ML PROJECT(CACSC17)

3D SEMANTIC SEGMENTATION OF BRAIN METASTASES

**Submitted by**

Twinkle Dhingra(2022UCA1918)

Vishwangi Raj(2022UCA1908)

## **INTRODUCTION**

Semantic segmentation is a computer vision task that involves classifying each pixel in an image into a specific category or class. This technique is crucial in applications where understanding the structure and layout of a scene is essential, such as autonomous driving, medical imaging, and satellite imagery analysis.

Semantic segmentation for brain tumor analysis involves using machine learning techniques to segment or classify each pixel of a brain image to identify the presence, location, and structure of a tumor.

Metastasis refers to the process by which cancer cells spread from their original (primary) site to other parts of the body, forming new (secondary) tumors in organs or tissues beyond the initial location. This process is a hallmark of advanced cancer and is often associated with more challenging treatment and management.

**METASTASIS OCCURENCE**

1. **Invasion**: Cancer cells first invade nearby normal tissue surrounding the primary tumor. These cells acquire the ability to break through the barriers that typically confine them within the tissue of origin.

2. **Intravasation** : Cancer cells enter the bloodstream or lymphatic system. This phase involves cells infiltrating blood vessels or lymph nodes, where they can travel to distant sites in the body.

3. **Circulation**: Once inside the circulatory systems, cancer cells move through the body, where they may evade immune detection and survive under conditions unlike their original environment.

4. **Extravasation**: Cancer cells exit the bloodstream or lymphatic system and invade new tissues. They usually attach to blood vessel walls at distant sites and then migrate through the vessel walls into surrounding tissue.

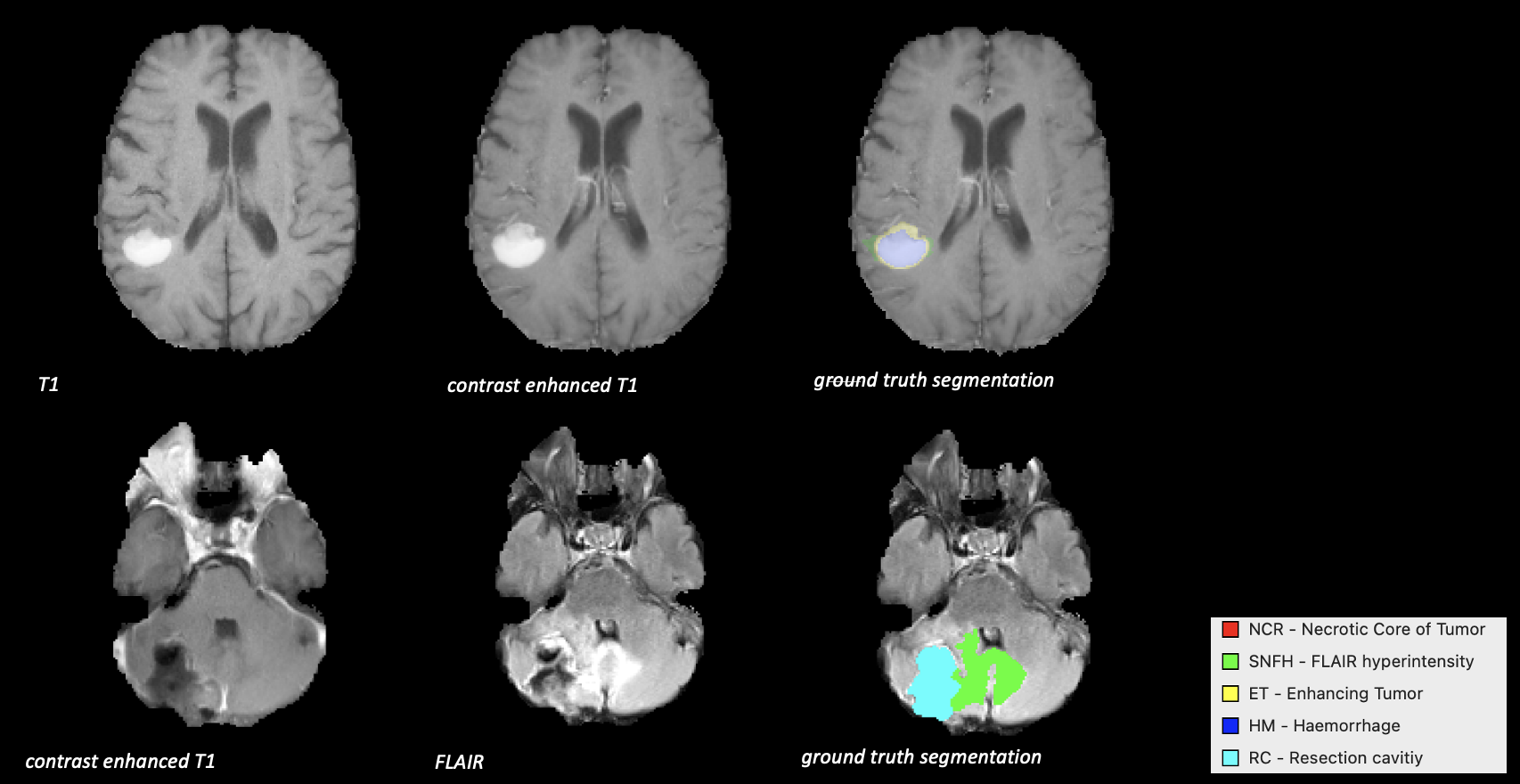
**USES OF SEMANTIC SEGMENTATION OF BRAIN TUMOR**

Metastatic tumors in the brain are often small, irregular, and may be distributed across multiple regions. Semantic segmentation allows clinicians to precisely identify these areas, distinguishing metastasis from surrounding healthy brain tissue.

Detailed segmentation helps determine the exact size, shape, and location of brain metastases, which is critical in planning surgeries, radiation therapy, or targeted therapies.

Semantic segmentation allows for consistent and quantifiable tracking of brain metastases over time. By segmenting and measuring tumors before and after treatment, clinicians can better understand how well a therapy is working.

## **DATASET SPECIFICATIONS**



The BraTS 2024 Metastasis dataset along with the NYUMets dataset taht we have used consists of 328 image folders, each containing five images. The images are labeled with specific suffixes:

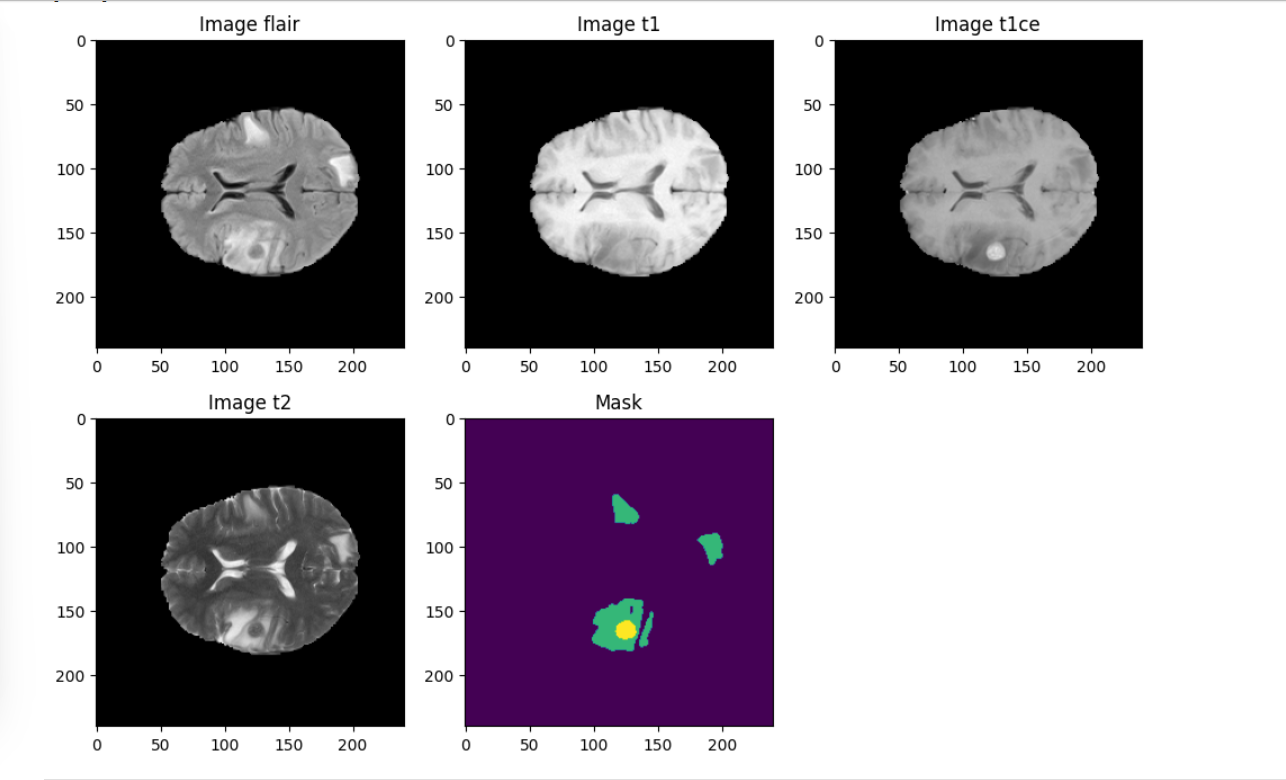
1. **post-contrast T1-weighted (T1W)**: This is the T1-weighted contrast-enhanced MRI image, providing high anatomical detail and highlighting areas where contrast agents accumulate (often indicating active tumor regions). These images show the anatomy of the brain clearly, with bright areas representing fat and some types of tissue. T1-weighted images provide a good view of the structure of the brain but may not show fluids or certain abnormalities as clearly as other types.

2. **pre-contrast T1-weighted (T1N)** : The T1-weighted non-contrast MRI image, useful for viewing anatomical structures without contrast enhancement.

3. **T2-weighted (T2w)**: The T2-weighted MRI image, helpful in detecting fluid accumulation and edema around the tumor, offering a broader view of affected brain tissue. In these images, fluids (like cerebrospinal fluid) appear bright, while solid tissues are darker. T2-weighted images are especially useful for spotting areas of swelling, tumors, or inflammation in the brain since these areas often contain extra fluid.

4. **T2-weighted FLAIR (T2W)** : T2-weighted fluid-attenuated inversion recovery (FLAIR) images, which suppress fluids in normal brain regions, accentuating abnormalities like edema and lesions near the tumor. This type of image is a T2-weighted scan with some adjustments to suppress fluid signals. By removing the bright areas from fluid, FLAIR makes it easier to see abnormal areas like lesions, swelling, or other irregularities in the brain tissue. It's particularly useful for spotting conditions like brain tumors or multiple sclerosis.

5. **Segmentation Mask** : The segmentation mask, which labels specific areas of the brain in different classes, identifying the tumor's presence and boundaries. In brain segmentation, the mask outlines parts of the brain, like tumors, in each scan. Each mask pixel is usually labeled by numbers to show different regions, with one number representing the background (non-tumor areas) and others indicating specific types or areas of the tumor. These labels help us know exactly where the regions of interest are.



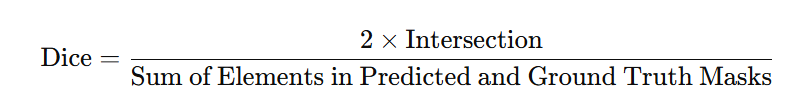
## 

## **METHODOLOGY**

### **Custom Loss Function**

Our loss function approach combines **Weighted Dice Loss** and **Categorical Focal Loss**, both highly suited for medical image segmentation tasks like brain metastasis due to their attention to class imbalance and difficult-to-classify regions.

**1. Weighted Dice Loss**

****

The Dice Loss measures the overlap between predicted and ground truth segmentation, which is useful in scenarios where class imbalance (e.g., small tumor regions versus large background areas) is common.

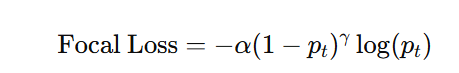
It is a metric used to gauge the similarity between two sets. It is widely used in segmentation tasks and ranges from 0 to 1, with 1 indicating perfect overlap and 0 indicating no overlap.

By assigning equal weights (wt0 = wt1 = wt2 = wt3 = 0.25) to each class, the Weighted Dice Loss ensures that each class, including smaller or less represented tumor regions, contributes equally to the learning process.

This weighted approach enhances segmentation accuracy by reducing the impact of class imbalance, which is particularly beneficial in medical imaging tasks where certain regions, like tumors, occupy much less space compared to background tissue.

### **2. Categorical Focal Loss**

The Focal Loss is designed to focus on challenging samples by down-weighting easy-to-classify voxels and giving more importance to voxels that are harder to classify.



where pt​ is the model's predicted probability for the true class, α is a balancing factor for the class, and γ is a focusing parameter to reduce the loss from well-classified examples.

By focusing more on difficult-to-classify voxels, this loss function is well-suited for detecting subtle anatomical variations in the brain, such as small metastasis regions that are often hard to capture.

This helps improve model sensitivity to less prominent or ambiguous areas, ensuring that the model becomes better at segmenting regions that could be easily misclassified as background or normal tissue.

### **3. Combined Loss Function**

This setup is particularly powerful for brain metastasis segmentation, where minor discrepancies can significantly impact clinical decisions. By leveraging both loss functions, the model optimizes for both high accuracy and enhanced detection of subtle anatomical variations.

### **DATASET PROCESSING**

The images from different modalities (e.g., T1, T1 contrast-enhanced, T2, and FLAIR) and the corresponding segmentation masks are loaded from a specified directory. These images are stored in medical image formats such as `.nii.gz` and are read using a medical imaging library like `nibabel`, which converts the images into NumPy arrays for further processing.The dataset is present in .nii format or NIfTI files. NIfTI stands for **Neuroimaging Informatics Technology Initiative**.

1. **Scaling the Images**

Each image is normalized to standardize the pixel values across different modalities. This is done by reshaping the image data into a 2D array where one dimension corresponds to all pixels and the other to the channels (if applicable). The pixel values are then scaled using a StandardScaler ensuring the images have zero mean and unit variance. After scaling, the images are reshaped back to their original dimensions.

1. **Processing the Segmentation Masks**

The segmentation masks are processed to adjust the class labels. The mask may contain multiple label values corresponding to different tissue types or background. Any unwanted or redundant label values are reassigned to a predefined label set to standardize the mask and align with the model's expectations.

1. **Cropping the Images**

To ensure the image dimensions are suitable for training (divisible by 64 for patch extraction), the images are cropped to a fixed size. This cropping process typically focuses on the central region of the image, ensuring that the dimensions of the image and mask are consistent. The cropped images are often resized to a smaller, uniform shape suitable for the model’s input layer.

1. **Random Slice Selection**

For volumetric data (3D images), a random slice from the third dimension (e.g., the z-axis) is selected to introduce variability. This random selection helps create a more diverse training set, simulating different perspectives of the same anatomical structures.

1. **Class Distribution Check**

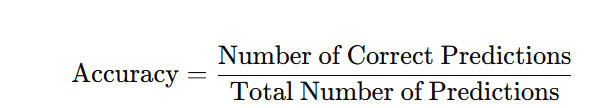
The class distribution of the mask is examined to ensure that the data contains a sufficient proportion of useful labels (i.e., labels that are not background). A threshold is set ( at least 1% of the data should contain non-background labels) to filter out instances with too much background or irrelevant data. If the data meets the threshold, it is considered "useful" and proceeds to the next step.

1. **Saving the Preprocessed Data**

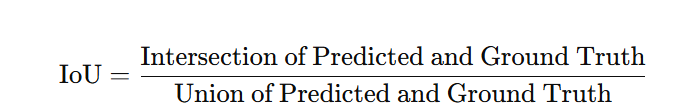
Once the data is deemed useful, it is saved as files for later use in model training. The images from different modalities are stacked together to form a multi-channel input tensor, while the masks are typically one-hot encoded (for multi-class segmentation tasks). These processed images and masks are saved in a format (e.g., `.npy` files) that can be efficiently loaded during training.

### **PERFORMANCE METRICS**

1. **Accuracy:** It calculates the proportion of correctly predicted pixels (or voxels in the case of 3D data) out of all the predictions made.



1. **IoU Score :**The **IoU score** is a more specific metric for evaluating segmentation tasks, particularly in multi-class segmentation. It measures the overlap between the predicted segmentation mask and the true mask.

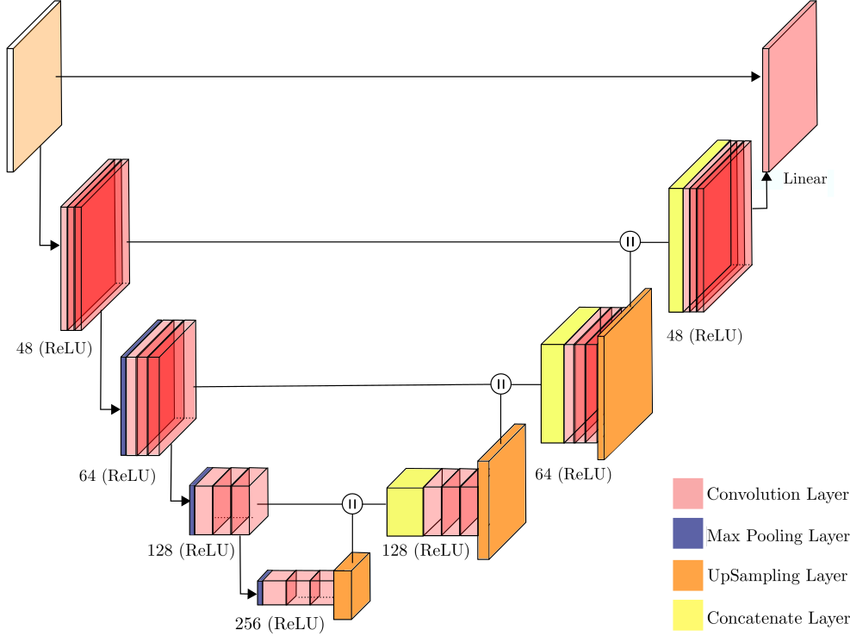
​

It calculates the ratio of the area where the predicted mask and the ground truth mask overlap (intersection) to the total area covered by either mask (union).

**In Context**: The IoU score is useful for assessing how well the model identifies the regions of interest in the segmentation task. For multi-class segmentation, the IoU score is often calculated per class, and a higher IoU indicates better segmentation performance. A threshold (e.g., 0.5) is typically set to determine whether a prediction is considered correct.

### ***3D U-Net Model***

**3D U-Net Model Training for Medical Image Segmentation**

****

The U-Net model has an **encoder-decoder** structure, resembling a "U" shape. The **encoder path** compresses the input into compact feature maps, while the **decoder path** gradually reconstructs the segmentation mask to the original resolution. Skip connections link corresponding layers in the encoder and decoder paths to retain fine-grained spatial information.

**1. Objective:**

* This model is designed for precise 3D segmentation of medical imaging data (e.g., brain MRI) across four classes. The segmentation assists in identifying various anatomical or pathological regions in a 3D space.

**2. Architecture:**

* The model is a 3D U-Net that accepts volumetric data with dimensions 128 x 128 x 128 and three channels per voxel, making it well-suited for spatially-consistent 3D imaging datasets.
* The output is a 4-class segmented image, allowing the model to classify each voxel into one of four categories.

**3. Optimization & Metrics:**

* Adam Optimizer is used with a learning rate of LR = 0.0001, providing efficient convergence.
* Metrics: Accuracy and Intersection over Union (IoU) are used to monitor model performance and assess segmentation accuracy.

**4. Training Pipeline**:

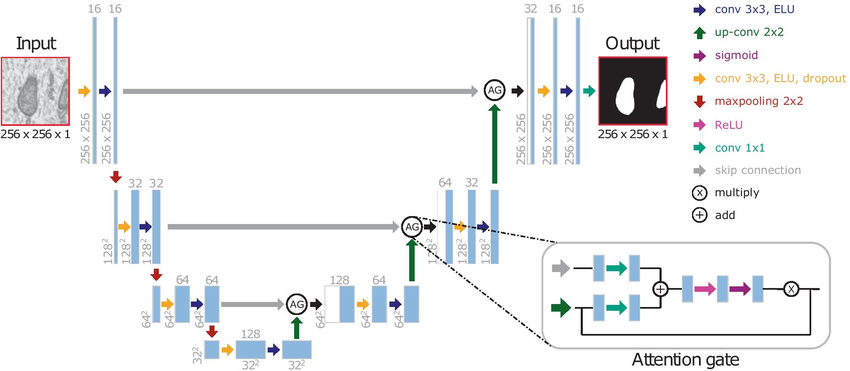
* Data Loading: imageLoader function is employed to feed batches of training and validation images and corresponding masks to the model.
* Batch Processing: The model processes the training data in batches, with steps\_per\_epoch and validation\_steps derived based on the dataset size and batch size.
* Training Execution: The model is trained over 50 epochs, with validation at each epoch to monitor overfitting and generalization.

**5. Model Saving:**

* The trained model is saved for future inference or fine-tuning. This model can be reloaded to segment 3D images effectively.

### ***Attention U-net Model***

**Attention U-Net for Enhanced Medical Image Segmentation**

****

**Attention** in the context of deep learning, particularly in tasks like image segmentation, refers to a mechanism that allows the model to focus more on relevant parts of the input while ignoring less relevant parts. This is especially useful for tasks where specific regions of an image are more important for making accurate predictions, such as in medical image segmentation. Attention mechanisms assign different weights to different parts of the input, effectively highlighting regions that are most likely to contain the object or feature of interest.

**Attention U-Net** builds on the original U-Net architecture by adding attention gates to its layers. The U-Net is popular in segmentation tasks for its encoder-decoder structure with skip connections. Attention U-Net enhances this by integrating attention mechanisms to help the model focus more on the relevant regions in the image.

**1. Objective:**

* Like the initial model, this 3D Attention U-Net is built for 3D medical image segmentation to distinguish between four different classes. However, it incorporates attention mechanisms for improved focus on relevant anatomical structures.

2. **Key Differences from the Previous Model:**

* Attention Mechanism: The addition of attention gates in the expansive path enables the model to selectively focus on significant features, suppressing less relevant information. This improvement helps the model better capture subtle boundaries and regions, especially valuable in complex medical data where small structures matter.

* Enhanced Feature Selection: The attention gates interact with the skip connections in the U-Net structure, refining the feature maps passed from the contracting to the expansive path. This targeted refinement improves segmentation accuracy for challenging regions compared to the standard U-Net, which does not have this focused attention mechanism.

3. **Architecture**:

* 3D Attention U-Net Structure: The architecture is a U-Net with an encoder-decoder design, incorporating attention blocks at each up-sampling layer in the expansive path. Input size remains 128 x 128 x 128 with three channels, and output size is a 4-class segmented volume.
* Standard 3D U-Net: In contrast, the previous model lacked attention gates, meaning all skip connection features were included equally in the expansive path, without emphasis on the most relevant features.

4. **Improved Precision in Segmentation:**

* The attention model’s selective focus on specific regions enhances its ability to distinguish between closely neighboring structures or fine boundaries, particularly in medical applications with complex or subtle features. This increased precision leads to a more robust segmentation than the standard 3D U-Net.

5. **Training Process & Performance:**

* Training parameters, batch size, and learning rate remain consistent across both models to allow for fair performance comparisons.
* With attention gates, the model is expected to converge on a more accurate solution, especially as observed in the IoU and accuracy metrics.

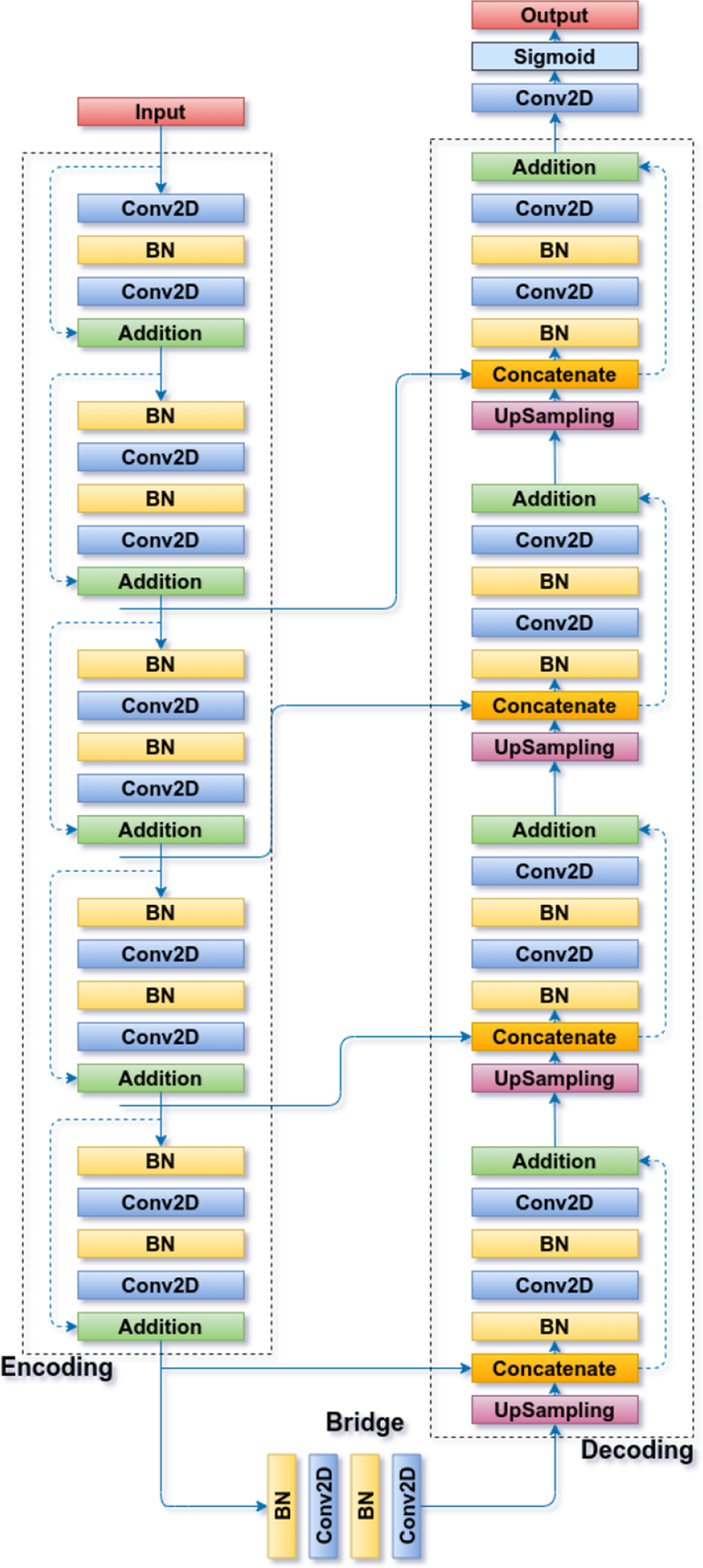
6. **Model Saving:**

* The 3D Attention U-Net model is saved as `ml2-brats\_attention\_unet\_3d.keras`, enabling future reuse with enhanced segmentation capabilities over the standard U-Net model.

### ***Res-U-Net Model***

**3D ResUNet for Medical Image Segmentation**

### 



**ResUNet** is a variant of the U-Net architecture that incorporates **Residual Connections** (from ResNet) into the standard U-Net model. The integration of residual connections helps improve training by addressing issues like the vanishing gradient problem and enables better feature reuse across layers.

In deep networks, during training, the distribution of activations (outputs of neurons) can change as the parameters (weights) of the earlier layers are updated. This can slow down training because each layer has to constantly adapt to these changing distributions.

Batch normalization reduces this shift by ensuring that the inputs to each layer remain more stable throughout training.

**1. Objective:**

* ResUNet is developed to segment complex 3D medical images more effectively by utilizing residual learning. The goal is to improve segmentation accuracy, particularly in cases where finer anatomical details are critical.

**2. Key Differences from Previous Models:**

* Incorporation of Residual Blocks:
* ResUNet includes residual blocks instead of regular convolutional layers, distinguishing it from the standard U-Net and Attention U-Net. These residual blocks use skip connections within each layer, enabling the model to "skip" certain layers and pass information more effectively through the network.
* This structure helps to mitigate the vanishing gradient problem, allowing ResUNet to train deeper and capture more detailed features than a standard U-Net without compromising training stability.
* Enhanced Gradient Flow:
* Unlike previous models, the residual connections in ResUNet allow gradients to backpropagate directly through the network, which improves convergence and helps the model focus on subtle details in medical images, leading to more accurate segmentation.
* Efficiency in Deep Feature Learning:
* Residual learning helps ResUNet avoid the degradation problem common in deep networks, where adding more layers can sometimes degrade performance. This characteristic makes ResUNet more effective than U-Net variants without residuals, particularly for handling complex 3D structures in medical data.

**3. Architecture:**

* 3D ResUNet Structure:
* The ResUNet maintains the traditional U-Net encoder-decoder structure but replaces standard convolutional layers with residual blocks in both the contracting and expansive paths.
* Each residual block performs two 3D convolutions with batch normalization and ReLU activation, followed by a shortcut connection that directly adds the input to the block’s output.
* Expansive Path with Residual Skip Connections:
* Similar to previous U-Net architectures, skip connections from the encoder to decoder are present, but in ResUNet, these are residual in nature, further enhancing feature reuse and enabling the model to learn intricate features.

4. **Loss Function and Optimization:**

* ResUNet employs a combination of Dice Loss with Categorical Focal Loss, which addresses class imbalance and ensures that challenging regions are segmented effectively. The model is trained with Adam optimizer and a low learning rate for stable convergence.

5. **Expected Advantages of ResUNet:**

* The residual structure in ResUNet is expected to provide better segmentation of fine anatomical structures and improved robustness over standard U-Net and Attention U-Net.
* Residual learning enables more precise segmentation, especially in deep 3D imaging data where detailed region distinction is essential.

6. **Model Saving:**

* The ResUNet model is saved providing an optimized model for detailed and reliable 3D medical image segmentation.

**MODULES USED**

1. **Keras Layers and Model Utilities**

* **Input**: Defines the input layer for the model, specifying image dimensions.
* **Conv3D**: Performs 3D convolution operations, essential for learning spatial features in 3D medical images.
* **MaxPooling3D**: Downsamples the spatial dimensions, reducing the feature map size and adding translational invariance.
* **Conv3DTranspose:** A 3D transposed convolution (upsampling) operation used in the expansive path to increase feature map size.
* **BatchNormalization:** Normalizes activations across batches, which helps with training stability and convergence.
* **Dropout:** Prevents overfitting by randomly setting a fraction of the input units to zero during training.
* **Add**: Adds two tensors element-wise, which is crucial for residual connections in ResUNet.
* **Activation**: Applies an activation function (e.g., ReLU) to introduce non-linearity in the network.
* **Lambda**: Wraps custom operations (when necessary) in a layer, enabling more flexible model design.

2.**Keras Model**

* Model: Combines inputs, outputs, and layers into a fully functional neural network model that can be trained, evaluated, and saved.

3. **Tifffile (from tifffile import imsave, imwrite)**

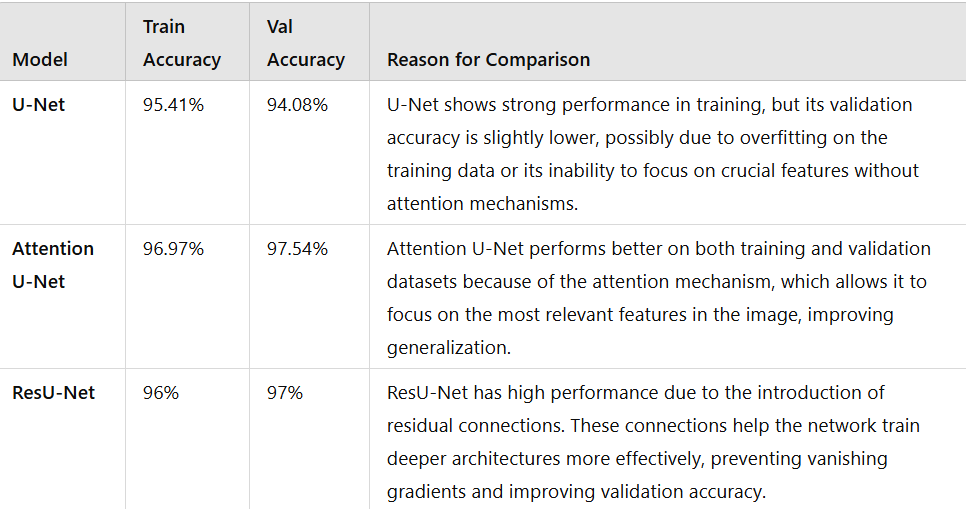
* Purpose: tifffile is used for reading and writing TIFF (Tagged Image File Format) images, which support high-resolution 3D and multi-page images.
* Common Usage: Useful for saving processed image slices or volumes as TIFF files, which are suitable for storing medical image stacks and often retain more detail than standard image formats.

4. **Nibabel (import nibabel as nib)**

Nibabel provides tools to read, write, and manipulate neuroimaging data in popular medical imaging file formats (such as NIfTI .nii and .nii.gz, commonly used for MRI and CT scans).

5. **Keras Optimizers**

**Adam**: Adaptive moment estimation optimizer, which adjusts learning rates for each parameter, balancing convergence speed and performance. It’s ideal for complex models like 3D ResUNet due to its efficiency.

**RESULTS**

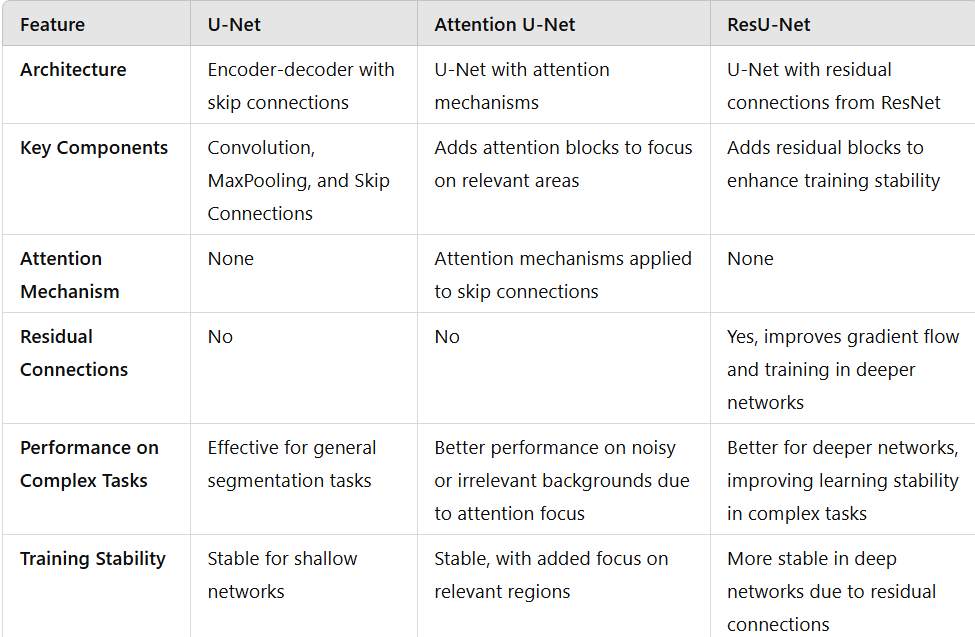
## 

## 

## 

## 

## **COMPARISON**



## CONCLUSION

The experiment shows that Attention U-Net outperformed U-Net and ResU-Net in validation accuracy, highlighting the effectiveness of attention mechanisms in focusing on relevant features and improving generalization to unseen data. ResU-Net, with its residual connections, demonstrated robust training accuracy by mitigating the vanishing gradient problem but slightly lagged in validation performance compared to Attention U-Net, indicating that while residuals aid feature learning, they lack the precision of attention. The baseline U-Net, though simpler, performed well but struggled in complex scenarios, showcasing its strengths as a foundational model.