0668	-0.0394	True
NCR	NET	-0.0506	≤ 0.001	-0.0643	-0.0369	True

Table 28: Tukey HSD post-hoc comparison table for the **FLAIR modality**

- In T1ce: NET mean is significantly different to NCR, ED and ET means. The non-enhancing tumor region appears distinctly darker than the enhancing tumor, slightly brighter than edema and brighter than necrosis.
- In Flair: NET mean is significantly different to NCR, ED and ET means. The non-enhancing tumor parts are darker than necrosis and slightly darker than edema and enhancing tumor.

All of these observations comply with the imaging characteristics described in Section 3.4 and Table 2.

A tumor region with a distinguished high or low intensity value within a given modality is represented as a bright or dark area compared to the remaining tumor regions in the corresponding MRI records. Since the non-enhancing part lies within the edema area and is likely to be surrounded by it, the similar means of both in the T2 modality will make the detection difficult. The same holds for the T1 modality, where NET and the enhancing tumor part have very near intensity values while also being located near or next to each other. This makes T1ce and Flair the two reasonable candidates for allocating the non-enhancing areas, although in practice the non-enhancing tumor can appear very variable on T1ce and the task of identification is still tough in many cases.

7.4 Unified 4-Label BraTS 18/21 Dataset

In the BraTS 2018 data, the non-enhancing regions could be isolated from necrosis as described in the previous Section 6.3. This will allow the creation of new segmentation data for BraTS 2018 containing the split versions of NET and NCR instead of the original merged version.

For the Brats 2021 data a similar split can be performed by applying the NET prediction model which was won from the BraTS 2018 NET reconstruction data. The edema parts can be cleaned from non-enhancing regions and the predicted NET regions themselves can be added as a new label.

This will result in both modified datasets, BraTS 2018 as well as BraTS 2021 providing the same labels *NCR*, *ED*, *NET* and *ET*. With the data part also providing the same structure (T1, T2, T1ce and Flair records) both modified datasets become compatible to be unified to one large combined dataset **BraTS 2018/21**. Table 29 shows the single and isolated data labels of the large combined dataset. The combined **BraTS 2018/21** dataset has a total number of 1536 records together with the four labels *NCR*, *ED*, *NET* and *ET*.

Since the dataset structure of the unified BraTS 2018/21 dataset is compatible with it's predecessors, the upscaling pre-activation U-Net which was introduced in section 6.2 can now be trained on this larger dataset without major modification. For retrieving better evaluation insights and for a more detailed outcome of the segmentation

Label	Data Name	Description
0	*BG*	Background
1	*NCR*	Necrosis
2	*ED*	Edema
3	*NET*	Non-Enhancing Tumor
4	*ET*	Enhancing Tumor

Table 29: Unified BraTS 2018/21 dataset labels. The combined BraTS 2018/21 dataset contains 1536 records together with the four labels *NCR*, *ED*, *NET* and *ET*.

results on the newly introduced NET part, two additional evaluation segments were introduced. Besides the three original evaluation segments for enhancing tumor *ET*, tumor core *TC* and whole tumor *WT*, the newly added non-enhancing tumor region *NET* itself is added as a fourth evaluation segment. This way the traditional *ET*, *TC* and *WT* can be compared towards reference models and the results for *NET* will be output as well.

The original tumor core part *TC* is combined similar to the one in BraTS 2021, without the non-enhancing parts being included. In BraTS 2021 the non-enhancing tumor part was counted towards edema which is a frequent way to treat the non-enhancing region. Also in the clinical context of treatment or resection, the focus lies on the tumor core parts including the necrotic parts as well as the active or enhancing region. The non-enhancing tumor part is a more neglected part, often due to identification problems as discussed in Section 2.4. But due to the NET's importance in medical treatment and tumor diagnosis, the tumor core part is increasingly suggested to be extended by the non-enhancing region. This extended tumor core part will be denoted *TCN* here and includes the necrotic, the enhancing and the non-enhancing parts. Tumor core with non-enhancing *TCN* will be added as a fifth evaluation segment in the training of the large upscaling U-Net model on the unified BraTS 2018/21 data. All evaluation segments and their composition based the single tumor parts are listed in Table 30.

Eval. Segment	BraTS 2018/21 Data	BraTS 2018/21 Labels	Description
ET	*ET*	4	Enhancing tumor
TC	*ET* + *NCR*	4, 1	Tumor core
WT	*ET* + *NCR* + *NET* + *ED*	4, 1, 3, 2	Whole tumor
NET	*NET*	3	Non-Enhancing tumor
TCN	*ET* + *NCR* + NET	4, 1, 3	Tumor core w. Non-Enhancing

Table 30: Unified BraTS 2018/21 evaluation segments as compositions of data labels. *NET* is added as a evaluation segment for itself. *TCN* is added as a new evaluation segment, which extends the traditional *TC* by the non-enhancing regions.

The unified BraTS 2018/21 dataset has a total number of 1536 records which was divided into a training dataset holding 1305 records (85% of total) and a validation

dataset with 231 records. The structure of the presented upscaling pre-activation U-Net was modified to have an output layer suitable for the five segments *ET*, *TC*, *WT*, *NET*, *TCN* and the training and validation data was pre-processed accordingly with a center cropped size of $4 \times 96 \times 192 \times 160$ for each record (channels, axial, coronal, sagittal). The metric and loss functions were kept as before, which means the Sørensen-Dice coefficient was used as a metric for each of the single segments. The Soft-Dice-Loss function was used as a combined loss function. The batch size was again set to 1 due to high memory demand, the learning rate was selected to start at 10^{-4} with an exponential decay of 5% per epoch.

The GPU accelerator used for training was a powerful NVIDIA A100-SXM4 with 40GB RAM which allowed a training speed of 1177ms/step or approximately 26 minutes per epoch. The training was run for 45 epochs which resulted in a total training time of 19.5 hours for the large prediction model.

The last column of Table 31 shows the performance results of the model as Sørensen-Dice coefficient, evaluated on the validation data. On the enhancing tumor part *ET* the new model reaches a Dice-score of 81.94%, for tumor core 89.97% and for whole tumor 91.51% is reached. This gives a mean Dice-coefficient over the *ET*, *TC*, *WT* of 87,75%. Table 31 also shows performance results of the previously discussed BraTS winning models of Section 5.2. The upscaling pre-activation U-Net model introduced on this work, which was trained on the unified BraTS 2018/21 dataset achieves competitive performance results between the winning models. It outperforms the original U-Net with Variational Autoencoder (SegResNetVAE) model by Myronenko from 2018 [87], the Two Stage Cascaded (TSC) U-Net by Jiang et al. from 2019 [59] as well as the original No-New U-Net (nnU-Net) by Isensee et al. from 2020 [56] in all three classical categories while additionally and consistently providing the *NET* predictions.

	SegResNetVAE 2018 [87]	**TSC U-Net 2019 [59]**	**nnU-Net 2020 [56]**	**SegResNet 2021 [105]**	**nnU-Net 2021 [78]**	**Ours**
NET	-	-	-	-	-	0.5320
TCN	-	-	-	-	-	0.8846
ET	0.8145	0.8021	0.7945	0.8600	0.8823	0.8194
TC	0.8596	0.8647	0.8524	0.8868	0.9235	0.8979
WT	0.9042	0.9094	0.9119	0.9265	0.9383	0.9151
Mean	**0.8594**	**0.8587**	**0.8529**	**0.8911**	**0.9147**	**0.8775**

Table 31: Validation results, combined 4-label BraTS 2018/21 data, upscaling pre-activation U-Net, learning rate 1e-4 with 5% decay, training data size 1305, validation data size 231, batch size 1, 45 epochs, NVIDIA A100-SXM4-40GB, training speed 1563s/epoch 1177ms/step, total training time 19.5h, 2.879M parameters. Values are Sørensen-Dice-coefficients on validation data. Mean values are calculated from ET, TC, WT for all models (NET and TCN segments don't count into mean value).

The performance results stand back behind the 2021 model of Myronenko et al. (Seg-

ResNet) [105] and the 2021 No-New U-Net model of Luu et al. (nnU-Net 2021) [78]. When compared to the SegResNet 2021 model the gap in *ET* Dice-score is high, while it shows minor performance drawback in the *WT* category and is even slightly superior in *TC*. The results of the improved nnU-Net version from 2021 are more far ahead.

In total the model presented in this work shows competitive performance results in the original BraTS segmentation categories without reaching or surpassing the state-of-the-art performance of the latest winning models. Concurrently, it offers two important advantages over existing models. On the one hand, it will automatically output resolution enhanced segmentation masks serving a more detailed and advanced view onto the tumor's structure. This can help radiologists and clinicians to visualize and understand the shape and extent of the tumor and it's parts. On the other hand it will, in addition to the regular segments, also predict the important non-enhancing regions, which are hard to locate on the conventional radiology records. This can give new options in radiology assistance and treatment planning. A further aspect in analysing and diagnosing the active tumor region is the extended *TCN* segment which grows the classical tumor core *TC* by the relevant non-enhancing part and can be a sophisticated complement in the hierarchical order of a glioma's extent:

$$ET \subset TC \subset TCN \subset WT$$

The overall performance value of *TCN* using the presented model on the unified BraTS 2018/21 dataset is 88.46% and therefore very comparable to that one of *TC*, which is 89.79%. The Dice-coefficient performance of *NET* on the given dataset is 53.20% and is low compared to the remaining segments. This is due to the subtle nature of the non-enhancing regions, which are only showing very slightly on the conventional MRI sequences. The lack of curated non-enhancing tumor data and the practical limitations in representative reconstruction from given data are also contributing to inhomogeneous NET data in the constructed dataset. This necessarily leads to an overall upper bound performance in NET identification, given a proper regularization behaviour and robustness of the model.

7.5 3D Model Reconstruction

3D models of tumors, derived from the medical imaging MRI data, can significantly aid in both the diagnosis and treatment of cancer in several ways. The models can be able to offer a detailed and accurate representation of tumors, providing insights that can enhance the precision and efficiency of the cancer therapy.

The major advantage of 3D tumor models lies in the ability to accurately visualize the size, shape, and volume of tumors. They provide a clear delineation of tumor

boundaries from surrounding tissues, aiding in understanding its invasiveness and potential spread. Additionally, the structure of the tumor seen from these models can offer valuable information about its type and aggressiveness. Also 3D models have become useful tools for preoperative planning, allowing surgeons to arrange the surgical removal of tumors with minimal impact on healthy tissue. The additional information on non-enhancing areas, which can be added by the developed model can therefore support the decision towards boundaries in surgical removal. The 3D models can also serve for an educational purpose, aiding in patient understanding of surgical procedures and enabling surgeons to practice and simulate complex operations and thereby reducing risks and improving surgical outcomes.

The raw MRI data comes in a number of slices, each representing a 2D pixel image. Stacking the slices of the binary segmentation masks using the given resolution of the MRI device gives a raw voxel representation of the tumor, where a voxel, or volumetric pixel, is the three-dimensional equivalent of a pixel in 2D images. Voxel representations are the very first and basic version of generating such a 3D model representation. A major advance in efficiently creating more enhanced 3D models was the introduction of the so called *Marching Cubes* algorithm by William E. Lorensen and Harvey E. Cline in 1987 [73]. The Marching Cubes algorithm is a computer graphics algorithm used for extracting a polygonal mesh of an isosurface from a three-dimensional scalar or binary field as given in the 3D volumetric MRI data.

For creating 3D volumetric models based on the segmentation masks generated from BraTS 2021 data by the model presented in the previous section, the tool **Medical Imaging Interaction Toolkit (MITK)** developed by the German Cancer Research Center DKFZ [39] was used. The MITK Workbench is a free open-source software for viewing, processing, and segmenting medical images.

The top left image of Figure 38 shows the marching cube model created from the original resolution version of the new BraTS 2021 segmentation mask. The bottom left version shows a smoothed polygon 3D model which can be calculated by the MITK software that applies algorithms like Laplacian smoothing to the volume mesh. This involves adjusting the position of each vertex in the mesh based on the average of its neighbors, leading to a smoother appearance.

The smoothed 3D volumes are providing a clearer view and improve the visual quality of the reconstructed surfaces on of the overall shape of the tumor, a drawback although is the loss of details that the smoothing process entails. This can be observed in small structures being lost after smoothing, as marked with arrows in the top left and bottom left images in Figure 38.

Increasing the resolution of the MRI data for the volume construction can therefore result in more detailed and smoother surfaces. The resolution enhanced output of the proposed upscaling U-Net model enables this and the generated 3D volumes can

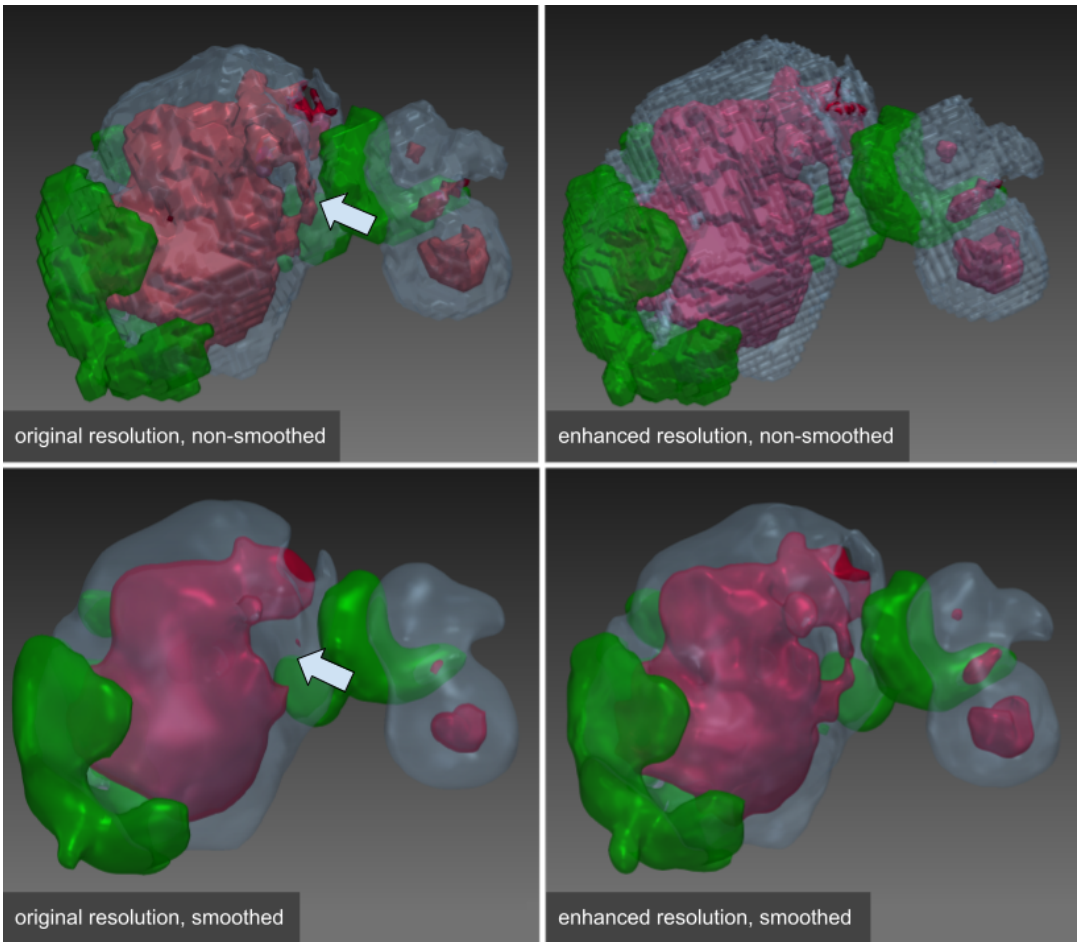

Figure 38: 3D volumetric model of BraTS2021 record no. 00209. The top row shows the marching cubes reconstructions, the bottom row shows the smoothed versions. The two models on the left side were created using the original, lower resolution of the segmentation masks. The two volumes on the right side were generated from the enhanced, higher resolution outputs of the presented upscaling pre-activation U-Net. All 3D models were created using the open source MITK Workbench, version v2023.04, published by the German Cancer Research Center DKFZ. The arrows on the left images mark small parts in the lower resolution versions, which can get lost due to the smoothing process. The tumor parts shown in the 3D model are necrosis NCR (red), enhancing tumor ET (blue) and non-enhancing tumor NET (green).

thus offer a richer and more detailed insight in the tumors structure. Another example visualizing the advantage of higher resolution segmentation outputs for 3D reconstruction is shown in Figure 39 which presents the whole tumor and the isolated necrosis part of a BraTS 2021 sample, transformed to a smoothed polygon model from original resolution (left) as well as the same one from the enhanced resolution version (right). Appendix 9.8 shows a collection of more generated 3D models.

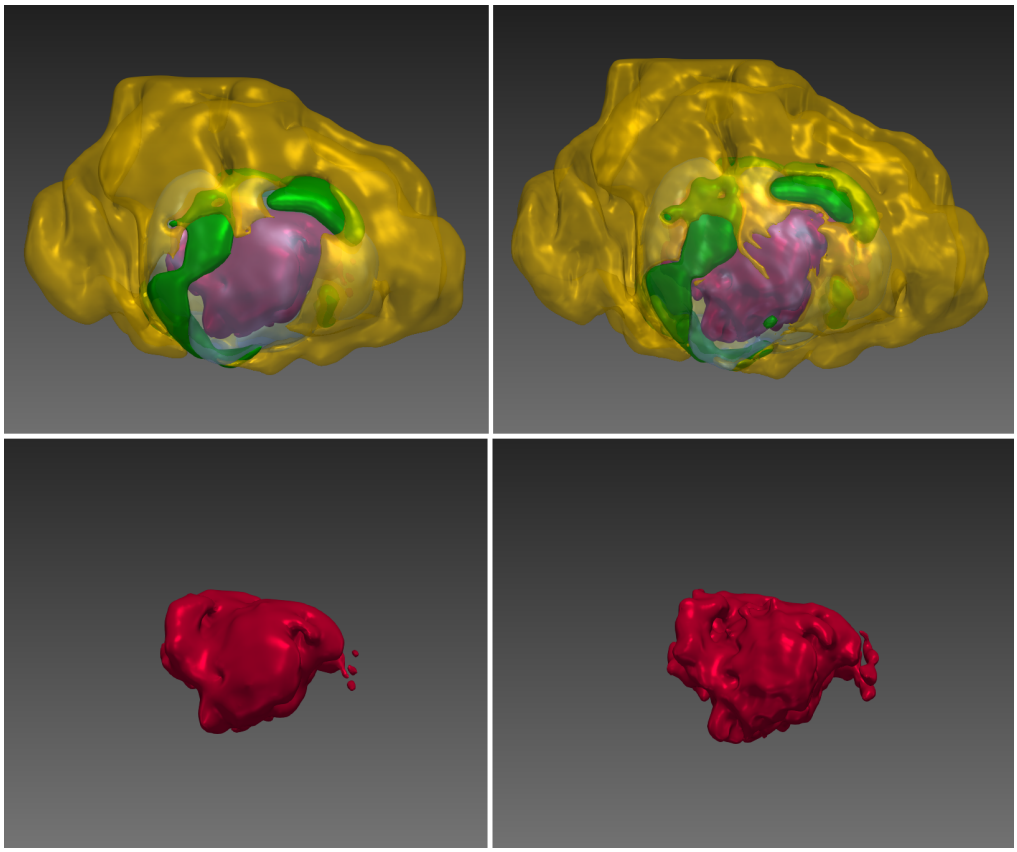

Figure 39: 3D volumetric model, smoothed version of WT and NCR, based on original resolution (left) and on enhanced resolution (right) BraTS2021 record 00216. Top left: Whole tumor, smoothed polygon model from original BraTS resolution. Bottom left: Necrosis, smoothed polygon model from original resolution. Top right: Whole tumor, smoothed polygon model from enhanced resolution output. Bottom right: Necrosis, smoothed polygon model from enhanced resolution output. Edema (ED) is shown yellow, necrosis (NCR) red, enhancing tumor (ET) is transparent blue, non-enhancing tumor (NET) is green.

A characteristic aspect of the non-enhancing tumor region which can be observed in many of the 3D visualizations is it's eccentric localization within the total tumor area. As stated in a quote from the Lasocki paper [65] in Section 3.4, edema typically occurs in a concentric distribution around the tumor core (unless it is obstructed in some way) while the NET regions often arise in an eccentric location towards the enhancing part in the tumor center. An example showing the tumor core (NCR + ET) with evenly spread surrounding edema (ED) is presented in the left image of Figure 40, with the NET region suppressed. The right image in Figure 40 shows the pre-

dicted NET region inside the same tumor, which is, in contrast to edema not spread evenly around the core part, but instead appears accumulated on one side. This characteristic of NET can be confirmed in more of the 3D models given in Appendix 9.8.

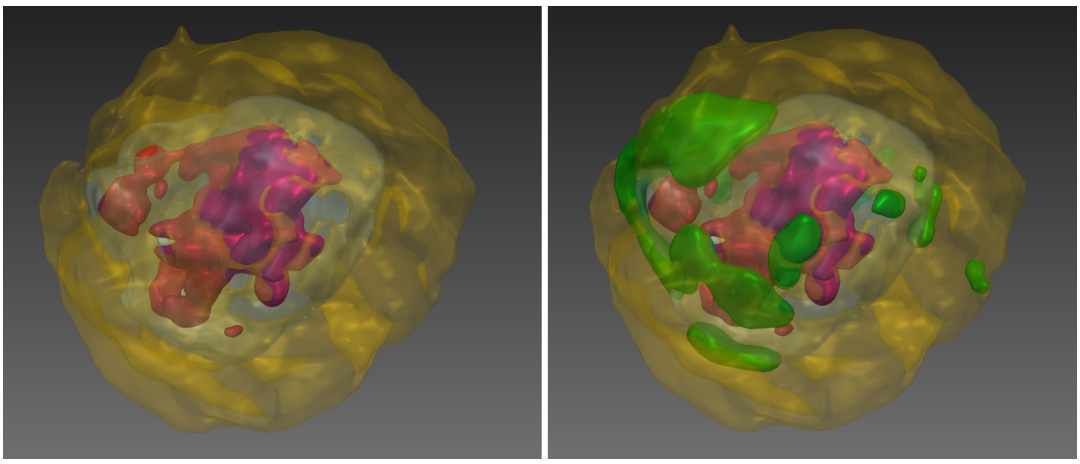

Figure 40: 3D volumetric model, smoothed version of WT of NCR, based on original resolution (left) and on enhanced resolution (right) BraTS2021 record 00413. Top left: Whole tumor, smoothed polygon model from original BraTS resolution. Bottom left: Necrosis, smoothed polygon model from original resolution. Top right: Whole tumor, smoothed polygon model from enhanced resolution output. Bottom right: Necrosis, smoothed polygon model from enhanced resolution output. Edema (ED) is shown yellow, necrosis (NCR) red, enhancing tumor (ET) is transparent blue, non-enhancing tumor (NET) is green.

8 Conclusion

This study has shown advancements in the identification and segmentation of non-enhancing tumor regions in glioblastomas using established MRI techniques together with sophisticated machine learning methods. The findings demonstrate the potential of the proposed upscaling pre-activation U-Net model in improving the detection and characterization of NET regions.

The core of this work was the development and validation of the upscaling pre-activation U-Net model, enhanced with depth regularization, which has demonstrated promising potential in detecting and characterizing NET regions, a task traditionally challenging due to the subtle and diffuse nature of these lesions.

The model's unique approach in integrating a resolution enhancing extension in the decoder branch has contributed towards a more accurate identification of NET regions. This is particularly crucial in glioblastoma, where accurate delineation of the non-enhancing regions can significantly impact treatment planning and prognosis assessment. The accurate segmentation of NET regions can aid in the better understanding of tumor boundaries, potentially leading to more targeted and effective treatment approaches, including surgery and radiotherapy.

However, the study is not without limitations. The primary challenge is the limited availability of appropriate data representing the subjected non-enhancing areas, which was addressed by using reconstructed data from existing datasets. The used data was extracted and prepared by a well-founded automatic process but still it is not hand-curated similarly to the original BraTS data. A major improvement to the automated identification and detection of non-enhancing tumor areas will therefore depend on the creation and availability of consistent, expert-based segmentation data for training.

Furthermore and to move forward, it has to be considered to translate these findings into clinical practice. Future studies could focus on integrating the model into clinical workflows, assessing its impact on patient outcomes and refining it's capabilities to serve in the finding and diagnosis of glioblastoma. Another promising field for future research is the exploration of the biological significance of NET regions. Understanding these areas allows new insights into glioblastoma's diverse and complex nature.

In conclusion, this work presents another step in the offering medical and radiological assistance applications using latest machine learning methods. The further development of refined prediction models has the potential to improve the diagnosis, treatment planning, and management of glioblastoma.

9 Appendix

9.1 Code: Dice Coefficient and Dice Loss

```
import tensorflow as tf
import tensorflow.keras.backend as K

def dice_coefficient(y_true, y_pred):
    y_pred = (y_pred > 0.5)
    y_pred = tf.dtypes.cast(y_pred, tf.float32)
    intersection = K.sum(K.abs(y_true * y_pred), axis=[-3,-2,-1])
    dn = K.sum(y_true, axis=[-3,-2,-1]) + K.sum(y_pred, axis=[-3,-2,-1]) + 1e-8
    return K.mean(2 * intersection / dn, axis=[0,1])

def soft_dice_loss(y_true, y_pred):
    y_true = tf.dtypes.cast(y_true, tf.float32)
    intersection = K.sum(K.abs(y_true * y_pred), axis=[-3,-2,-1])
    dn = K.sum(K.square(y_true) + K.square(y_pred), axis=[-3,-2,-1]) + 1e-8
    return 1 - K.mean(2 * intersection / dn, axis=[0,1])
```

9.2 Code: Data pre-processing for evaluation segments

```
def preprocess_labels(seg, output_shape=None):

    """
    Brats 2018
    1: Necrosis and non-enhancing tumor NCR + NET
    2: Edema ED
    3: not used
    4: Enhancing tumor ET
    ---
    Brats 2021
    1: Necrosis NCR
    2: Edema and non-enhancing tumor ED + NET
    3: not used
    4: Enhancing tumor ET
    """

    ncr = seg == 1
    ed = seg == 2
    et = seg == 4
    tc = (et | ncr)
    wt = (et | ncr | ed)

    if output_shape is not None:
        et = crop(et, output_shape)
        tc = crop(tc, output_shape)
        wt = crop(wt, output_shape)

    return np.array([et, tc, wt], dtype=np.uint8)
```

9.3 Code: Filter blocks (Tensorflow/Keras)

```
def residual_block(inp, filters, type = 'plain'):

    assert type in ['plain', 'resnet', 'resnetn', 'preact'], "Unknown type"
    cf = 'channels_first'

    if type == 'plain':

      # Plain U-Net filter block

      x = Conv3D(filters=filters, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(inp)
      out = Activation('relu')(x)

    elif type == 'resnet':

      # ResNet like filter block

      x = Conv3D(filters=filters, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(inp)
      x = Activation('relu')(x)
      x = Conv3D(filters=filters, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(x)
      out = Add()([x, inp])

    elif type == 'resnetn':

      # ResNet like filter block with normalization

      x = Conv3D(filters=filters, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(inp)
      x = GroupNormalization(groups=8, axis=1)(x)
      x = Activation('relu')(x)
      x = Conv3D(filters=filters, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(x)
      x = GroupNormalization(groups=8, axis=1)(x)
      x = Add()([x, inp])
      out = Activation('relu')(x)

    elif type == 'preact':

      # Pre-activation filter block

      x = GroupNormalization(groups=8, axis=1)(inp)
      x = Activation('relu')(x)
      x = Conv3D(filters=filters, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(x)
      x = GroupNormalization(groups=8, axis=1)(x)
      x = Activation('relu')(x)
      x = Conv3D(filters=filters, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(x)
      out = Add()([x, inp])

    else:

      out = inp

    return out
```

9.4 Code: Upscaling 4-layer U-Net (Tensorflow/Keras)

```
# Input layer
inp = Input((4, 96, 192, 160))
cf = 'channels_first'

# Initial block
x = Conv3D(filters=32, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(inp)

# Dropout
x = SpatialDropout3D(0.2, data_format=cf)(x)

# Encoder block
x0 = residual_block(x, 32, filter_block_type)

# Branch for upscaling path
xu = UpSampling3D(size=2, data_format='channels_first')(x0)
xu = Conv3D(filters=16, kernel_size=(3, 3, 3), strides=1, padding='same', data_format=cf)(xu)
xu = residual_block(xu, 16, filter_block_type)

# Downsampling
x = Conv3D(filters=64, kernel_size=(3, 3, 3), strides=2, padding='same', data_format=cf)(x0)

# Encoder blocks
x = residual_block(x, 64, filter_block_type)
x1 = residual_block(x, 64, filter_block_type)

#  Downsampling
x = Conv3D(filters=128, kernel_size=(3, 3, 3), strides=2, padding='same', data_format=cf)(x1)

# Encoder blocks
x = residual_block(x, 128, filter_block_type)
x = residual_block(x, 128, filter_block_type)

#  Upsampling
x = Conv3D(filters=64, kernel_size=(1, 1, 1), strides=1, data_format='channels_first')(x)
x = UpSampling3D(size=2, data_format=cf)(x)
x = Add()([x, x1])

# Decoder block
x = residual_block(x, 64, filter_block_type)

#  Upsampling
x = Conv3D(filters=32, kernel_size=(1, 1, 1), strides=1, data_format='channels_first')(x)
x = UpSampling3D(size=2, data_format=cf)(x)
x = Add()([x, x0])

# Decoder block
x = residual_block(x, 32, filter_block_type)

#  Upsampling
x = Conv3D(filters=16, kernel_size=(1, 1, 1), strides=1, data_format=cf)(x)
x = UpSampling3D(size=2, data_format=cf)(x)
x = Add()([x, xu])

# Decoder block
x = residual_block(x, 16, filter_block_type)

# Decoder end
out = Conv3D(filters=3, kernel_size=(1, 1, 1), strides=1, data_format=cf, activation='sigmoid')(x)

# Create and compile the model
model = tf.keras.Model(inp, outputs=[out])
model.compile(Adam(), loss=[loss_gt(soft_dice_loss)], metrics=[dice_coefficient])
```

9.5 Slices Showing Resolution Enhancement and NET

Left: BraTS 2021 ground truth segmentation, original resolution, edema (yellow), enhancing tumor (blue) and necrosis (red). Right: Prediction using the presented upscaling pre-activation U-Net, including the predicted non-enhancing tumor region (green).

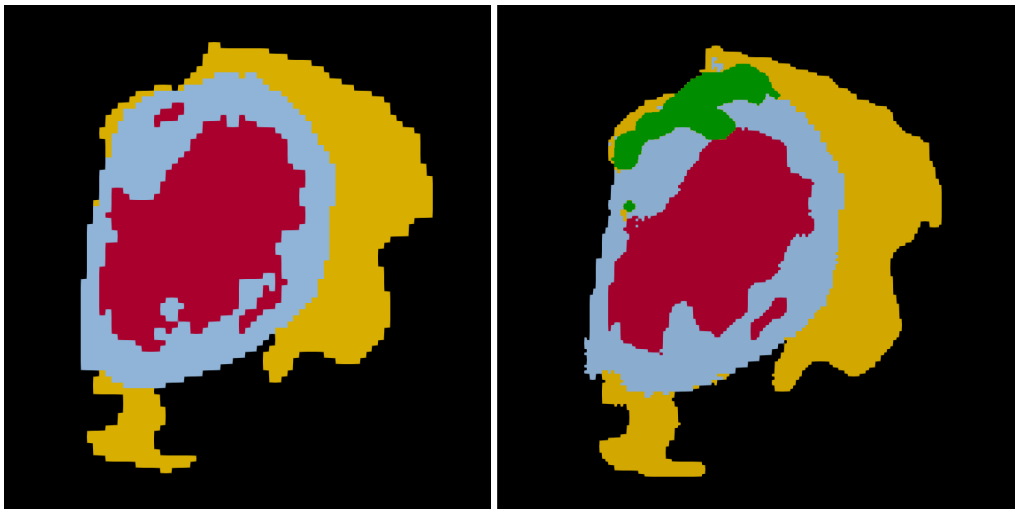

Figure 41: Sample using resolution enhancing model including NET prediction, BraTS 2021 record 00211.

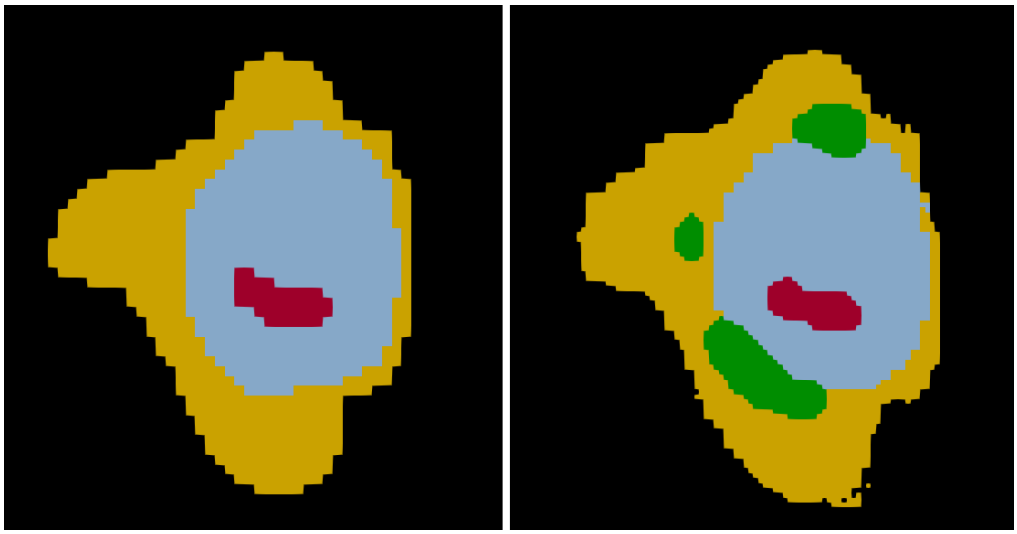

Figure 42: Sample using resolution enhancing model including NET prediction, BraTS 2021 record 00206.

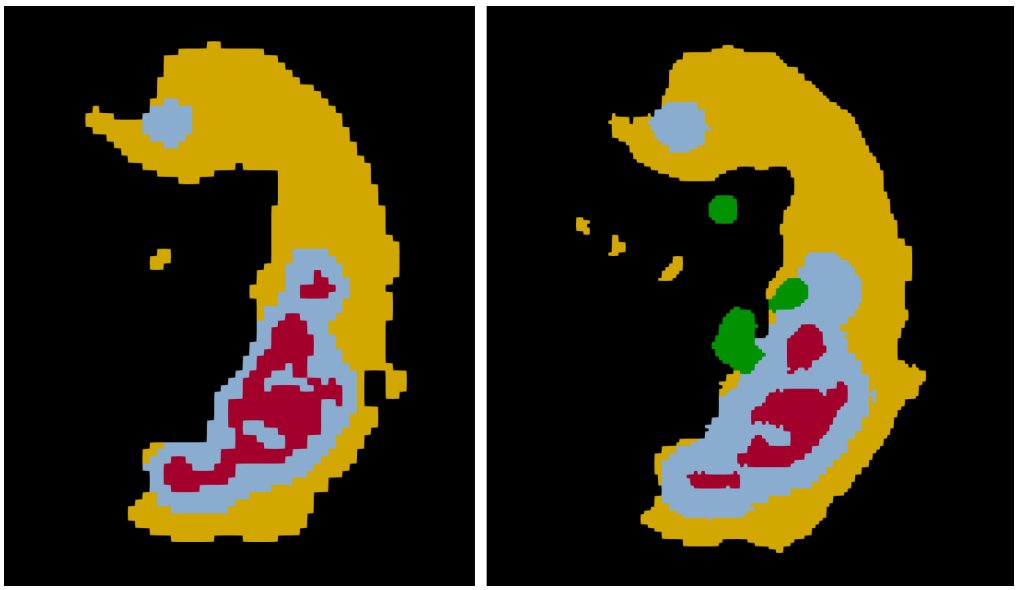

Figure 43: Sample using resolution enhancing model including NET prediction, BraTS 2021 record 00207.

9.6 BraTS2021 00589 Segmentation Slices

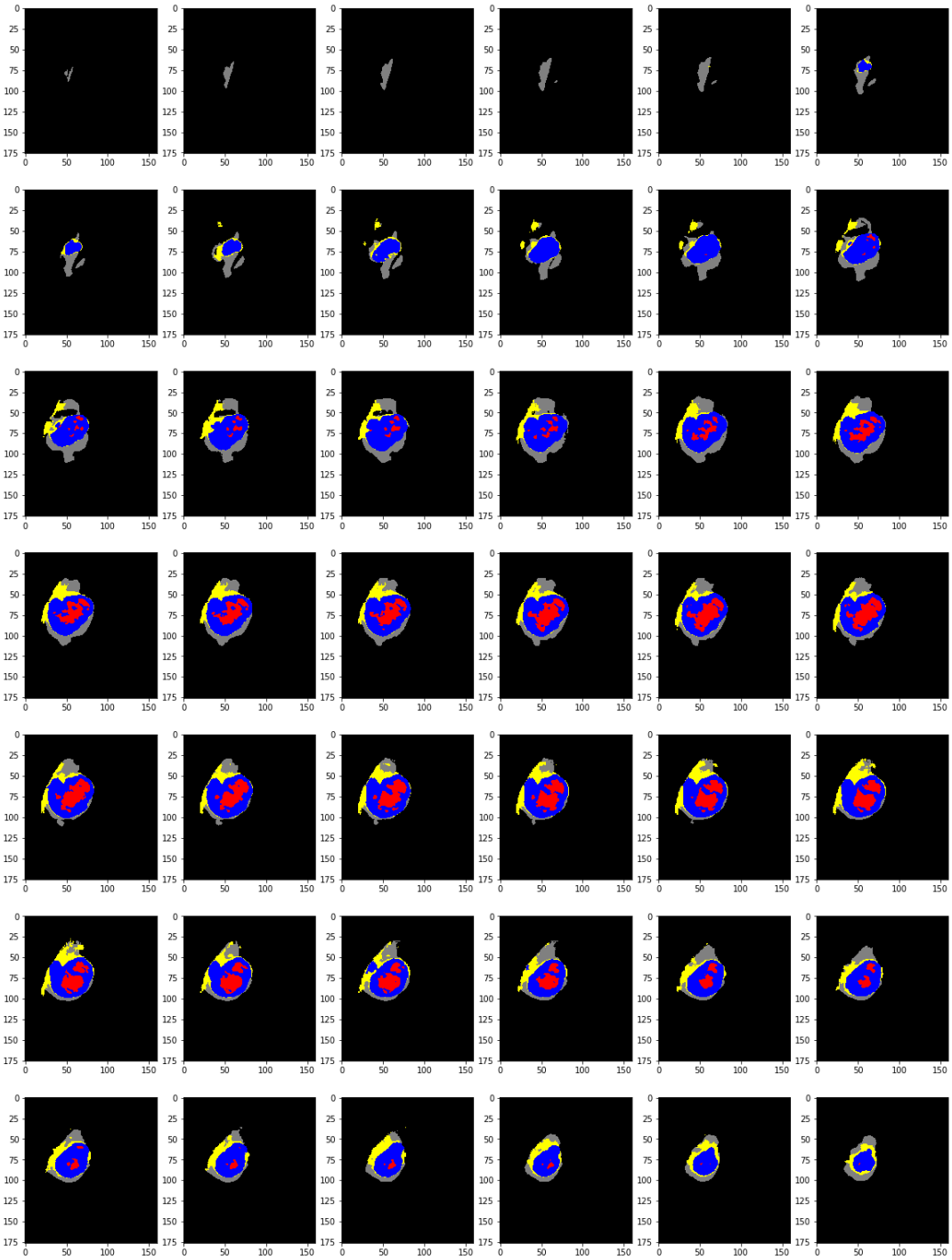

Figure 44: BraTS2021-00589 Segmentation, gray: Edema (from Brats2021), red: Necrosis (from Brats2021), blue: Enhancing tumor (from Brats2021), yellow: Non-Enhancing tumor (new, predicted by model based on Brats2018 data)

9.7 BraTS2021 Segmentation Samples

In each sample, the left column shows the original segmentation (top) of enhancing tumor (ET blue), edema (ED yellow) and necrosis (NCR red) as well as the new segmentation mask extended by the prediction for non-enhancing NET (bottom image, NET green). The middle and right columns show the T1, T2, T1ce and Flair sequence images. The arrows are pointing onto shaded areas indicating NET.

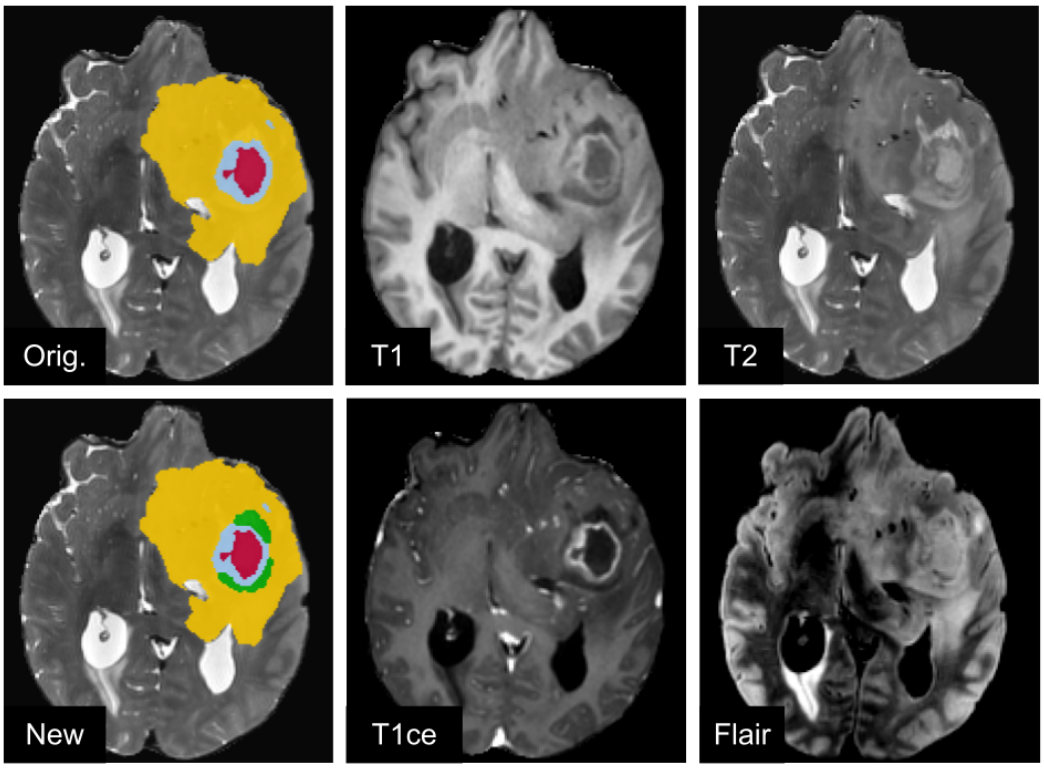

Figure 45: BraTS 2021, record 00501, axial slices.

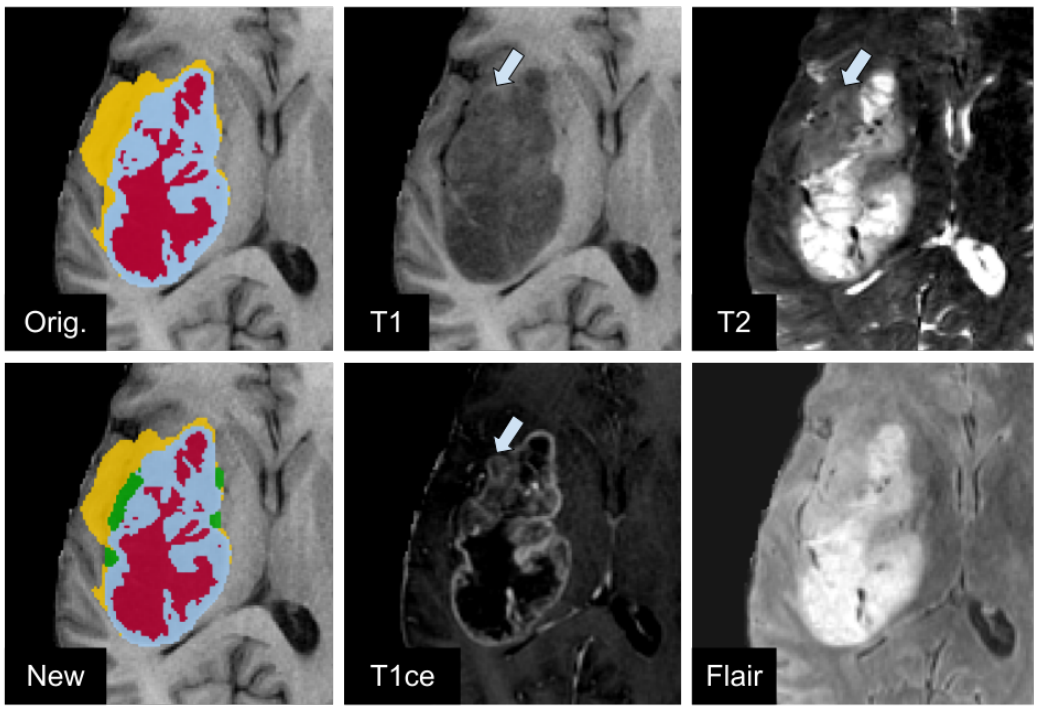

Figure 46: BraTS 2021, record 00513, axial slices.

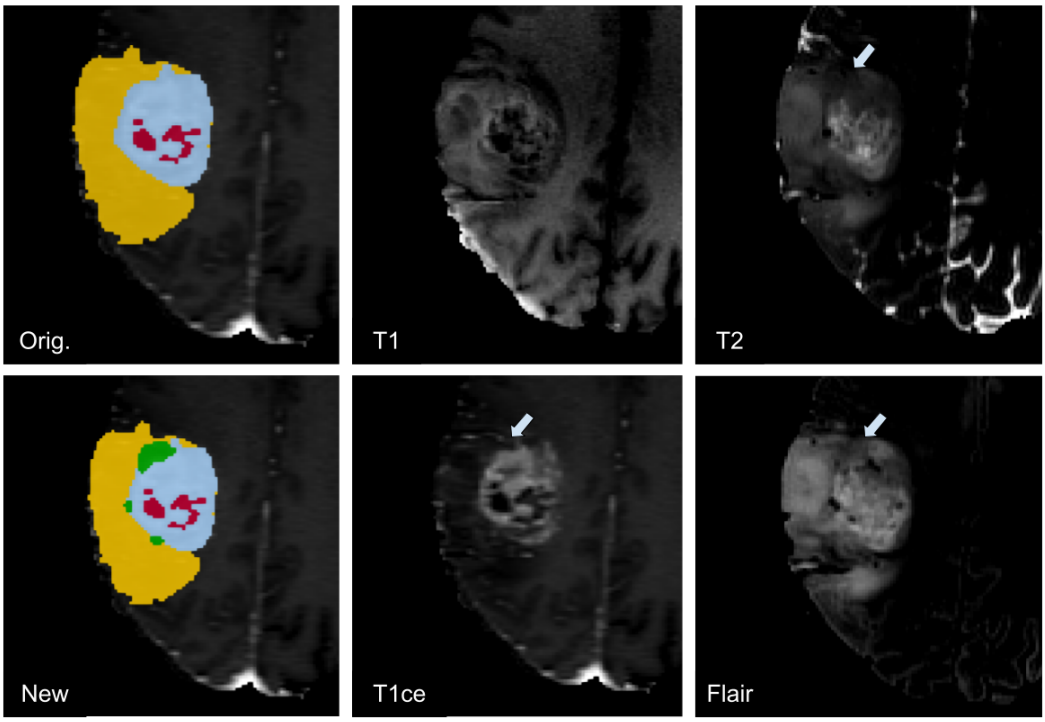

Figure 47: BraTS 2021, record 00084, axial slices.

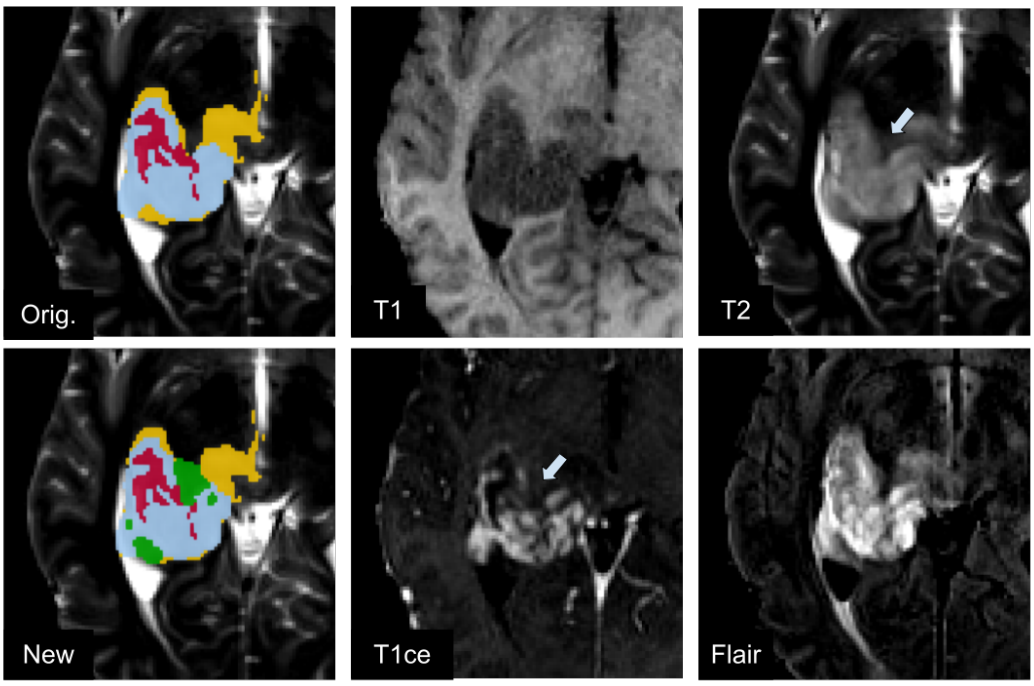

Figure 48: BraTS 2021, record 00334, axial slices.

9.8 3D Tumor Models

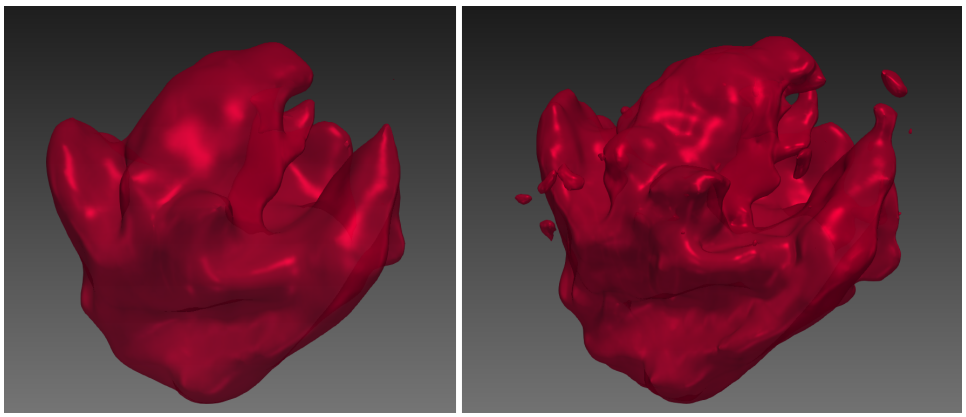

Figure 49: 3D volumetric model, smoothed version of NCR (red), BraTS 2021 record 00201. Left: Based on model prediction in original BraTS resolution. Right: Based on enhanced resolution output from presented upscaling U-Net.

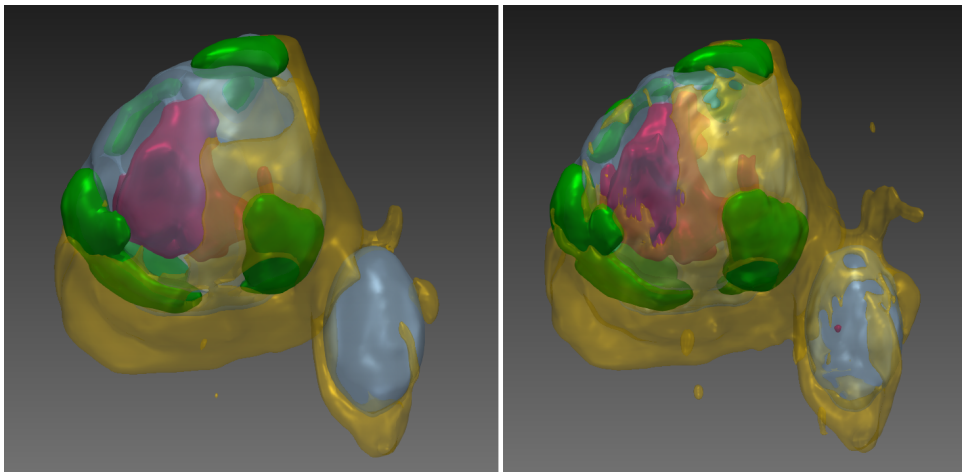

Figure 50: 3D volumetric model, smoothed version of NCR (red), ET (blue), NET (green) and ED (yellow), BraTS 2021 record 00214. Left: Based on model prediction in original BraTS resolution. Right: Based on enhanced resolution output from presented upscaling U-Net.

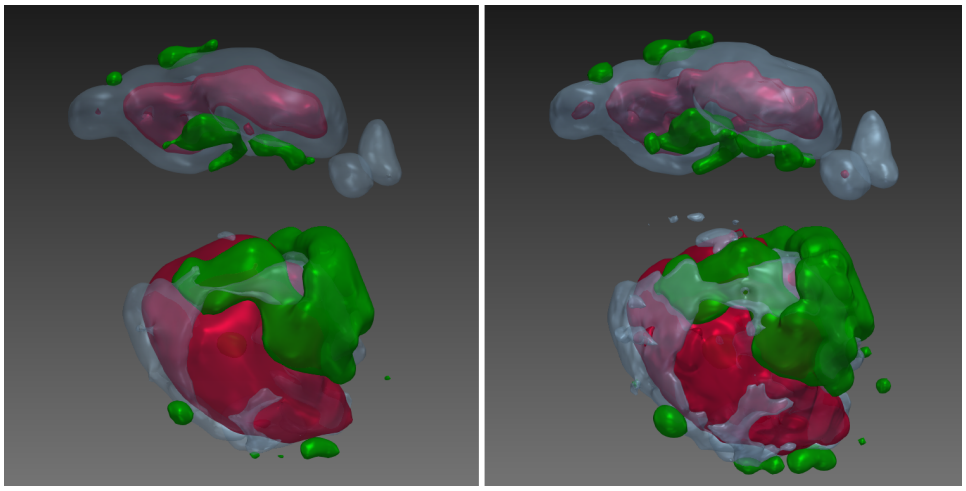

Figure 51: 3D volumetric model, smoothed version of NCR (red), ET (blue), NET (green) and ED (yellow), BraTS 2021 record 00207. Left: Based on model prediction in original BraTS resolution. Right: Based on enhanced resolution output from presented upscaling U-Net.

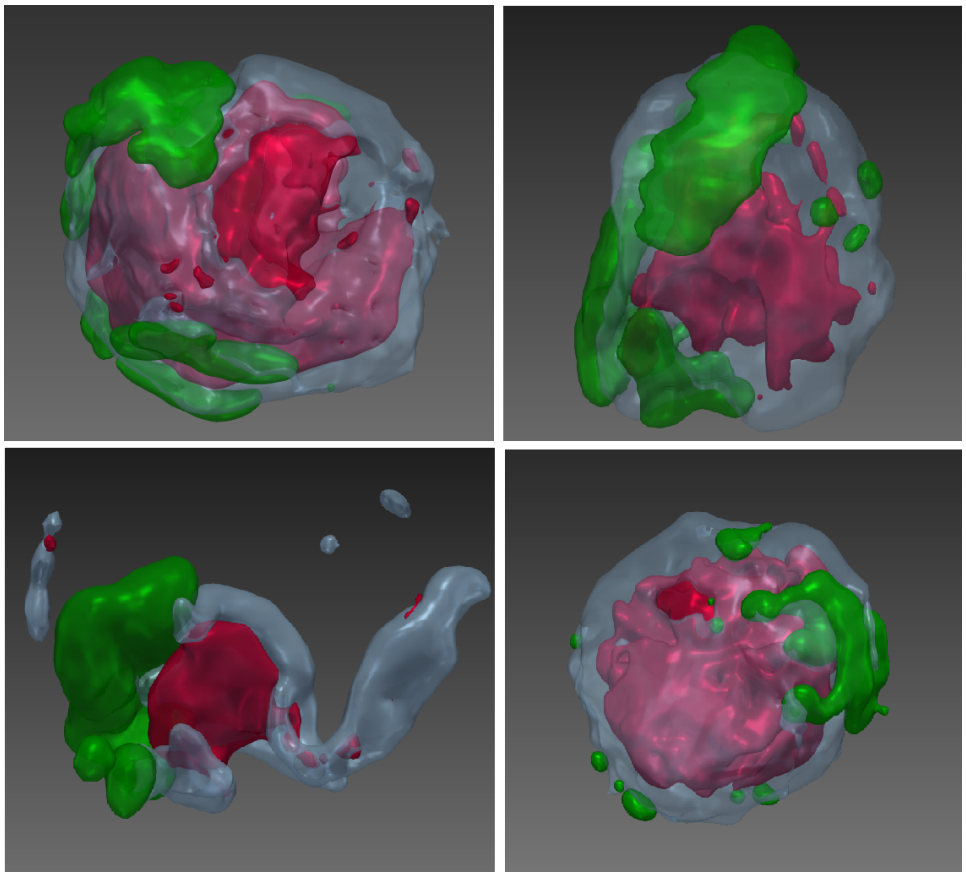

Figure 52: 3D volumetric model, smoothed version of NCR (red), ET (blue), NET (green) and ED (yellow), BraTS 2021 record 00207. Left: Based on model prediction in original BraTS resolution. Right: Based on enhanced resolution output from presented upscaling U-Net.

References

[1] R. Agravat and M. S. Raval. Brain tumor segmentation and survival prediction, 2019.

[2] V. K. Anand, S. Grampurohit, P. Aurangabadkar, A. Kori, M. Khened, R. S. Bhat, and G. Krishnamurthi. Brain tumor segmentation and survival prediction using automatic hard mining in 3d cnn architecture, 2021.

[3] R. Ansorge and M. Graves. *The Physics and Mathematics of MRI*. 2053-2571. Morgan & Claypool Publishers, 2016.

[4] R. Ayachi, M. Afif, Y. Said, and M. Atri. *Strided Convolution Instead of Max Pooling for Memory Efficiency of Convolutional Neural Networks*, pages 234–243. 01 2020.

[5] W. Baccouch, S. Oueslati, B. Solaiman, and S. Labidi. A comparative study of cnn and u-net performance for automatic segmentation of medical images: application to cardiac mri. *Procedia Computer Science*, 219:1089–1096, 2023. CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022.

[6] U. Baid, S. Ghodasara, M. Bilello, S. Mohan, E. Calabrese, E. Colak, K. Farahani, J. Kalpathy-Cramer, F. C. Kitamura, S. Pati, L. M. Prevedello, J. D. Rudie, C. Sako, R. T. Shinohara, T. Bergquist, R. Chai, J. A. Eddy, J. Elliott, W. Reade, T. Schaffter, T. Yu, J. Zheng, B. Annotators, C. Davatzikos, J. Mongan, C. Hess, S. Cha, J. E. Villanueva-Meyer, J. B. Freymann, J. S. Kirby, B. Wiestler, P. Crivellaro, R. R. Colen, A. Kotrotsou, D. S. Marcus, M. Milchenko, A. Nazeri, H. M. Fathallah-Shaykh, R. Wiest, A. Jakab, M. Weber, A. Mahajan, B. H. Menze, A. E. Flanders, and S. Bakas. The RSNA-ASNR-MICCAI brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *CoRR*, abs/2107.02314, 2021.

[7] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, M. Prastawa, E. Alberts, J. Lipkova, J. Freymann, J. Kirby, M. Bilello, H. Fathallah-Shaykh, R. Wiest, J. Kirschke, B. Wiestler, R. Colen, A. Kotrotsou, P. Lamontagne, D. Marcus, M. Milchenko, A. Nazeri, M.-A. Weber, A. Mahajan, U. Baid, E. Gerstner, D. Kwon, G. Acharya, M. Agarwal, M. Alam, A. Albiol, A. Albiol, F. J. Albiol, V. Alex, N. Allinson, P. H. A. Amorim, A. Amrutkar, G. Anand, S. Andermatt, T. Arbel, P. Arbelaez, A. Avery, M. Azmat, P. B., W. Bai, S. Banerjee, B. Barth, T. Batchelder, K. Batmanghelich, E. Battistella, A. Beers, M. Belyaev, M. Bendszus, E. Benson, J. Bernal, H. N. Bharath, G. Biros, S. Bisdas, J. Brown, M. Cabezas, S. Cao, J. M. Cardoso, E. N. Carver, A. Casamitjana, L. S. Castillo, M. Catà, P. Cattin, A. Cerigues, V. S.

Chagas, S. Chandra, Y.-J. Chang, S. Chang, K. Chang, J. Chazalon, S. Chen, W. Chen, J. W. Chen, Z. Chen, K. Cheng, A. R. Choudhury, R. Chylla, A. Clérigues, S. Colleman, R. G. R. Colmeiro, M. Combalia, A. Costa, X. Cui, Z. Dai, L. Dai, L. A. Daza, E. Deutsch, C. Ding, C. Dong, S. Dong, W. Dudzik, Z. Eaton-Rosen, G. Egan, G. Escudero, T. Estienne, R. Everson, J. Fabrizio, Y. Fan, L. Fang, X. Feng, E. Ferrante, L. Fidon, M. Fischer, A. P. French, N. Fridman, H. Fu, D. Fuentes, Y. Gao, E. Gates, D. Gering, A. Gholami, W. Gierke, B. Glocker, M. Gong, S. González-Villá, T. Grosges, Y. Guan, S. Guo, S. Gupta, W.-S. Han, I. S. Han, K. Harmuth, H. He, A. Hernández-Sabaté, E. Herrmann, N. Himthani, W. Hsu, C. Hsu, X. Hu, X. Hu, Y. Hu, Y. Hu, R. Hua, T.-Y. Huang, W. Huang, S. V. Huffel, Q. Huo, V. HV, K. M. Iftekharuddin, F. Isensee, M. Islam, A. S. Jackson, S. R. Jambawalikar, A. Jesson, W. Jian, P. Jin, V. J. M. Jose, A. Jungo, B. Kainz, K. Kamnitsas, P.-Y. Kao, A. Karnawat, T. Kellermeier, A. Kermi, K. Keutzer, M. T. Khadir, M. Khened, P. Kickingereder, G. Kim, N. King, H. Knapp, U. Knecht, L. Kohli, D. Kong, X. Kong, S. Koppers, A. Kori, G. Krishnamurthi, E. Krivov, P. Kumar, K. Kushibar, D. Lachinov, T. Lambrou, J. Lee, C. Lee, Y. Lee, M. Lee, S. Lefkovits, L. Lefkovits, J. Levitt, T. Li, H. Li, W. Li, H. Li, X. Li, Y. Li, H. Li, Z. Li, X. Li, Z. Li, X. Li, W. Li, Z.-S. Lin, F. Lin, P. Lio, C. Liu, B. Liu, X. Liu, M. Liu, J. Liu, L. Liu, X. Llado, M. M. Lopez, P. R. Lorenzo, Z. Lu, L. Luo, Z. Luo, J. Ma, K. Ma, T. Mackie, A. Madabushi, I. Mahmoudi, K. H. Maier-Hein, P. Maji, C. Mammen, A. Mang, B. S. Manjunath, M. Marcinkiewicz, S. McDonagh, S. McKenna, R. McKinley, M. Mehl, S. Mehta, R. Mehta, R. Meier, C. Meinel, D. Merhof, C. Meyer, R. Miller, S. Mitra, A. Moiyadi, D. Molina-Garcia, M. A. B. Monteiro, G. Mrukwa, A. Myronenko, J. Nalepa, T. Ngo, D. Nie, H. Ning, C. Niu, N. K. Nuechterlein, E. Oermann, A. Oliveira, D. D. C. Oliveira, A. Oliver, A. F. I. Osman, Y.-N. Ou, S. Ourselin, N. Paragios, M. S. Park, B. Paschke, J. G. Pauloski, K. Pawar, N. Pawlowski, L. Pei, S. Peng, S. M. Pereira, J. Perez-Beteta, V. M. Perez-Garcia, S. Pezold, B. Pham, A. Phophalia, G. Piella, G. N. Pillai, M. Piraud, M. Pisov, A. Popli, M. P. Pound, R. Pourreza, P. Prasanna, V. Prkovska, T. P. Pridmore, S. Puch, Élodie Puybareau, B. Qian, X. Qiao, M. Rajchl, S. Rane, M. Rebsamen, H. Ren, X. Ren, K. Revanuru, M. Rezaei, O. Rippel, L. C. Rivera, C. Robert, B. Rosen, D. Rueckert, M. Safwan, M. Salem, J. Salvi, I. Sanchez, I. Sánchez, H. M. Santos, E. Sartor, D. Schellingerhout, K. Scheufele, M. R. Scott, A. A. Scussel, S. Sedlar, J. P. Serrano-Rubio, N. J. Shah, N. Shah, M. Shaikh, B. U. Shankar, Z. Shboul, H. Shen, D. Shen, L. Shen, H. Shen, V. Shenoy, F. Shi, H. E. Shin, H. Shu, D. Sima, M. Sinclair, O. Smedby, J. M. Snyder, M. Soltaninejad, G. Song, M. Soni, J. Stawiaski, S. Subramanian, L. Sun, R. Sun, J. Sun, K. Sun, Y. Sun, G. Sun, S. Sun, Y. R. Suter, L. Szilagyi, S. Talbar, D. Tao, D. Tao, Z. Teng, S. Thakur, M. H. Thakur, S. Tharakan, P. Tiwari, G. Tochon, T. Tran, Y. M. Tsai, K.-L. Tseng, T. A. Tuan, V. Turlapov, N. Tustison, M. Vakalopoulou, S. Valverde, R. Vanguri, E. Vasiliev, J. Ventura, L. Vera, T. Vercauteren, C. A. Verrastro, L. Vidyaratne, V. Vilaplana, A. Vivekanandan, G. Wang, Q. Wang, C. J. Wang,

W. Wang, D. Wang, R. Wang, Y. Wang, C. Wang, G. Wang, N. Wen, X. Wen, L. Weninger, W. Wick, S. Wu, Q. Wu, Y. Wu, Y. Xia, Y. Xu, X. Xu, P. Xu, T.-L. Yang, X. Yang, H.-Y. Yang, J. Yang, H. Yang, G. Yang, H. Yao, X. Ye, C. Yin, B. Young-Moxon, J. Yu, X. Yue, S. Zhang, A. Zhang, K. Zhang, X. Zhang, L. Zhang, X. Zhang, Y. Zhang, L. Zhang, J. Zhang, X. Zhang, T. Zhang, S. Zhao, Y. Zhao, X. Zhao, L. Zhao, Y. Zheng, L. Zhong, C. Zhou, X. Zhou, F. Zhou, H. Zhu, J. Zhu, Y. Zhuge, W. Zong, J. Kalpathy-Cramer, K. Farahani, C. Davatzikos, K. van Leemput, and B. Menze. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge, 2019.

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

[9] V. Baliyan, C. J. Das, R. Sharma, and A. K. Gupta. Diffusion weighted imaging: Technique and applications. *World J. Radiol.*, 8(9):785–798, Sept. 2016.

[10] A. V. Belikov. The number of key carcinogenic events can be predicted from cancer incidence. *Sci. Rep.*, 7(1):12170, Sept. 2017.

[11] A. V. Belikov, A. Vyatkin, and S. V. Leonov. The erlang distribution approximates the age distribution of incidence of childhood and young adulthood cancers. *PeerJ*, 9(e11976):e11976, Aug. 2021.

[12] S. Bionetworks. Brats 2022 announcement. `https://www.synapse.org/#!Synapse:syn27046444/wiki/616992`, 2022. (Accessed: 2023-01-16).

[13] C. M. Bishop. *Pattern recognition and machine learning*. Springer, New York, 2006.

[14] A. Bonhage, M. Eltaher, T. Raab, M. Breuß, A. Raab, and A. Schneider. A modified mask region-based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the north german lowland. *Archaeological Prospection*, 28(2):177–186, 2021.

[15] M. Brant-Zawadzki, D. Atkinson, M. Detrick, W. G. Bradley, and G. Scidmore. Fluid-attenuated inversion recovery (flair) for assessment of cerebral infarction. *Stroke*, 27(7):1187–1191, 1996.

[16] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 10 2001.

[17] S. Brenner, S. Hartzendorf, P. Vogt, E. Maier, N. Etminan, E. Jung, W. Wick, F. Sahm, F. Winkler, and M. Ratliff. Progression patterns in non-contrast-enhancing gliomas support brain tumor responsiveness to surgical lesions. *Pathol. Oncol. Res.*, 28:1610268, May 2022.

[18] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, page 93–104, New York, NY, USA, 2000. Association for Computing Machinery.

[19] G. Brix, H. Kolem, W. R. Nitz, M. Bock, A. Huppertz, C. J. Zech, and O. Dietrich. Basics of magnetic resonance imaging and magnetic resonance spectroscopy. In *Magnetic Resonance Tomography*, pages 3–167. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[20] N. Bulakbasi, M. Kocaoglu, F. Ors, C. Tayfun, and T. Uçöz. Combination of single-voxel proton MR spectroscopy and apparent diffusion coefficient calculation in the evaluation of common brain tumors. *AJNR Am. J. Neuroradiol.*, 24(2):225–233, Feb. 2003.

[21] S. Cai, J. Gao, M. Zhang, W. Wang, G. Chen, and B. C. Ooi. Effective and efficient dropout for deep convolutional neural networks. *CoRR*, abs/1904.03392, 2019.

[22] M. Castillo, J. K. Smith, L. Kwock, and K. Wilber. Apparent diffusion coefficients in the evaluation of high-grade cerebral gliomas. *AJNR Am. J. Neuroradiol.*, 22(1):60–64, Jan. 2001.

[23] CBICA Center for Biomedical Image Computing & Analytics. RSNA-ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge 2021. `http://braintumorsegmentation.org/`, 2021. (Accessed: 2023-10-10).

[24] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv e-prints*, page arXiv:1606.06650, Jun 2016.

[25] S. Cha. Update on brain tumor imaging: from anatomy to physiology. *AJNR Am. J. Neuroradiol.*, 27(3):475–487, Mar. 2006.

[26] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert. Deep learning for cardiac image segmentation: A review. *Frontiers in Cardiovascular Medicine*, 7, 2020.

[27] G. Chen, Q. Li, F. Shi, I. Rekik, and Z. Pan. Rfdcr: Automated brain lesion segmentation using cascaded random forests with dense conditional random fields. *NeuroImage*, 211:116620, 2020.

[28] X. Chen, M. Young, M. Kimmel, and O. Gorlova. Modeling the natural history and detection of lung cancer based on smoking behavior. *PloS one*, 9:e93430, 04 2014.

[29] T. S. community. Scipy v1.11.4 manual. `https://docs.scipy.org/doc/scipy/tutorial/ndimage.html`, 2023. (Accessed: 2023-12-01).

[30] B. Dey, E. Dehaerne, K. Khalil, S. Halder, P. Leray, and M. A. Bayoumi. Deep learning based defect classification and detection in sem images: A mask r-cnn approach, 2022.

[31] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[32] A. Dubey, R. Kataria, and V. Sinha. Role of diffusion tensor imaging in brain tumor surgery. *Asian J. Neurosurg.*, 13(02):302–306, June 2018.

[33] M.-P. Dubuisson and A. Jain. A modified hausdorff distance for object matching. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 1, pages 566–568 vol.1, 1994.

[34] B. M. Ellingson, P. Y. Wen, and T. F. Cloughesy. Modified criteria for radiographic response assessment in glioblastoma clinical trials. *Neurotherapeutics*, 14(2):307–320, Apr. 2017.

[35] M. Essig, M. S. Shiroishi, T. B. Nguyen, M. Saake, J. M. Provenzale, D. Enterline, N. Anzalone, A. Dörfler, A. Rovira, M. Wintermark, and M. Law. Perfusion MRI: the five most frequently asked technical questions. *AJR Am. J. Roentgenol.*, 200(1):24–34, Jan. 2013.

[36] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.

[37] M. Futrega, A. Milesi, M. Marcinkiewicz, and P. Ribalta. Optimized u-net for brain tumor segmentation, 2021.

[38] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation, 2017.

[39] German Cancer Research Center (DKFZ), Division of Medical Image Computing. The Medical Imaging Interaction Toolkit (MITK). `https://www.mitk.org/wiki/The_Medical_Imaging_Interaction_Toolkit_(MITK)`, 2021. (Accessed: 2024-01-04).

[40] R. Gonzalez and R. Woods. *Digital Image Processing Global Edition*. Pearson Deutschland, 2017.

[41] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[42] V. P. B. Grover, J. M. Tognarelli, M. M. E. Crossey, I. J. Cox, S. D. Taylor-Robinson, and M. J. W. McPhail. Magnetic resonance imaging: Principles and techniques: Lessons for clinicians. *J. Clin. Exp. Hepatol.*, 5(3):246–255, Sept. 2015.

[43] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan. Sa-unet: Spatial attention u-net for retinal vessel segmentation, 2020.

[44] N. Gyori, J. Hutter, V. Nath, M. Palombo, M. Pizzolato, and F. Zhang. *Computational Diffusion MRI, International MICCAI Workshop, Lima, Peru, October 2020*. Springer, 2021.

[45] J. V. Hajnal, D. J. Bryant, L. Kasuboski, P. M. Pattany, B. De Coene, P. D. Lewis, J. M. Pennock, A. Oatridge, I. R. Young, and G. M. Bydder. Use of fluid attenuated inversion recovery (FLAIR) pulse sequences in MRI of the brain. *J. Comput. Assist. Tomogr.*, 16(6):841–844, Nov. 1992.

[46] R. Haralick, S. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-9:532 – 550, 08 1987.

[47] F. Hausdorff. Grundzüge der Mengenlehre. 1914.

[48] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN, 2018.

[49] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[50] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks, 2016.

[51] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[52] J. Ho, N. Kalchbrenner, D. Weissenborn, and T. Salimans. Axial attention in multidimensional transformers, 2019.

[53] H. Huang, W. Zhang, Y. Fang, J. Hong, S. Su, and X. Lai. Overall survival prediction for gliomas using a novel compound approach. *Frontiers in Oncology*, 11, 2021.

[54] Z. Huang, X. Wang, Y. Wei, L. Huang, H. Shi, W. Liu, and T. S. Huang. Ccnet: Criss-cross attention for semantic segmentation, 2020.

[55] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[56] F. Isensee, P. F. Jaeger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein. nnu-net for brain tumor segmentation, 2020.

[57] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein. No New-Net. *arXiv e-prints*, page arXiv:1809.10483, Sep 2018.

[58] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Köhler, T. Norajitra, S. J. Wirkert, and K. H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *CoRR*, abs/1809.10486, 2018.

[59] Z. Jiang, C. Ding, M. Liu, and D. Tao. *Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task*, pages 231–241. 05 2020.

[60] R. Kaur, H. GholamHosseini, R. Sinha, and M. Lindén. Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images. *BMC Medical Imaging*, 22(1):103, May 2022.

[61] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2022.

[62] E. A. Knopp, S. Cha, G. Johnson, A. Mazumdar, J. G. Golfinos, D. Zagzag, D. C. Miller, P. J. Kelly, and I. I. Kricheff. Glial neoplasms: dynamic contrast-enhanced t2*-weighted MR imaging. *Radiology*, 211(3):791–798, June 1999.

[63] D. Lachinov, E. Vasiliev, and V. Turlapov. Glioma Segmentation with Cascaded Unet. *arXiv e-prints*, page arXiv:1810.04008, Oct 2018.

[64] C. Laiton-Bonadiez, G. Sanchez-Torres, and J. Branch-Bedoya. Deep 3D neural network for brain structures segmentation using self-attention modules in MRI images. *Sensors (Basel)*, 22(7):2559, Mar. 2022.

[65] A. Lasocki and F. Gaillard. Non-contrast-enhancing tumor: A new frontier in glioblastoma research. *AJNR Am. J. Neuroradiol.*, 40(5):758–765, May 2019.

[66] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[67] Z. Li and J. Chen. Superpixel segmentation using linear spectral clustering. pages 1356–1363, 06 2015.

[68] C.-Y. Lin, C.-C. Chang, P.-L. Su, C.-C. Lin, Y.-L. Tseng, W.-C. Su, and Y.-T. Yen. Brain mri imaging characteristics predict treatment response and outcome in patients with de novo brain metastasis of egfr-mutated nsclc. *Medicine*, 98:e16766, 08 2019.

[69] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[70] Z. Liu, J. Wang, J. Li, P. Liu, and K. Ren. A novel improved mask rcnn for multiple targets detection in the indoor complex scenes, 2023.

[71] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation, 2015.

[72] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation, 2015.

[73] W. Lorensen and H. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21:163–, 08 1987.

[74] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison. The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.*, 131(6):803–820, June 2016.

[75] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. K. Ng, S. M. Pfister, G. Reifenberger, R. Soffietti, A. von Deimling, and D. W. Ellison. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro. Oncol.*, 23(8):1231–1251, Aug. 2021.

[76] R. P. Ltd. Radiopaedia: Brain tumor imaging. `https://radiopaedia.org/`, 2023. (Accessed: 2023-11-21).

[77] F. Lugauer, J. Wetzl, A. Maier, S. Steidl, V. Christlein, and J. Hornegger. *Magnetic Resonance Imaging*, pages 91–118. Springer International Publishing, Cham, 2018.

[78] H. M. Luu and S.-H. Park. Extending nn-unet for brain tumor segmentation, 2021.

[79] P. C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55, 1936.

[80] A. Maier, S. Steidl, and V. Christlein. *Medical Imaging Systems: An Introductory Guide*. Springer, 2018.

[81] M. Maier. Quantitative mri of the brain—measuring changes caused by disease. *Journal of Neurology Neurosurgery & Psychiatry*, 75, 10 2004.

[82] T. Martín-Noguerol, S. Mohan, E. Santos-Armentia, A. Cabrera-Zubizarreta, and A. Luna. Advanced MRI assessment of non-enhancing peritumoral signal abnormality in brain lesions. *Eur. J. Radiol.*, 143(109900):109900, Oct. 2021.

[83] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.

[84] E. Michael, H. Ma, H. Li, F. Kulwa, and J. Li. Breast cancer segmentation methods: Current status and future potentials. *Biomed Res. Int.*, 2021:9962109, July 2021.

[85] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016.

[86] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. Image segmentation using deep learning: A survey. *CoRR*, abs/2001.05566, 2020.

[87] A. Myronenko. 3D MRI brain tumor segmentation using autoencoder regularization. *arXiv e-prints*, page arXiv:1810.11654, Oct 2018.

[88] A. Myronenko and A. Hatamizadeh. Robust semantic segmentation of brain tumor regions from 3d mris, 2020.

[89] S. A. J. Naqvi, A. Tauqeer, R. Bhatti, and S. B. Ali. Improved lung segmentation based on u-net architecture and morphological operations, 2022.

[90] M. Neema, Z. Guss, J. Stankiewicz, A. Arora, B. Healy, and R. Bakshi. Normal findings on brain fluid-attenuated inversion recovery mr images at 3t. *American Journal of Neuroradiology*, 30(5):911–916, 2009.

[91] T. Oh, S. Park, and B.-J. Yang. Theory of transverse magnetization in spin-orbit coupled antiferromagnets, 2022.

[92] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.

[93] A. G. Osborn, G. L. Hedlund, and K. L. Salzman. *Osborn's brain : imaging pathology and anatomy (Second)*. Elsevier, 2018.

[94] A. R. Padhani, G. Liu, D. M. Koh, T. L. Chenevert, H. C. Thoeny, T. Takahara, A. Dzik-Jurasz, B. D. Ross, M. Van Cauteren, D. Collins, D. A. Hammoud, G. J. S. Rustin, B. Taouli, and P. L. Choyke. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. *Neoplasia*, 11(2):102–125, Feb. 2009.

[95] J. Polzehl. *Magnetic Resonance Brain Imaging*. Springer International Publishing, 2023.

[96] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016.

[97] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints*, page arXiv:1505.04597, May 2015.

[98] R. Savjani. NnU-net: Further automating biomedical image autosegmentation. *Radiol. Imaging Cancer*, 3(1):e209039, Jan. 2021.

[99] P. Scheikl, S. Laschewski, A. Kisilenko, T. Davitashvili, B. Müller, M. Capek, B. Müller, M. Wagner, and F. Mathis-Ullrich. Deep learning for semantic segmentation of organs and tissues in laparoscopic surgery. *Current Directions in Biomedical Engineering*, 6:20200016, 09 2020.

[100] S. S. Sengara, C. Meulengrachtb, M. P. Boesenb, A. F. Overgaardb, H. Gudbergsenb, J. D. Nybingb, and E. B. Dam. Unet architectures in multiplanar volumetric segmentation – validated on three knee mri cohorts, 2022.

[101] J. P. F. Serra. Image analysis and mathematical morphology. 1983.

[102] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[103] S. Shekhar, R. Nandi, and H. S. Kamath. Unet based pipeline for lung segmentation from chest x-ray images, 2022.

[104] J. Shotton, J. M. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 81:2–23, 2007.

[105] M. M. R. Siddiquee and A. Myronenko. Redundancy reduction in semantic segmentation of 3d brain tumor mris, 2021.

[106] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer Berlin Heidelberg, 2013.

[107] M. Soltaninejad, L. Zhang, T. Lambrou, G. Yang, N. M. Allinson, and X. Ye. MRI brain tumor segmentation using random forests and fully convolutional networks. *CoRR*, abs/1909.06337, 2019.

[108] T. Sørensen, T. Sørensen, T. Biering-Sørensen, T. Sørensen, and J. T. Sorensen. A method of establishing group of equal amplitude in plant sociobiology based on similarity of species content and its application to analyses of the vegetation on danish commons. 1948.

[109] E. Stejskal and J. Tanner. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. 42(1):288–292, 1965.

[110] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks, 2014.

[111] A. A. Taha and A. Hanbury. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med. Imaging*, 15(1):29, Aug. 2015.

[112] H. Tariq and S. Burney. Brain mri literature review for interdisciplinary studies. *Journal of biomedical graphics and computing*, 4:41–53, 10 2014.

[113] H. Tariq and S. Burney. Low level segmentation of brain mr slices and quantification challenges. 03 2020.

[114] T. S.-I. D. Team. Scikit-image, image processing in python, documentation. `https://scikit-image.org/docs/stable/auto_examples/applications/plot_morphology.html`, 2023. (Accessed: 2023-12-01).

[115] J. W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5(2):99–114, 1949.

[116] J. W. van der Graaf, M. L. van Hooff, C. F. M. Buckens, M. Rutten, J. L. C. van Susante, R. J. Kroeze, M. de Kleuver, B. van Ginneken, and N. Lessmann. Lumbar spine segmentation in mr images: a dataset and a public benchmark, 2023.

[117] J. Walsh, A. Othmani, M. Jain, and S. Dev. Using u-net network for efficient brain tumor segmentation in mri images, 2022.

[118] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation, 2020.

[119] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi. Medical image segmentation using deep learning: A survey. *IET Image Processing*, 16(5):1243–1267, jan 2022.

[120] B. Xu, W. Wang, G. Falzon, P. Kwan, L. Guo, G. Chen, A. Tait, and D. Schneider. Automated cattle counting using mask r-cnn in quadcopter vision system. *Computers and Electronics in Agriculture*, 171:105300, 2020.

[121] A. Y. Yuan, Y. Gao, L. Peng, L. Zhou, J. Liu, S. Zhu, and W. Song. Hybrid deep learning network for vascular segmentation in photoacoustic imaging. *Biomed. Opt. Express*, 11(11):6445–6457, Nov. 2020.

[122] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. C. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells, III, F. A. Jolesz, and R. Kikinis. Statistical validation of image segmentation quality based on a spatial overlap index1. *Acad. Radiol.*, 11(2):178–189, Feb. 2004.