AUT Journal of Mathematics and Computing

AUT J. Math. Comput., 6(1) (2025) 19-30
https://doi.org/10.22060/AJMC.2023.22562.1172

Original Article

Efficient deep learning algorithms for lower grade gliomas cancer MRI image segmentation: A case study

AmirReza BabaAhmadi*[a], Zahra FallahPour[b]

[a]*School of Mechanical Engineering, College of Engineering, University of Tehran, Tehran, Iran*
[b]*School of Computer Engineering, College of Engineering, University of Shiraz, Shiraz, Iran*

ABSTRACT: This study explores the use of efficient deep learning algorithms for segmenting lower grade gliomas (LGG) in medical images. It evaluates various pre-trained atrous-convolutional architectures and U-Nets, proposing a novel transformer-based approach that surpasses traditional methods. DeepLabV3+ with MobileNetV3 backbone achieved the best results among pre-trained models, but the transformer-based approach excelled with superior segmentation accuracy and efficiency. Transfer learning significantly enhanced model performance on the LGG dataset, even with limited training samples, emphasizing the importance of selecting appropriate pre-trained models. The transformer-based method offers advantages such as efficient memory usage, better generalization, and the ability to process images of arbitrary sizes, making it suitable for clinical applications. These findings suggest that advanced deep learning techniques can improve diagnostic tools for LGG and potentially other cancers, highlighting the transformative impact of deep learning and transfer learning in medical image segmentation.

Review History:

Received:18 July 2023
Revised:01 October 2023
Accepted:13 October 2023
Available Online:01 December 2024

Keywords:

Brain tumor segmentation
Lower grade gliomas
Transformers
SegFormer
MRI images

MSC (2020):

65D18; 68U10; 94A08

1. Introduction

1.1. background and objectives

Lower Grade Gliomas (LGGs), which are tumors in the brain that originate from glial cells, the supportive cells of the brain, are categorized as WHO grade II and III. They tend to grow at a slower rate when compared to higher-grade gliomas. However, they can still cause significant morbidity and mortality if not treated promptly.

LGGs are a diverse set of tumors that exhibit different histopathological and molecular characteristics. They can occur in any part of the brain and can cause a variety of symptoms, such as seizures, headaches, cognitive impairment, and neurological deficits. Due to their slow-growing nature and often nonspecific symptoms, LGGs are often diagnosed incidentally on imaging studies or after prolonged observation of symptoms.

The standard treatment for LGGs usually involves a combination of surgical removal, radiation therapy, and chemotherapy. The degree of surgical removal required is dependent on the size and location of the tumor, as well

*Corresponding author.
E-mail addresses: babaahmadi.amir@ut.ac.ir, farzanehhh.fl@gmail.com

as its proximity to important brain structures. Radiation therapy and chemotherapy are frequently employed in conjunction with surgery to decrease the likelihood of recurrence.

LGGs are typically associated with a more favorable prognosis compared to higher-grade gliomas, with a median survival rate of 5-10 years. Nevertheless, the outcome can be influenced by multiple factors, such as age, tumor position, degree of removal, histopathological subtype, and molecular features. Furthermore, LGGs can advance to higher-grade gliomas over time, which can complicate their treatment and prognosis.

Advances in imaging techniques and molecular profiling have improved our understanding of LGGs and have led to the development of more personalized treatment strategies. However, the management of LGGs remains challenging due to their heterogeneity and variable clinical course. Further research is needed to improve our understanding of LGGs and to develop more effective diagnostic and therapeutic approaches for these tumors.

Deep learning has become a potent technology for medical image segmentation, allowing for the automated and precise detection of areas of interest within intricate medical images. Deep learning algorithms, such as convolutional neural networks (CNNs), have demonstrated encouraging outcomes in diverse medical imaging domains, such as tumor identification, organ segmentation, and disease categorization [4]. These models can acquire intricate representations of the input data and can generalize effectively to new data, making them a highly efficient method for analyzing medical images.

A significant benefit of deep learning models for medical image segmentation is their capacity to learn from vast datasets without necessitating manually-crafted features or domain-specific knowledge [15]. This enables the automated and reliable segmentation of medical images, which can be a time-consuming and error-prone process when performed manually. Moreover, deep learning models can be trained using various imaging modalities, including MRI, CT, and PET, making them adaptable and useful for a broad range of medical imaging tasks [3].

Deep learning has demonstrated significant potential in automating and precisely segmenting brain tumors, including LGGs. These tumors are notoriously challenging to identify and delineate on medical images due to their diverse and infiltrative nature. Deep learning algorithms, such as convolutional neural networks (CNNs), have shown exceptional accuracy and efficiency in segmenting LGGs from MRI scans.

The U-Net architecture is a commonly used deep learning model for medical image segmentation. U-Net [15] is composed of a contracting path that captures context and an expanding path that enables accurate localization of the region of interest. Other deep learning models, like DeepLabV3, have also demonstrated encouraging outcomes for brain tumor segmentation [2].

One challenge in LGG segmentation is the high degree of variability in tumor morphology and location. To address this, transfer learning can be applied to deep learning models, where a pre-trained model is fine-tuned on a smaller dataset [6]. This technique has demonstrated the capacity to enhance segmentation precision while minimizing the requirement for extensive annotated datasets.

Another important consideration in tumor segmentation is the inclusion of different imaging techniques, such as magnetic resonance imaging (MRI) sequences, including those known as T_1, T_2, and FLAIR. Multi-modal segmentation can be achieved using $3D$ CNNs, which can learn from the spatiotemporal relationships between imaging modalities [11]. $3D$ CNNs have been shown to improve segmentation accuracy compared to $2D$ CNNs, particularly in regions where tumors overlap with normal brain tissue [14].

In addition to segmentation, deep learning models can also be used for automated tumor grading, which is important for treatment planning and prognostication. Several research studies have described the utilization of deep learning algorithms for categorizing LGG subtypes using MRI characteristics [18]. These models have shown high accuracy in predicting the histological subtype of LGGs, which can inform treatment decisions.

While deep learning models have shown great promise for LGG segmentation and classification, there are several challenges that must be addressed. The requirement for extensive annotated datasets is one of the primary obstacles, which can be difficult to obtain in the context of rare diseases such as LGGs. Another challenge is the interpretability of deep learning models, which can make it difficult to understand how the model arrived at its segmentation or classification decisions. To tackle this problem, various techniques have been suggested, including saliency maps and gradient-based methods [16].

Deep learning models have demonstrated high accuracy and efficiency in LGG segmentation and classification. These models have the potential to improve treatment planning and prognostication for patients with LGGs. Nevertheless, additional research is necessary to overcome the obstacles associated with deep learning models, such as the necessity for extensive annotated datasets and the comprehensibility of the models.

Recently, a modified version of DeepLab, known as DeepLabV3+, has been proposed for image segmentation. DeepLabV3+ integrates an encoder-decoder architecture with an atrous spatial pyramid pooling module to capture multi-scale contextual information [7]. This model has been shown to outperform U-Net in several medical image segmentation tasks, including LGG segmentation based on our study in later sections. Another deep learning model, known as V-Net, uses a 3D CNN architecture for volumetric segmentation and has shown promising results for LGG segmentation [8].

Overall, deep learning models have shown great promise for LGG segmentation and classification. These models have the potential to improve diagnosis, treatment planning, and prognostication for patients with LGGs. Further research is necessary to effectively address the challenges at hand associated with these models and to develop more interpretable models that can be used in clinical practice.

The purpose of this research article is to investigate the use of effective deep learning algorithms for the segmentation of LGG cancer images. The study compares the performance of various pre-trained models and proposes a transformer-based approach to achieve rapid and efficient LGG segmentation. Mean Intersection over Union (mean IoU) and Dice coefficient are employed as evaluation metrics on a publicly accessible LGG MRI dataset to assess these models, highlighting the potential of transfer learning techniques to enhance segmentation accuracy. The ultimate objective of this paper is to contribute to the development of precise and efficient diagnostic tools for LGG cancer detection and treatment planning.

The present study comprises seven sections, which are designed to address different aspects of our research on image segmentation. Section one provides an overview of the background and objectives of our work, while section two delves into the methodology and explores various types of loss functions that are relevant to image segmentation. In section three, we provide detailed information on the dataset used in this study, as well as the various data augmentation techniques employed. Section four outlines the training phase and evaluation metrics employed in our experiments. Section five presents the final results of our study, while section six and seven offers a concluding discussion and future works directions of the findings and their implications.

2. Method

One of the advantages of DeepLabV3+ over U-Net is its ability to capture multi-scale contextual information, which is particularly useful for LGG segmentation. The atrous spatial pyramid pooling module in DeepLabV3+ allows the model to process the image at multiple scales and capture both local and global contextual information. Additionally, DeepLabV3+ uses dilated convolutions, which broadens the model's receptive field without augmenting the number of parameters. On the other hand, U-Net is computationally more efficient and has a simpler architecture, which makes it easier to train and deploy.

Overall, both DeepLabV3+ and U-Net are powerful deep learning models for medical image segmentation, including LGG segmentation. While DeepLabV3+ has shown better performance in some studies, U-Net remains a popular choice, especially for applications with limited datasets. The choice of the model depends on the specific requirements of the application and the availability of resources. In this study, we have used both DeepLabV3+ and Unet with various efficient backbones. The chosen backbones are MobileNetV3 [1] and EfficientNets [17] for their both accuracy and light computational costs. They have been designed to be embedded in electronics devices and clinical experts' cell phones. Henceforth, they seem to be eligible candidates for our research purpose. In addition, we proposed a new fast and efficient architecture based on transformer which generalizes well to medical image segmentation task. It had an acceptable performance on the available dataset which makes it an ideal candidate for embedded systems and clinical experts' phones.

The Segformer architecture was first introduced in [19] that proposed a hierarchical design for the encoder part and multi-layer perceptron (MLP) networks for the decoder part. The encoder part is designed to produce multi-scale features. However, Segformer discards the use of positional encoding, which can lead to decreased performance when the testing resolution differs from the training resolution.

We believe that the Segformer encoder part is a suitable choice for the task at hand. However, we have decided to remove the MLP layers from the decoder part and replace them with transposed convolutional layers. Convolutional layers are much better for efficient deployment in real-time systems. Our decoder architecture is similar to the architecture proposed in [10] . Fig 1 depicts our modified architecture.

For more information, we kindly ask our readers to study the aforementioned papers for further details. Transformers have emerged as a superior alternative to convolutional networks in computer vision tasks due to their ability to capture long-range dependencies and handle sequential data efficiently. Unlike convolutional networks, which excel at local feature extraction, Transformers leverage self-attention mechanisms to attend to all input positions simultaneously, enabling them to model global relationships effectively. This global context modeling is particularly advantageous in tasks such as object detection, image segmentation, and image generation, where understanding the interactions between distant regions is crucial. Additionally, Transformers have demonstrated remarkable performance in tasks requiring fine-grained details and complex reasoning, thanks to their inherent positional encoding and multi-head attention mechanisms. The self-attention mechanism of Transformers allows them to dynamically attend to relevant parts of the input, enabling adaptive feature selection and learning. This property makes Transformers highly effective in capturing context and reducing information loss, leading to improved performance on various computer vision tasks and establishing them as a state-of-the-art choice in the field. To make Transformers more suitable for image segmentation tasks, multiple novel changes have been applied in the architecture, which will be explained in the following paragraphs. Additionally, the methodology for designing a novel decoder section will also be discussed.

2.1. Segformer Encoder

The SegFormer's encoder consists of four Transformer blocks that acquire hierarchical feature maps. Just like the Vision Transformer(ViT), each Transformer block can be divided into three primary components: the patch embedding layer, the attention layer, and the position embedding layer.

- Patch embedding layer: The patch embedding layer in SegFormer serves the purpose of dividing the image into smaller patches and converting them into embedding vectors. Unlike conventional Vision Transformers that divide the input image into non-overlapping patches, which disrupts the local continuity, SegFormer introduces overlapped patch embedding and merging techniques to tackle this issue. The overlapped patch embedding layer is implemented using a convolution layer, specifically the `nn.Conv2D` function in the PyTorch library. The degree of overlap is determined by the stride value of the convolution operation. To capture hierarchical features at both high and low resolutions, SegFormer employs four Transformer blocks—T1, T2, T3, and T4—where their feature maps have dimensions of $H \times 2^{(i+1)} \times H \times 2^{(i+1)} \times 2^{(i+5)}$ (where $i \in \{1, 2, 3, 4\}$). These blocks have kernel sizes of {7, 3, 3, 3} and strides of {4, 2, 2, 2}, respectively.

- Attention Layer: The attention layer in ViT consists of a multi-head self-attention module (MSA) that plays a crucial role in capturing dependencies among image patches or embedding vectors, specifically global dependencies. However, ViT suffers from significant computational complexity. To overcome this challenge, SegFormer introduces a technique called efficient multi-head self-attention (EMSA), which incorporates sequence reduction to decrease the number of embedding vectors. EMSA reduces the number of embedding vectors from N to N/r, where N represents the product of the image height (H) and width (W), and r is a hyperparameter known as the reduction ratio. EMSA is implemented using a convolution layer, specifically the `nn.Conv2D` function in PyTorch. Initially, a feature map with dimensions $H \times W \times C$ is reshaped to $N/r \times r \times C$. This step reduces the number of embedding vectors to N/r while increasing the length of each embedding vector to $r \times C$. Next, a fully connected layer is employed to reduce each embedding vector back to its original size C. Finally, the reduced feature map undergoes conventional MSA.

- Position embedding layer: The position embedding layer in ViT explicitly encodes and appends positional information of each patch to the patch embedding vector. However, encoding positional information directly into different levels of hierarchical feature maps poses a challenge. To address this, SegFormer introduces a 3×3 convolution operation that implicitly learns the positional information of patches. Additionally, a skip connection is employed to incorporate the positional information into the feature map.

2.2. SegFormer Revised Decoder

The decoder in image segmentation plays a crucial role in the overall process by refining the high-resolution feature maps generated by the encoder and pro-ducing the final segmentation output. Its primary function is to recover spatial details and localize object boundaries. However, the decoder often faces several bottlenecks that need to be addressed for optimal performance. One bottleneck is the loss of fine-grained information due to down sampling operations in the encoder. The decoder needs to effectively recover these details to ensure precise object segmentation. Another bottleneck arises from the limited receptive field of individual decoder layers, which restricts their ability to capture long-range dependencies and context. This limitation can result in incomplete object understanding and segmentation errors. Furthermore, the decoder must effectively fuse information from multiple encoder layers to leverage both low-level and high-level features. Achieving this fusion while maintaining spatial coherence and avoiding information loss is a significant challenge. Additionally, the decoder needs to handle class imbalance issues, as segmentation tasks often involve imbalanced object distributions, where certain classes are more prevalent than others. Ensuring accurate segmentation for all classes requires specialized techniques to address this imbalance. Addressing these bottlenecks requires careful architectural design and the integration of advanced techniques. Up sampling operations, such as transposed convolutions or interpolation, can help recover spatial details lost during down sampling. Incorporating skip connections that bridge the encoder and decoder layers enables the effective fusion of multi-scale features, enhancing segmentation accuracy. Additionally, employing dilated convolutions or spatial pyramid pooling can expand the receptive field, allowing the decoder to capture broader contextual information. Finally, techniques like class-balanced loss functions or data augmentation strategies can help mitigate the effects of class imbalance and improve segmentation performance across all object classes. In original SegFormer, the Decoder part consists of multiple MLPs which incurs high computational costs

Using MLPs (Multi-Layer Perceptron) in the decoder for image segmentation tasks has several disadvantages. Firstly, MLPs lack the ability to capture spatial relationships between neighboring pixels, leading to fragmented segmentation results. Without considering the context of surrounding pixels, MLPs struggle to maintain spatial coherence in the segmentation output. Moreover, MLPs are parameter-intensive, requiring a large number of parameters, which can result in high computational and memory requirements. This can make training and inference with MLP-based decoders computationally expensive and resource-intensive, limiting their practicality for image segmentation tasks. Another drawback of using MLPs in the decoder is their difficulty in handling varying image resolutions. MLPs are designed for fixed-size inputs, and resizing or reshaping images to fit the fixed input size can lead to the loss of spatial in-formation and introduce distortions that negatively impact segmentation accuracy. Additionally, MLPs are limited in their ability to capture long-range dependencies and global context, which are crucial for accurate image segmentation. The lack of efficient modeling of complex spatial relationships between distant regions hinders the decoder's understanding of the overall scene and context. In the context of image segmentation, transposed convolutions offer advantages over MLP layers in the decoder. Transposed convolutions, also known as deconvolutions or up sampling operations, help recover spatial details lost during down sampling in the encoder. They enable the decoder to reconstruct higher-resolution feature maps, preserving spatial information and improving segmentation accuracy. The use of transposed convolutions facilitates the effective fusion of low-level and high-level features, enabling the decoder to leverage multi-scale information for more precise object localization and segmentation. Overall, transposed convolutions provide a more suitable and efficient approach for image segmentation tasks compared to MLP layers, addressing the limitations associated with MLP-based decoders. We employed several transposed convolutional layers, which progressively upsample the image to match the original image's size before downsampling. These layers utilize skip connections to preserve spatial information from the hierarchical representations of the encoder. The skip connections allow the decoder to access high-resolution details from earlier stages of the encoder, maintaining fine-grained spatial information during the upsampling process. This fusion of features from multiple levels enhances the model's ability to capture both local and global context, improving the accuracy in localizing objects and boundaries. The model also benefits from gradient flow directly from the decoder to the encoder layers, mitigating the vanishing gradient problem and facilitating effective learning and faster convergence. Overall, these design choices increase the model's capacity, enabling it to capture a broader range of patterns and variations in the input data, leading to improved segmentation performance and more accurate results. Then, we used a dense layer with 256 neurons to fuse all the information from the decoder, which was concatenated channel-wise. After that, we applied a 1x1 convolution with 1024 filters to enhance the details obtained from the previous layer. Following that, a Conv2d layer and an Identity layer were used to resize the image and pass it to the sigmoid function, which makes decisions about the pixel-wise classification task at hand. More details are available in Figure1.

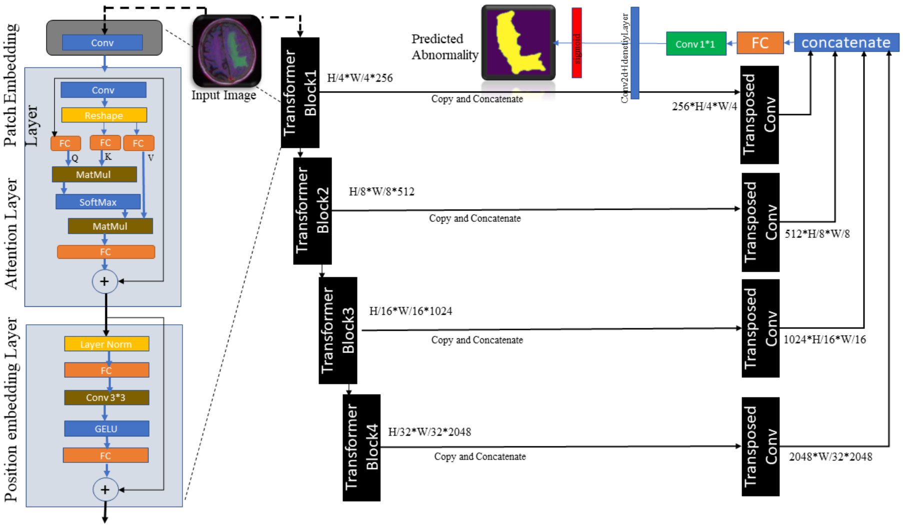

Figure 1: Modified Transformer-based Architecture for Brain Tumor Segmentation

Medical image segmentation is a complex undertaking, and the selection of an appropriate loss function is pivotal for precise segmentation. Multiple loss functions are available for medical image segmentation, and the selection of a suitable loss function is dependent on the particular task and dataset. This response will explore some frequently employed loss functions for medical image segmentation.

2.3. Different types of loss functions for image segmentation

- Binary Cross-Entropy Loss: Binary cross-entropy (BCE) loss is the most widely employed loss function for binary segmentation/classification challenges. BCE loss quantifies the discrepancy between predicted and actual probabilities and is computed as the negative logarithm of the predicted probability for the correct class. BCE loss is straightforward to implement and computationally efficient.

$$Binary - CrossEntropyLoss = -\frac{1}{N}\sum_{i=1}^{n} y_i \cdot \log(p(y_i)) + (1-y_i)\cdot \log(1-p(y_i)) \tag{1}$$

- Dice Loss: Dice loss is a metric that calculates the agreement between predicted and actual masks and is especially advantageous for datasets that are imbalanced, with one class being significantly more prevalent than the other. Dice loss is computed by taking twice the intersection between predicted and actual masks and dividing it by the sum of pixels in both masks. Dice loss is a symmetric measure and has demonstrated encouraging outcomes for medical image segmentation.

$$Dice = 2 * y \cap ypredy + ypred \tag{2}$$

$$DiceLoss = 1 - Dice \tag{3}$$

- Jaccard Loss: Jaccard loss, commonly referred to as intersection over union (IoU) loss, is a measure that quantifies the resemblance between predicted and actual masks. Jaccard loss is computed by dividing the intersection between predicted and actual masks by their union. Jaccard loss is also a symmetric measure and is closely linked to Dice loss

$$JaccardCoefficient = |A \cap B||A \cup B| = |A \cap B||A| + |B| - |A \cap B| \tag{4}$$

$$JaccardLoss = 1 - JaccardCoefficient \tag{5}$$

BCE-Dice loss: it's a combination of the binary cross-entropy (BCE) loss and the Dice coefficient. It is commonly used in image segmentation tasks, particularly for binary segmentation. The BCE-Dice loss combines these two losses (Binary Cross entropy and Dice Loss) to encourage the model to simultaneously optimize for pixel-level binary classification and segmentation accuracy. The formula for BCE-Dice loss is:

$$BCE - Diceloss = (1-a)BCEloss + \alpha * Dicecoefficient \tag{6}$$

where α is a weighting factor that balances the contribution of the BCE loss and Dice coefficient.

3. Dataset

We have used the public dataset available at the link "url/https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation" . The dataset comprises FLAIR (Fluid-Attenuated Inversion Recovery), T_1-weighted, T_1-weighted with contrast-enhancement, and T_2-weighted sequences of Magnetic Resonance Imaging (MRI) scans of the brain. Each sequence consists of multiple slices, and the dataset contains 2D axial slices of MRI scans from a total of 110 patients specifically diagnosed with lower-grade gliomas (LGG).

The dataset is labeled with binary masks for each MRI slice indicating the presence or absence of a glioma. The masks were created using a combination of manual segmentation and semi-automated methods. The dataset has been partitioned into training (80%) and testing (20%) subsets, with the testing set consisting of Magnetic Resonance Imaging (MRI) scans of patients who were not present in the training set. The purpose of using a separate testing set is to evaluate the performance of the model on previously unseen data and to determine if it can generalize to new cases. The dataset is provided in the form of NIfTI files (. nii.gz) for the MRI scans and PNG files for the binary masks. The number of training data samples is 2750, the validation set consists of 590 samples, and there are 589 samples in the test set.

This dataset has become a popular benchmark dataset for testing the performance of various deep learning models for medical image segmentation, especially for brain tumor segmentation. The use of this dataset has enabled researchers "To assess and contrast the effectiveness of various models and evaluate the effectiveness of different techniques for brain tumor segmentation.

However, it is important to note that this dataset has certain limitations. Firstly, the dataset only contains images of LGG tumors, and therefore, the models trained on this dataset may not generalize well to other types of brain tumors. Secondly, the masks provided in the dataset may not be perfect, and there may be some errors or inaccuracies in the ground truth labels. Regardless, this dataset has served as a significant asset for researchers to appraise and enhance the efficacy of deep learning models for the segmentation of brain tumors, and has spurred further research in this area. Fig.2 depicts samples from the dataset at hand.

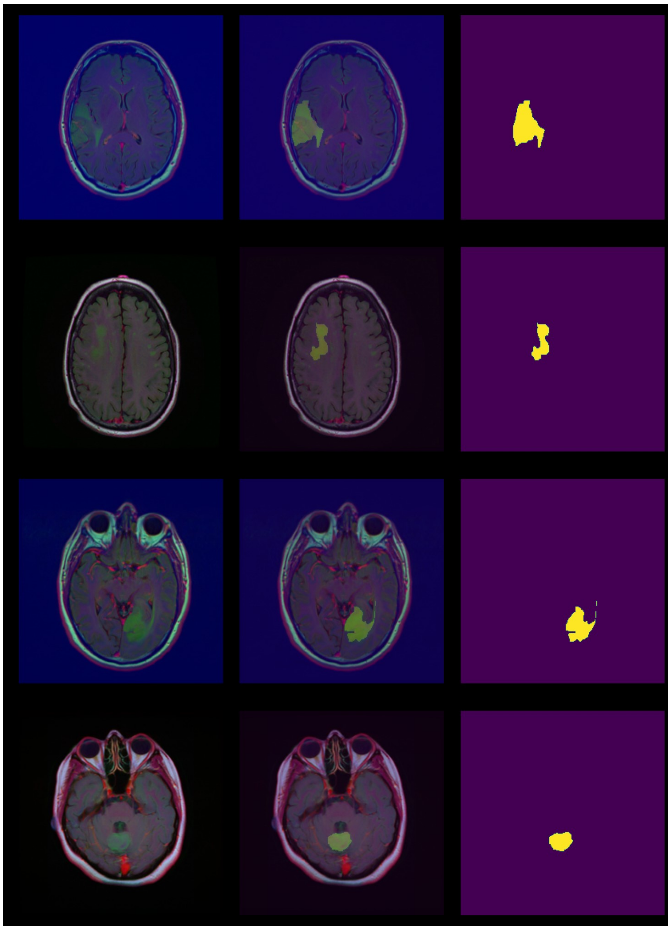

Figure 2: LGG samples with masks and their highlighted abnormality

3.1. Data Augmentation

Techniques for augmenting image data are utilized to expand the size of training datasets by implementing diverse transformations on the images. This can aid in enhancing the resilience of deep learning models to fluctuations in the input data. Some of the image data augmentation approaches employed in our study include channel dropout, random brightness, and color jitter.

- **Channel Dropout:** Channel dropout is a technique used to randomly set entire color channels of an image to zero. This can help to prevent overfitting By compelling the model to acquire more resilient features that are not dependent on specific color channels. Channel dropout can be applied by randomly selecting one or more channels to drop for each image in the training dataset.
- **Random Brightness and Contrast Adjustment:** Random brightness and contrast adjustment is a technique used to adjust the brightness and contrast of an image by adding random values to the pixel values. This can help to Enhance the ability of the model to generalize by creating additional training data with different levels of brightness and contrast. The amount of adjustment can be randomly selected within a specified range.
- **Color Jitter:** Color jitter is a technique used to randomly modify the hue, saturation, and brightness of an image. This approach can assist in augmenting the variety of the training dataset and enhancing the model's generalization capability. The extent of jitter can be randomly chosen from a predefined range.

Data augmentation techniques, such as channel dropout, random brightness and contrast adjustment, and color jitter are widely employed in medical image analysis tasks, including tumor segmentation. In brief, these techniques can expand the size and diversity of training datasets, thereby enhancing the performance of deep learning models for medical image analysis tasks

4. Training Phase

4.1. Learning Phase

Initially, we normalized the dataset and then fed it into all available architectures, as well as our designed architecture. The loss function used was BCE-Dice, the optimizer was Adam, the batch size was 16, and the initial learning rate was set to 0.001. We used a learning rate scheduler ”Reduce LR on Plateau”, with a patience value of 3 and a factor of 0.2. All models were implemented in PyTorch. We trained all models for 60 epochs and evaluated them using the test set. It is important to note that we were unable to increase the batch size due to hardware limitations. All the models were run on Nvidia Tesla T_4 with 16 GB of memory. Increasing the batch size can dramatically influence the performance of algorithms. Hence, we leave it to future studies to explore the impact of larger batch sizes on the performance of our models. Multiple architectures were trained using DeepLabV3+ and U-Net with different backbones in this study.

4.2. Evaluation Metrics and Inference Phase

The evaluation metrics used in this case study were Mean Intersection over Union (mean IoU) and Dice coefficient. In this work, we utilized the BCE-Dice loss, which is a weighted combination of Binary Cross Entropy and Dice Loss. The BCE loss serves as a binary classifi-cation loss and quantifies the disparity between the predicted and actual binary labels. It is computed by summing the negative logarithm of the predicted proba-bility for the correct class. On the other hand, the Dice coefficient is commonly employed in image segmentation to assess the similarity between two sets. It measures the overlap between the predicted and ground truth segmentation masks and is calculated as twice the intersection of these masks divided by the sum of their respective areas. The BCE-Dice loss offers several advantages when used as a loss function, particularly in image segmentation tasks. By combining the strengths of Binary Cross Entropy (BCE) and Dice Loss, it provides a balanced and comprehensive objective function. The BCE loss focuses on accurate binary classification by penalizing discrepancies between predicted and ground truth labels. It accomplishes this by summing the negative logarithm of the predicted probability for the correct class. On the other hand, the Dice coefficient is commonly employed in image segmentation to assess the similarity between two sets, specifically the overlap between predicted and ground truth segmentation masks. One key benefit of using BCE-Dice loss is its ability to address class imbalance in segmentation tasks. Class imbalance occurs when the number of pixels belonging to one class significantly outweighs the other. This imbalance can negatively impact model performance. However, BCE-Dice loss helps mitigate this issue by considering both binary cross-entropy and the Dice coefficient. By incorporating these components, the loss function can alleviate the effects of class imbalance and improve the segmentation results. Another advantage of BCE-Dice loss is its effectiveness in handling ambiguous boundaries. Accurate delineation of object boundaries is crucial in image segmentation. BCE-Dice loss encourages the model to produce sharper and more precise predictions by incorporating the Dice coefficient. The Dice coefficient evaluates the overlap between the predicted and ground truth masks, and by optimizing this measure, the loss function helps the model generate improved segmentation results, particularly in scenarios with complex or ambiguous boundaries. Furthermore, the adaptability and customization of BCE-Dice loss make it a valuable choice for researchers and practitioners. The weighting factor between the BCE and Dice components can be adjusted to suit the specific requirements of the task. This flexibility allows fine-tuning of the loss function based on the dataset's characteristics and the desired trade-off between different evaluation criteria. Researchers can experiment with different weightings to find the optimal balance that leads to improved segmentation performance in their particular applications. To put in a nutshell, the BCE-Dice loss provides a balanced approach to image segmentation by combining BCE and Dice Loss. It addresses class imbalance, encourages accurate boundary delineation, and allows for customization based on specific needs. These benefits make BCE-Dice loss a valuable tool in improv-ing segmentation performance across various applications. Mean Intersection over Union (IoU) and Dice Coefficient are widely used evaluation metrics in image segmentation tasks. Both metrics assess the similarity between the predicted and ground truth segmentation masks, providing valuable insights into the quality of segmentation results. The Mean IoU, also known as the Jaccard Index, measures the intersection over the union of the predicted and ground truth masks for each class and then calculates the average across all classes. It quantifies the extent of overlap between the predicted and ground truth regions, providing a measure of accuracy and completeness in segmentation. A higher Mean IoU indicates a better alignment between the predicted and ground truth masks, suggesting a more accurate segmentation. On the other hand, the Dice Coefficient, also known as the F1 score, is another popular metric used in image segmentation. It evaluates the overlap between the predicted and ground truth masks by calculating twice the intersection of the two masks divided by the sum of their areas. The Dice Coefficient measures the similarity between the two sets, with a value of 1 indicating a perfect match and 0 indicating no overlap. Similar to the Mean IoU, a higher Dice Coefficient signifies a better agreement between the predicted and ground truth segmentation masks. Both Mean IoU and Dice Coefficient provide valuable information about the quality of segmentation results. However, they have different characteristics and interpretations. Mean IoU considers the overall overlap between the predicted and ground truth masks across all classes, providing a global assessment of segmentation accuracy. It is particularly useful when dealing with imbalanced datasets or when evaluating multi-class segmentation tasks. On the other hand, the Dice Coefficient focuses on the local agreement between the masks, emphasizing the similarity between the predicted and ground truth regions on a per-pixel ba-sis. While Mean IoU and Dice Coefficient are commonly used evaluation metrics, it's important to note that they have their limitations. Both metrics treat each class equally, regardless of class size or importance. This can be problematic when dealing with imbalanced datasets, where certain classes may dominate the overall evaluation. Additionally, these metrics only provide a numerical measure of segmentation quality and do not consider other factors such as boundary accuracy or semantic understanding. In summary, Mean IoU and Dice Coefficient are both valuable metrics for evaluating image segmentation results. Mean IoU provides a global measure of overlap and accuracy across all classes, while Dice Coefficient focuses on the per-pixel similarity between predicted and ground truth masks. Understanding these metrics and their interpretations can help researchers and practitioners assess and compare the performance of different segmentation models and tech-niques.

5. Results and Discussion

This section provides the outcomes of evaluating all the algorithms discussed in the preceding sections, which are presented and documented in TABLE 1, along with the parameter count of each model. Based on the results presented in TABLE 1, our model has outperformed all the efficient models with a small number of parameters. It appears that the DeepLab V3+ architecture is superior in most cases compared to the U-Net architecture. The two best candidates among the efficient algorithms are DeepLabV3+ with Mo-bileNetV3 and EfficientNet-b0 as backbones. We believe that our modified Trans-former-based solution, with about 5.5E6 parameters, is appropriate for implementation in devices with restricted computational capacity. Furthermore, we found that our model is better at segmenting small-sized tumors than other algorithms. However, we believe that further investigation is required on other benchmarks as well. Our proposed model has been compared to state-of-the-art published papers in terms of the Dice Score. Our Proposed method outperforms previous works us-ing vanilla Unet[13] by 11 % improvement and SynthSeg [5] by 7 % margin.

Table 1: Brain Tumor Segmentation Results

Architecture	Encoder (Backbone)	Squeeze & Excitation Block (SE Block)	Mean IoU	Dice Coefficient	No. Parameters
DeepLabV3+	EfficientNet-b7	No	0.89	0.92	67.1E6
DeepLabV3+	EfficientNet-b0	No	0.90	0.92	4.90E6
DeepLabV3+	MobileNetV3-Large	No	0.90	0.92	4.7E6
U-Net	EfficientNet-b0	Yes	0.86	0.86	6.25E6
U-Net	EfficientNet-b7	Yes	0.89	0.91	67.09E6
U-Net	MobileNetV3-Large	Yes	0.88	0.91	6.68E6
Vanilla Unet [15]	-	-	-	0.82	-
SynthSeg [5]	-	-	-	0.86	-
SWTRU [21]	Tranforemr with reinforced Unet	-	-	0.86	-
Unet++ [20]	-	-	-	0.89	-
TransU^2-Net [12]	Tranformers	-	-	0.86	-
Unet with new residual connections [9]	-	-	0.87	0.92	-
Modified SegFormer(ours)	Four Transformers Block	No	0.91	0.93	5.54E6

In terms of specifics, our model outperformed DeepLabV3+ with MobileNetV3 and EfficientNet-b0 by around 1% in mean Intersection over Un-ion (IoU) and Dice Coefficient. Although this improvement may seem insignificant at first glance, our experiments indicate that even a slight enhancement in the medical field can have a significant impact on saving lives. Figure 3 displays the predicted abnormalities using our Transformer-based approach.

The main challenge we aimed to address was the high computational costs associated with utilizing transformer-based models on embedded devices, such as smartphones and conventional computers, in both the training and inference phases for medical experts. To a certain extent, the innovations presented in this article have mitigated this challenge. However, we believe that there is still ample room for further improvement in reducing the computational requirements of training this algorithm and similar ones. While certain modifications have been applied to the decoder section, additional research is needed to achieve computational complexity approaching linearity by focusing on the encoder part, which exhibits greater complexity compared to the decoder part. Another limitation of our work lies in its inability to handle different modalities of medical images, an aspect that we did not explore in this paper.

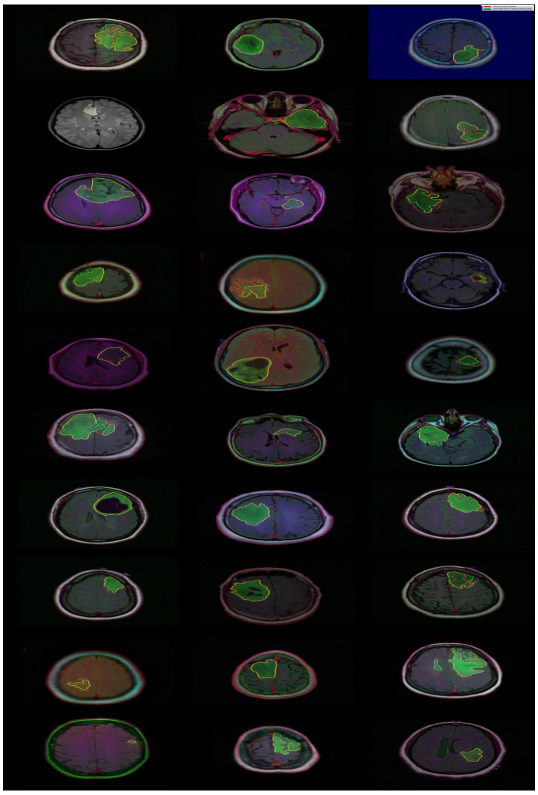

Figure 3: Abnormality prediction using the proposed Transformer-based solution.

6. Future Works

We believe that there exist numerous potential solutions to enhance the efficiency of transformers in the computer vision domain and enable their deployment on edge devices for medical experts. Exploring the encoder to achieve linear computational complexity represents a promising avenue for researchers to pursue. Furthermore, tackling the challenge of handling diverse modalities of medical images and incorporating text notes from medical experts can unveil valuable insights and render models more informative within the realm of multi-modal domains.

7. Conclusion

In this paper, we have presented a transformer-based solution for the medical image segmentation task that outperforms several state-of-the-art algorithms, including DeepLabV3+ and U-Net with different backbones. Our results demonstrate that our solution is suitable for use in devices with limited computational power and is particularly effective in segmenting small-sized tumors. Our proposed approach yielded a mean IoU and Dice Coefficient of 91% and 93% respectively, which is higher than that achieved by DeepLabV3+ with MobileNetV3 and EfficientNet-b0. Additionally, our proposed method outperforms previous works by 11% improvement against Vanilla Unet and 7% against SynthSeg method in terms of Dice Score. In the medical domain, even a slight improvement in these metrics can have a significant impact on people's lives.

Our study highlights the potential of transformer-based solutions in the medical image segmentation field. We believe that this work can serve as a basis for future research on this topic. Overall, our solution has demonstrated promising results and opens up new avenues for exploring the potential of transformer-based models in medical image segmentation.

References

[1] H. Andrew, S. Mark, C. Grace, C. Liang-Chieh, C. Bo, T. Mingxing, W. Weijun, Z. Yukun, P. Ruoming, V. Vijay, et al., *Searching for mobilenetv3*, in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1314–1324.

[2] R. Azad, M. Heidari, M. Shariatnia, E. K. Aghdam, S. Karimijafarbigloo, E. Adeli, and D. Merhof, *Transdeeplab: Convolution-free transformer-based deeplab v3+ for medical image segmentation*, in International Workshop on PRedictive Intelligence In MEdicine, Springer, 2022, pp. 91–102.

[3] A. BabaAhmadi, S. Khalafi, and F. M. Esfahani, *Designing an improved deep learning-based classifier for breast cancer identification in histopathology images*, in 2022 International Conference on Machine Vision and Image Processing (MVIP), IEEE, 2022, pp. 1–4.

[4] A. BabaAhmadi, S. Khalafi, M. ShariatPanahi, and M. Ayati, *Designing an improved deep learning-based model for covid-19 recognition in chest x-ray images: A knowledge distillation approach*, arXiv preprint arXiv:2301.02735, (2023).

[5] B. Billot, D. Greve, K. Van Leemput, B. Fischl, J. E. Iglesias, and A. V. Dalca, *A learning strategy for contrast-agnostic mri segmentation*, arXiv preprint arXiv:2003.01995, (2020).

[6] M. Buda, E. A. AlBadawy, A. Saha, and M. A. Mazurowski, *Deep radiogenomics of lower-grade gliomas: Convolutional neural networks predict tumor genomic subtypes using mr images*, Radiology: Artificial Intelligence, 2 (2020), p. e180050.

[7] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, *Encoder-decoder with atrous separable convolution for semantic image segmentation*, in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.

[8] R. Hua, Q. Huo, Y. Gao, H. Sui, B. Zhang, Y. Sun, Z. Mo, and F. Shi, *Segmenting brain tumor using cascaded v-nets in multimodal mr images*, Frontiers in Computational Neuroscience, 14 (2020), p. 9.

[9] C. Huang and M. Wan, *Automated segmentation of brain tumor based on improved u-net with residual units*, Multimedia Tools and Applications, 81 (2022), pp. 12543–12566.

[10] V. Iglovikov and A. Shvets, *Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation*, arXiv preprint arXiv:1801.05746, (2018).

[11] S. Lefkovits, L. Lefkovits, and L. Szilágyi, *Hgg and lgg brain tumor segmentation in multi-modal mri using pretrained convolutional neural networks of amazon sagemaker*, Applied Sciences, 12 (2022), p. 3620.

[12] X. Li, X. Fang, G. Yang, S. Su, L. Zhu, and Z. Yu, *Transu^2-net: An effective medical image segmentation framework based on transformer and u^2-net*, IEEE Journal of Translational Engineering in Health and Medicine, 11 (2023), pp. 441–450.

[13] S. Mishra and M. Manish, *Studies on computational grafting of malarial epitopes in serum albumin*, Computers in Biology and Medicine, 102 (2018), pp. 126–131.

[14] H. Mzoughi, I. Njeh, A. Wali, M. B. Slima, A. BenHamida, C. Mhiri, and K. B. Mahfoudhe, *Deep multi-scale 3d convolutional neural network (cnn) for mri gliomas brain tumor classification*, Journal of Digital Imaging, 33 (2020), pp. 903–915.

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

[16] H. Saleem, A. R. Shahid, and B. Raza, *Visual interpretability in 3d brain tumor segmentation network*, Computers in Biology and Medicine, 133 (2021), p. 104410.

[17] M. Tan and Q. Le, *Efficientnet: Rethinking model scaling for convolutional neural networks*, in International conference on machine learning, PMLR, 2019, pp. 6105–6114.

[18] S. Vidyadharan, B. V. V. S. N. Prabhakar Rao, Y. Perumal, K. Chandrasekharan, and V. Rajagopalan, *Deep learning classifies low-and high-grade glioma patients with high accuracy, sensitivity, and specificity based on their brain white matter networks derived from diffusion tensor imaging*, Diagnostics, 12 (2022), p. 3216.

[19] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, *Segformer: Simple and efficient design for semantic segmentation with transformers*, Advances in Neural Information Processing Systems, 34 (2021), pp. 12077–12090.

[20] D. Xu, X. Zhou, X. Niu, and J. Wang, *Automatic segmentation of low-grade glioma in mri image based on unet++ model*, 1693 (2020), p. 012135.

[21] J. Zhang, Y. Liu, Q. Wu, Y. Wang, Y. Liu, X. Xu, and B. Song, *Swtru: Star-shaped window transformer reinforced u-net for medical image segmentation*, Computers in Biology and Medicine, 150 (2022), p. 105954.

Please cite this article using:

AmirReza BabaAhmadi, Zahra FallahPour, Efficient deep learning algorithms for lower grade gliomas cancer MRI image segmentation: A case study, AUT J. Math. Comput., 6(1) (2025) 19-30
https://doi.org/10.22060/AJMC.2023.22562.1172