ITC 4/53
Information Technology and Control
Vol. 53 / No. 4/ 2024
pp. 1074-1087
DOI 10.5755/j01.itc.53.4.37719

Imaging Segmentation of Brain Tumors Based on the Modified U-net Method

Received 2024/06/22 | Accepted after revision 2024/09/13

HOW TO CITE: Zhang, Y., Ngo, H. C., Zhang, Y., Yusof, N. F. A., Wang, X. (2024). Imaging Segmentation of Brain Tumors Based on the Modified U-net Method. *Information Technology and Control*, 53(4), 1074-1087. https://doi.org/10.5755/j01.itc.53.4.37719

Imaging Segmentation of Brain Tumors Based on the Modified U-net Method

Yajie Zhang
School of Artificial Intelligence, Wenzhou Polytechnic, Wenzhou, 325000, Zhejiang, China
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

Hea Choon Ngo
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

Yifan Zhang
Division of Nephrology, Wenzhou Central Hospital, The Dingli Clinical Institute of Wenzhou Medical University, Wenzhou, 325000, China

Noor Fazilla Abd Yusof
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

Xiaohan Wang
Division of Nephrology, Wenzhou Central Hospital, The Dingli Clinical Institute of Wenzhou Medical University, Wenzhou, 325000, China

Corresponding author: Yajie Zhang, e-mail: zhangyajie98@outlook.com

Brain tumor segmentation in medical image analysis is a challenging task. Deep learning techniques have recently shown promise in resolving a variety of computer vision problems, such as semantic segmentation and image classification. Brain MRI (magnetic resonance imaging) requires precise brain image segmentation for effective, rapid diagnosis and treatment planning. However, it is quite difficult to manually segment the brain image rapidly and accurately in low-quality, noisy images. This paper proposes a U-Net and combined attention mechanism-based method. This research enhances the segmentation of images of tumors in the brain using modified U-net. Traditional U-net segmentation techniques are still widely used in the medical field, but they have a number of limitations when dealing with small targets and fuzzier boundaries. To address this issue, we made the following modifications to U-net: We propose attention mechanisms to assist the network in concen-

trating on important regions. The multiscale feature fusion strategy improves the efficacy of network segmentation at various scales. Cross-entropy loss function and data augmentation improve the performance of the network further. Our method was validated using the Brats2019 dataset. The experimental results demonstrate that our proposed methodology exhibits superior speed and efficiency compared to existing techniques in the context of brain image segmentation. The dice coefficients for the multiple branch TS-U-net model were 0.876, 0.868, and 0.814 in the tumor subregions of WT, TC, and ET, respectively. This exemplifies the feasibility and potential of our methodology for the segmentation of medical images.

KEYWORDS: Brain Tumor, Deep Learning, Image Segmentation, U-net.

1. Introduction

Brain tumors are one of the most severe brain diseases, and malignant glioma is the most common and mortality primary brain tumor. According to American Brain Tumor Association, approximately 80,000 Americans are diagnosed with brain tumor each year.

The World Health Organization categorized gliomas into four grades (I, II, III, IV) based on their malignancy [2, 5]. Additionally, it is worth noting that Gliomas tumors are the predominant primary brain tumors in the adult population [4]. These tumors account for approximately 81% of all malignant brain tumors [8] and 45% of all primary brain tumors [6]. According to Ostrom et al., the survival rates for some people diagnosed with Gliomas tumors range from 0.05% to 4.7% [9]. This data indicates that Gliomas tumors are the second leading cause of mortality. According to Glas et al. [3], individuals who are afflicted with low-grade gliomas, specifically oligodendrogliomas and astrocytomas, exhibit a 5-year survival rate of 57%. However, the 5-year survival rate for high-grade gliomas, or grade IV glioblastoma, is about 5%. In this paper, our method is oriented to the segmentation of Glioblastomas which are brain tumors belonging to the category of Gliomas tumors.

The health of patients is highly dependent on timely detection and an accurate assessment of their prognosis. The gold standard for treating grade I and higher tumors in clinical practice is surgical excision. In the process of analyzing brain MRI images, a highly trained radiologist use a manual segmentation technique that incorporates data from MRI images together with their extensive anatomical and physiological expertise [5]. It is well known that the manual segmentation of MRI images is a time-consuming and arduous procedure. The primary obstacle is in the interpretation and segmentation of brain MRI images, which is contingent upon the proficiency of individual radiologists. Furthermore, the task of a radiologist becomes significantly challenging when dealing with tumor regions that contain intra-tumoral structures, such as Glioblastomas tumors. These tumors exhibit three distinct structures within the tumor region, in addition to healthy tissue: Necrotic and Non-Enhancing tumor, Peritumoral Edema, and Enhancing tumor. Given the objective of getting a segmentation of brain tumors, it is obvious that the utilization of completely automated segmentation methods is of great benefit in both clinical settings and research endeavours.

2. Related Work

2.1. Medical Image Segmentation Based on Machine Learning

In recent years, machine learning based segmentation algorithms have been utilized extensively in medical image segmentation. The earliest learning-based segmentation algorithms are manual feature extraction and classifier classification. A disadvantage of this approach is the requirement for meticulous and extensive feature extraction, as well as the potential impact of feature quality on segmentation accuracy. Convolutional neural network (CNN) models are extensively employed in the field of medical image segmentation due to their ability to perform feature extraction without the need for intricate methodologies. Considerable progress has been achieved in the segmentation of several anatomical regions, including the liver, prostate, brain, pancreas, and other tissues, as well as in the segmentation of brain tumors, liver cancer, and breast cancer. The medical imaging methods utilized for this objective include magnetic resonance imaging (MRI), computed tomography (CT),

X-ray, ultrasound (US), and neuropathological optical images. CNN can be used to classify two distinct categories based on their calculation methodology: image block-based convolution segmentation models and full convolution-based models. Although there are benefits associated with augmenting the training sample size and enhancing computational efficiency using picture block-based convolution segmentation, further refinement of the method is necessary due to the requirement for sufficient computational memory resources. Additionally, the presence of redundant image information within the training image block results in insufficient accuracy for the segmentation process. Usually, the segmentation probability distribution is obtained by convolution calculation. Then, the predicted probability distribution is the initial segmentation result. Finally, the results are optimized with the help of the traditional model segmentation algorithm. This type of convolutional neural network could be more computationally efficient. At the same time, there is no good use of spatial information. To overcome this problem, Ronneberger et al. [10] have proposed a computational segmentation model based on fully convolutional networks, UNet.

One perspective is the reconstruction of the dimension-reduction image through the process of deconvolution. Simultaneously, in order to get complete automation of segmentation, it is necessary to consider the data from the convolution reduction layer that corresponds to the task. Presently, medical photographs are extensively utilized in the field. Milletari et al. [8] enhanced the UNet architecture by incorporating 3D convolution for their calculations. Promising results are achieved in the field of medical picture segmentation. In a recent study, HanX introduced a novel approach that combines the concepts of U-Net and ResNet. This approach, referred to as DCNN, achieved the top position in liver tumor segmentation during the ISBI2017 competition. In contrast to the image block-based segmentation approach, the fully convolutional network (FCN) achieves end-to-end learning and streamlines the computational process. However, the primary obstacle lies in addressing the issues of imbalanced training samples, limited computing resources, and accurate boundary positioning. These challenges serve as the foundation for the deep learning segmentation model proposed in this research paper.

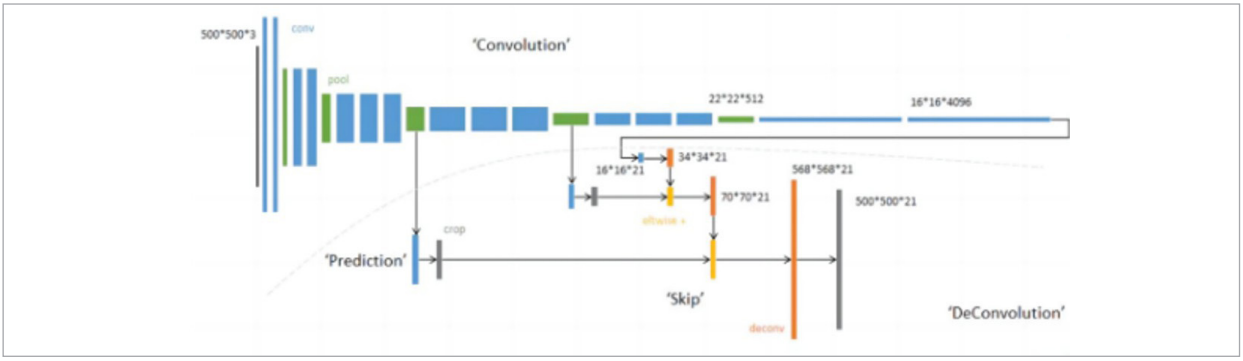
2.2. Traditional Neural Network Model

2.2.1. Establishment of the FCN model

The utilization of FCN enables the classification of individual pixels and the subsequent completion of image segmentation. The fully connected layer at the conclusion of the network is sometimes referred to as CNN. The feature graph is transformed into one-dimensional vectors using the fully connected layer, and subsequently classified by computing the ultimate probability. In contrast to CNN, FCN incorporates convolutional layers throughout its architecture, allowing for pixel-wise classification capabilities. FCN eliminates the need for the link layer, hence rendering the network independent of the input image size. The level of abstraction in Fully Convolutional Networks (FCNs) is positively correlated with the depth of the network, whereas shallower feature graphs contain more detailed information. The utilization of both fine-grained information from shallow feature networks and semantic information from deep feature graphs in semantic segmentation leads to improved accuracy in segmentation. Figure 1 shows the network architecture of FCN, where in the final fully connected layer of the conventional CNN is omitted.

The Fully Convolutional Network (FCN) is composed of a series of convolutional layers and pooling layers that are coupled in a sequential manner. Following each pooling process, the dimensions of the feature map are halved, resulting in a reduction in size compared to the original. Additionally, the receptive field of the corresponding pixel is doubled in relation to the original. Hence, it can be inferred that every pixel within the fifth layer encapsulates the semantic details pertaining to a specific region of the original image, measuring 32 by 32 in dimensions. The Fusion Convolutional Network (FuCNN) can be categorized into three distinct classical models, which are determined by the feature graph information of varying dimensions in the fusion convolutional layers. The FCN-32s architecture does not incorporate the shallow features in the convolution layer. Instead, it employs direct upsampling of the final output feature map to restore it to the original image size, hence facilitating the completion of the segmentation process. The FCN-16s model integrates the feature graph produced by the network with the feature graph derived from the fourth convolutional layer, resulting in the

Figure 1
FCN network structures



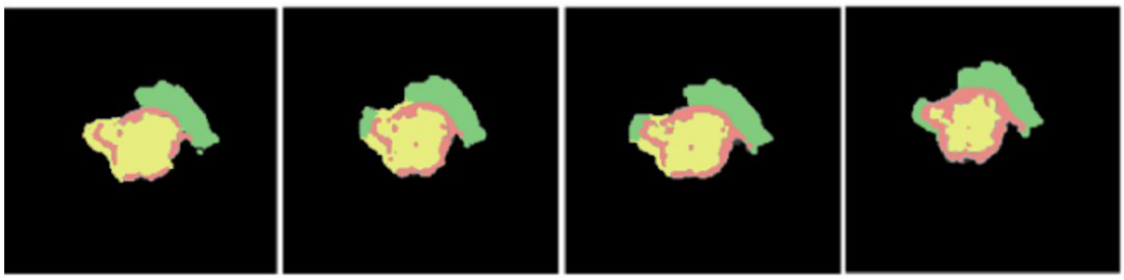
generation of a novel feature graph. The fusion mode involves the summation of pixels at related positions in the feature graph, which is subsequently restored to the original picture size through upsampling in order to accomplish segmentation. Like FCN-16s, FCN-8s incorporates the feature map information from both the fourth and third layers. This integration allows for a more comprehensive utilization of detailed and semantic information from the feature map, resulting in enhanced fitting capabilities of FCN.

This paper will use FCN in conjunction with various upsampling and fusion techniques to perform image segmentation of brain tumors. Figure 2 presented below depicts the partial segmentation results of the three models, with the rightmost figure representing the manually annotated gold standard.

The evolution of Fully Convolutional Networks (FCNs) from FCN-32s to FCN-16s, and subsequently to FCN-8s, epitomizes a sophisticated advancement in the architecture of neural networks for semantic image segmentation tasks. These variants exemplify a strategic progression in mitigating the inherent trade-offs between capturing high-level semantic content and preserving spatial detail crucial for precise boundary delineation.

The FCN-32s architecture, the foundational model among the three, generates feature maps with a considerable stride of 32 pixels. This configuration facilitates the encapsulation of high-level semantic information across expansive contexts, attributable to the extensive receptive fields of its feature maps. However, this approach invariably compromises the precision of

Figure 2
The partial segmentation results of the three models. (a) FCN-32s; (b) FCN-16s; (c) FCN-8s; (d) Manually annotated gold standard. Red represents necrotic and non-enhancing glioma, green represents peritumoral edema, and yellow represents enhancing glioma



boundary definitions within the segmentation output, primarily due to the relatively coarse nature of the feature maps, which lack granular spatial details.

In an endeavor to surmount this limitation, the FCN-16s model integrates semantic information from the feature maps of the final layer (characterized by a stride of 32 pixels) with spatial information derived from the upsampled feature maps of the pool4 layer, which exhibits a reduced stride of 16 pixels. This amalgamation harnesses the strengths of both semantically enriched features and localized, spatially detailed features, thereby enhancing the accuracy of boundary localization beyond the capabilities of FCN-32s through the inclusion of additional spatial information from an intermediary network layer.

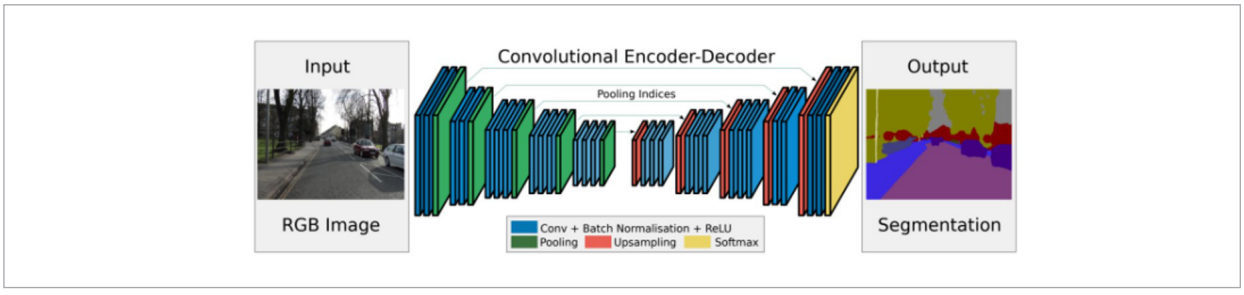
Further refinement is achieved with the FCN-8s model, which amalgamates feature maps from the ultimate layer, the upsampled pool4, and further upsampled pool3 layers at a minimal stride of 8 pixels. This strategy introduces an even higher resolution of spatial details into the output of the model, facilitating a segmentation output that more closely approximates the manually annotated gold standard. By leveraging information from shallower feature maps, FCN-8s markedly improves the fitting capabilities of the model, resulting in superior segmentation outcomes in comparison to both FCN-32s and FCN-16s.

2.2.2 Establishment of the SegNet Model

The University of Cambridge team has suggested an open-source project called SegNet, which facilitates pixel-level image segmentation. Figure 3 illustrates the architecture of SegNet [1], which has two main components: encoding and decoding. The left half of the structure encompasses the coding process, which involves the utilization of a convolution layer and a pooling layer. These layers are capable of extracting intricate information and local features from the image. Additionally, the pooling layer performs a downsampling operation, which effectively enlarges the receptive field of the feature map. On the other hand, the right half of the structure involves the decoding process, which comprises the utilization of an upsampling. The process of upsampling in the upper sampling stage aims to restore the feature map to the dimensions of the original input image. This is important because during the downsampling process, certain information may have been lost. To address this, the convolution layer in the network is designed to learn and capture the missing details. In order to maximize the utilization of shallow feature maps, the author included a crucial innovation in SegNet by incorporating pooling index connections between the encoding and decoding layers at matching positions. This addition plays a significant role in enhancing the overall performance of the model. The proposed methodology involves preserving the positional information associated with the highest value in the encoding layer, and subsequently reinstating this positional information during the upsampling phase of the decoding layer. In contrast to the deconvolution technique used in FCN upsampling, the pooling index immediately samples the data value at the appropriate index. This process requires minimal storage space, without the need for learning. Consequently, it minimizes the number of model parameters and enhances the models' capacity to fit boundaries.

There are no fully connected layers and hence it is only convolutional. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map(s). It then performs convolution with a trainable filter bank to densify the

Figure 3
An illustration of the SegNet architecture



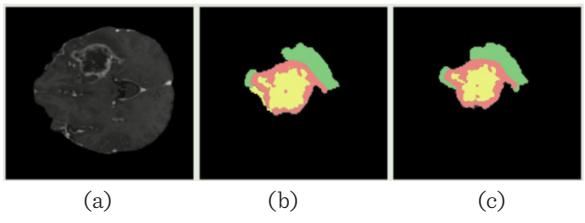
feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification [1].

The architecture diagram in Figure 3 includes direct connection lines from the input to the output, signifying the implementation of skip connections within the convolutional encoder-decoder framework. These skip connections are employed to mitigate the spatial information loss that occurs during the pooling operations in the encoder. By transferring feature maps from various encoder layers to the corresponding decoder layers, the network is able to utilize both high-level semantic information and low-level textural details. This dual usage of features facilitates the reconstruction of fine-grained spatial details in the output segmentation map, resulting in a more precise delineation of object boundaries. The inclusion of pooling indices along these connections further enhances the upsampling process, allowing the decoder to more accurately reconstruct the spatial hierarchy of the input image.

In this work, SegNet was employed to perform tests pertaining to the segmentation of brain tumors. Figure 4(a) displays the magnetic resonance imaging (MRI) part of the brain tumor T1 modality, and Figure 4(b) shows the segmentation results of this network are presented, while Figure 4(c) represents the gold standard of hand annotation for educational purposes. It demonstrates that the SegNet model exhibits more accuracy in segmentation findings when compared to FCN-8s. Notably, the SegNet model achieves more exact boundaries between different tumor subregions, particularly in the peripheral edema region. This improvement in delineating the overall shape and detail enhances the performance of the SegNet model. Nevertheless, the segmentation performance of SegNet is not perfect in certain tumor locations and exhibits segmentation mistakes in specific subregions, likely due to the intricate and diverse nature of brain tumor images.

Figure 4

(a) MRI slice of a brain tumor in T1 mode; (b) The segmentation results of this network are presented; (c) Manual annotation gold standard for learning



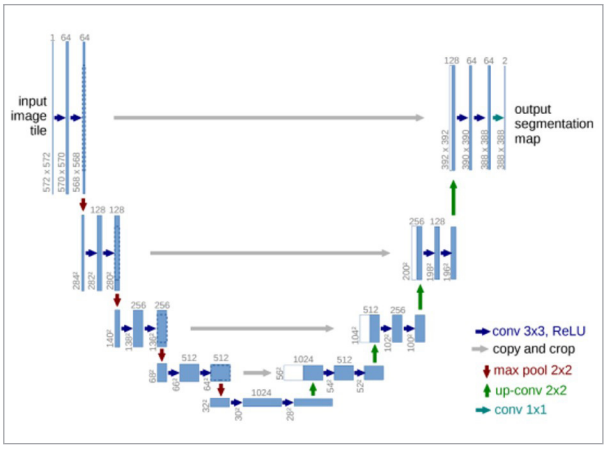
(a) (b) (c)

2.2.3. U-net Model Building

The schematic representation of the U-net architecture is depicted in Figure 5. In the architecture of U-net, the process of feature fusion, particularly exemplified by the technique of "feature splicing" represents a critical innovation. This process is characterized by dynamically adjusting the number of channels, specifically by doubling the feature channels during the upsampling phase, thereby enhancing the capability on the network to process and interpret complex information.

Figure 5

U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations



The process of feature fusion within the architecture of U-net encompasses two fundamental phases: downsampling (encoding) and upsampling (decoding), each serving a distinct purpose in the operation of the network. The downsampling phase is characterized by the ability of the network to extract and refine information from the input image. This phase is critical for the capacity of the network to identify and retain essential details and textures, thereby facilitat-

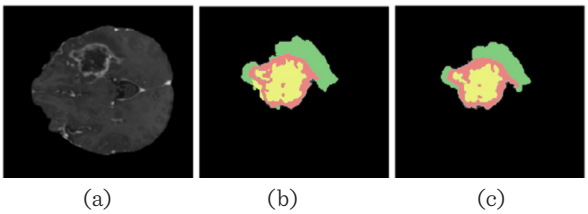
ing a granular comprehension of the data. Through the strategic reduction of data dimensions, the network efficiently focuses on salient features, enabling a more profound analysis and understanding of the composition of the input image. Conversely, the upsampling phase is dedicated to the reconstruction of the image from its processed, condensed state. The objective during this stage is to accurately predict the final output, such as a segmented version of the original image. Central to this phase is the mechanism of feature fusion—also referred to in prior discussions as "feature splicing." This mechanism is integral to the ability of the network to amalgamate and refine the distilled information obtained during downsampling with the goal of generating a coherent and detailed output.

This dual-phase approach, consisting of both downsampling and upsampling, is foundational to the U-net architecture's efficacy in performing complex image processing tasks, such as segmentation. The strategic interleaving of feature extraction and reintegration through feature fusion is pivotal, enabling the network to achieve high precision in its output, thereby underscoring the significance of each phase in the overall operation of the U-net architecture.

The U-net model was utilized in this study to conduct research on brain tumor image segmentation. Figure 6 displays the segmentation sample of U-net, where the number of initial convolution channels is configured to be 64, denoted as U-net (64) throughout the training process. The brain images in the left-to-right sequence consist of the raw images, the segmentation results obtained using U-net (64), and the conventional segmentation results that were manually annotated.

Figure 6

Segmentation Sample of U-net. (a) The original brain images; (b) U-net(64) segmentation results; (c) The manually annotated standard segmentation results



(a) (b) (c)

When comparing the segmentation results of FCN and SegNet with U-net, it is observed that U-net exhibits improved segmentation effects for necrotic and non-enhancing glioma, peritumoral edema, and enhancing glioma. This improvement can be attributed to U-net having a greater number of characteristic channels. U-net demonstrates a more precise and accurate boundary delineation between subregions, and it possesses specific segmentation fitting ability for both the entire tumor and local tiny structures within tumor subregions. Despite the great segmentation accuracy exhibited by U-net, the model does exhibit certain segmentation mistakes in localized regions. These errors can be attributed to the constraints imposed by the small dataset and the loss of tumor multimodal data information.

2.2.4. CAM and SAM

Channel Attention Module (CAM) and Spatial Attention Module (SAM) have emerged as significant advancements in the realm of deep learning research, particularly in the enhancement of image processing tasks. CAM is designed to prioritize different channels within the input feature maps, thus augmenting the responsiveness of the model to channels that contain more informative attributes. In contrast, SAM is engineered to optimize the utilization of spatial location information. It accomplishes this by assigning varied weights to different spatial locations within the input feature maps, thereby accentuating crucial spatial characteristics. This enhancement is crucial for improving performance in the intricate task of segmenting brain tumor images, where precision in identifying and delineating tumor boundaries is paramount.

3. Method

In recent years, numerous methodologies have been proposed and implemented with the aim of enhancing the precision of medical image segmentation. The aforementioned techniques encompass the utilization of enhanced algorithms, augmentation of the training dataset size, and enhancement of the training dataset quality.

Through the examination of segmentation experiments conducted on conventional classical full convolutional network models, it was shown that the comprehensive convolutional neural network

exhibits a distinct capability in achieving high segmentation accuracy for brain tumor segmentation tasks. The FCN-8s architecture incorporates the fusion of shallow and deep feature maps in order to enhance the capacity of the model for fitting and afterwards achieve semantic segmentation of brain tumor images. In the process of upsampling in SegNet, the incorporation of position information from the shallow pooling process results in a reduction in the parameters of the model while only slightly increasing storage space. This enhancement improves the learning efficiency of the model and enhances its ability to differentiate between boundaries in distinct subregions. The U-net architecture, which is a variant of the conventional convolutional neural network model, demonstrates superior segmentation performance. It employs a distinctive approach for feature concatenation and fusion, wherein the number of feature channels is augmented during the upsampling phase. This enables the utilization of more intricate information from shallow feature maps, thereby facilitating effective image segmentation.

Nevertheless, there is still a need for improvement in the segmentation result of the above three models. Specifically, there are issues with the accuracy of local subregion segmentation and the classification of boundaries. Upon thorough review, it has been determined that there are three distinct reasons.

1 The conventional approach for segmenting brain tumors in multimodal data involves feeding the multimodal data into CNN. However, during the fusion of multimodal input data in the first convolutional layer, there is a substantial loss of information. This results in a limited ability of the network to differentiate between different modes of data, thereby preventing the full utilization of all four modalities. To achieve high-quality segmentation results, it is crucial to fully exploit the information provided by the multimodal data.
2 The availability of medical-labeled data is limited, and the convolutional neural network requires a substantial volume of data to assure effective training of the network. The segmentation accuracy of a neural network is expected to exhibit a notable enhancement with the effective expansion of the data.
3 The presence of brain tumor images exhibits a significant disparity in class distribution. Table 1 presents the relative proportions of several categories, namely background (label0), necrotic and non-enhanced gliomas (label1), peritumoral edema (label2), and enhanced gliomas (label4), inside the Brats2019 dataset.

Table 1
The proportion of labels of different tumor categories to total pixels (Brats2019)

Tumor Categories	Label0	Label1	Label2	Label4
HGG	0.98934	0.00155	0.00658	0.00253
LGG	0.9875	0.00587	0.006	0.00061
Total brain glioma	0.98893	0.00253	0.00645	0.00209

Valuable insights can be derived from the label percentage table pertaining to various tumor classifications. In the study conducted on HGG, it was seen that the proportion of label0 was almost 99%. The remaining tumor pixels were found to be limited in quantity. Additionally, the ratio of background pixels to tumor pixels was determined to be 92.46:1. Similar to high-grade gliomas (HGG), the label0 of low-grade gliomas (LGG) achieved an accuracy of 98.75%, with the remaining pixels representing tumor tissue. In both HGG and LGG, label0 continues to dominate the entire tumor image, accounting for 98.89% of the total. Furthermore, a noticeable disparity in class distribution was seen among the three distinct tumor subregions. Within the context of high-grade gliomas (HGG), the tumor subregions designated as label1, label2, and label4 accounted for 14.5% and 61.7% of the total tumor volume, respectively. In low-grade gliomas (LGG), these three tumor subregions constituted 47%, 48.1%, and 4.9% of the overall tumor volume, respectively, with proportions of 22.9%, 58.3%, and 18.9% of the total tumor volume. Based on the findings of the research, it is evident that there is a considerable magnitude in both the proportion of tumor and background, as well as the proportion among different subregions of the tumor. Training the model with such data may result in an overabundance of background information being learned, hence potentially hindering the complete acquisition of tumor-related

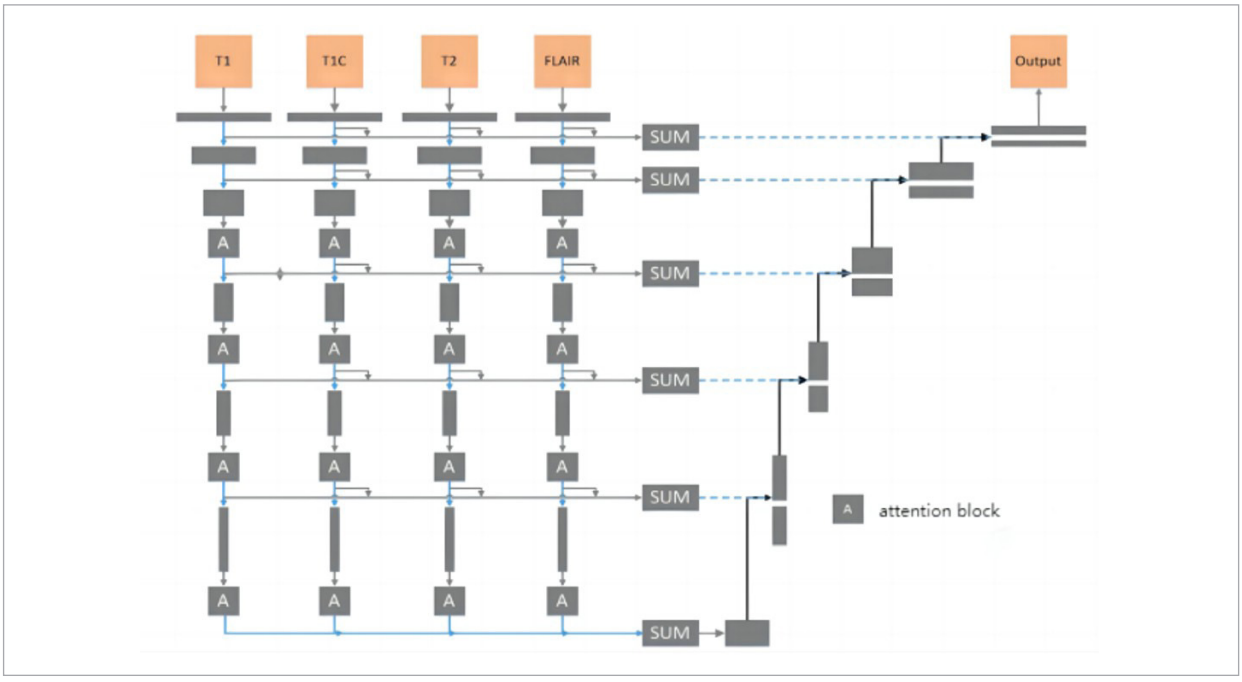
information. As a consequence of the aforementioned challenges, the U-net, which exhibits superior segmentation capabilities within the conventional convolutional neural network framework, is nevertheless prone to some segmentation inaccuracies. Consequently, this study proposes novel convolutional neural networks with the aim of enhancing the segmentation performance of the network.

In order to address the issue of information loss in multimodal data, we have developed a multi-branch network architecture inspired by the U-net model. It is shown as Figure 7. The coding portion encompasses various branches, each serving the purpose of processing data derived from distinct modalities. Every branch possesses a same framework while exhibiting distinct parameters. The decoding component of the system likewise employs the U-net architecture, enabling the effective integration of both high-level semantic information and low-level detail information. In order to enhance the extraction of features and effectively utilize the global context information of the image, an attention module was incorporated into each branch of the encoding component. In this study, an attention mechanism is incorporated into the four convolutions within the encoding component in order to enhance the segmentation performance of the model.

In order to address the issue of inadequate data volume, a data-enhanced network model called TS-U-net (Teacher Student U-net) was developed. Conventional techniques for data augmentation, such as mirror imaging and scale transformations, solely induce deformations in the original data without substantially expanding the dataset. In addition to exclusively utilizing the available data information, the TS-U-net model has the capability to incorporate additional data through the utilization of pseudo-labels. The procedures employed in this research are outlined below:

1. The data is categorized into two distinct groups: labelled data and unlabelled data.
2. Train a U-net as a teacher network using data with labels.
3. After the completion of training the teacher network, the unlabelled data is fed into the teacher network to generate data with inaccurate labels.
4. Subsequently, a new U-net network, referred to as the student network, is retrained using a combination of labelled data and pseudo-label data from the previous step.

Figure 7
Multi-branch network architecture



5 Following the training of the student network, it is considered as a new teacher network. The unlabelled data is then input into the student network to generate new data with inaccurate labels.

6 This process of steps 2 to 5 is iterated until there is no further improvement in the accuracy of segmentation.

The aforementioned data augmentation approach allows for the generation of several data instances with inaccurate labels based on the available unlabeled data. In theoretical terms, the potential for unlimited expansion of data presents a partial resolution to the issue of limited availability of medical picture data. To address the issue of class imbalance in tumor labelling, we employed the weighted cross-entropy as the loss function. Weighting refers to the process of assigning varying weights to different categories based on their relative importance or significance. Greater emphasis will be placed on categories that possess higher weights during the process of network training. Typically, in statistical analysis, weights assigned to small sample sizes are increased, whereas weights assigned to big sample sizes are decreased.

In addressing the challenge of class imbalance within the BraTS2019 dataset, the relative weights for the loss function were judiciously calibrated to amplify the influence of underrepresented classes. The background class (Label0), which constitutes the majority of the dataset, was assigned a baseline weight of 1. The weights for necrotic and non-enhanced glioma (Label1), peritumoral edema (Label2), and enhanced glioma (Label4) were proportionately escalated to 391, 153, and 473, respectively. This weighting scheme was devised by inverting the prevalence of each class to yield a set of preliminary weights, which were then normalized against the weight of the background class to preserve computational balance. Through this methodology, the sensitivity of the model to rarer classes is enhanced, encouraging a more equitable distribution of the predictive focus and an improved segmentation performance across all classes.

4. Experiment Results and Analysis

4.1. Experimental Data

4.1.1. Data Set

Our work uses data from the MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS2019) to conduct glioma segmentation and prognostic analysis. The training dataset consisted of 335 patients, with 76 samples from patients diagnosed with low-grade glioma and 259 samples from patients diagnosed with high-grade glioma. The validation set consisted of 125 patient samples that were devoid of any labels. All patients included in the analysis for predicting overall survival have provided their clinical age measurements.

Each example comprises four modalities of three-dimensional magnetic resonance imaging (MRI) data, including the T1-weighted image (post-contrast), T1-weighted image, T2-weighted image, and liquid decay inversion recovery sequence. These diverse methods offer differing levels of attention on distinct subregions of brain tumors. Each instance is accompanied by an annotation, which is methodically performed by professional physicians using manual procedures. Moreover, each annotation is subjected to a meticulous evaluation process in order to ensure its precision. Each unique image of a brain tumor was annotated with four distinct parts, which include the background as well as three specific subregions of the tumor within the brain. The subregions encompass the following components: the backdrop (designated as 0), necrotic and non-enhanced glioma (designated as 1), peritumoral edema (designated as 2), and enhanced glioma (designated as 4). The entire understanding of malignancies cannot be adequately represented by data obtained from a single modality. The segmentation model developed in this study integrates four modalities, namely T1, T1C, T2, and FLAIR, as input data to assess the impact of brain tumors.

4.1.2. Data Processing

In order to address the issue of the original MR image's large size, this study employs a resizing technique to reduce its dimensions to 128 * 160 * 192, based on the brain volume. This resizing process serves two purposes: firstly, it allows for the removal of redundant information (specifically, non-brain information represented by intensity value 0), and secondly, it ensures that the resulting image block volume is compatible with the available computer video memory. The imaging results of brain tumor patients can be influenced by various factors, including the features of the imaging machines used and the contrast agents administered. Consequently, it is imperative to preprocess

the multimodal MRI data to ensure a harmonized data distribution and mitigate any disparities. Initially, the three-dimensional (3D) data inside the dataset is transformed into two-dimensional (2D) data with dimensions of 240 * 240 along the horizontal axis. Subsequently, the data lacking tumor information is removed to prevent the inclusion of background data in the training process, which could potentially impact the efficacy of the learning of the model. The conventional normalization process was applied to each image using the formula (3-1) subtracted by the mean and then divided by the standard deviation.

$$V_{out}^{k}(x,y,z)=\frac{V_{in}^{k}(x,y,z)-mean(I^{k})}{std(I^{k})} \quad (1)$$

4.2. Experimental Environment and Setting

The Pytorch deep learning framework was used in the UBUNTU environment and trained on an NVIDIA 1080ti GPU equipped with 16G RAM, adding Adam, the adaptive moment estimation, as the optimization function.

4.3. Evaluation Indicators

The input of the network consists of four channels, each representing the MRI data of one of the four modalities. The slice data of each modality is spatially aligned with the individual case and the 3D brain tumor MRI. The initial learning rate is initialized at 10-7, and subsequently, the learning rate undergoes a progressive increase during the training process of the network. In instances where the model convergence is sluggish, it is common practice to augment the learning rate by a factor of 5-10 until the model ceases to converge. The metric used to assess the accuracy of segmentation is the dice coefficient, which is calculated for each person and then averaged across all individuals. The evaluation indexes for enhanced glioma (Enhance Tumor, ET), glioma nucleus (Tumor Core, TC), and all glioma (Whole Tumor, WT) were computed using the dice coefficient. The classification of gliomas encompasses various types, including necrotic, non-enhancing, and enhancing gliomas. Additionally, the term "WT" refers to the inclusion of all tumor subregions within this classification.

In the process of partitioning the dataset for model training and evaluation, a meticulous stratification strategy was employed to accommodate the multimodal composition inherent within the dataset. Each constituent data point is an aggregation of spatially aligned slices derived from four distinct MRI modalities: T1, T1C, T2, and FLAIR. Each modality contributes uniquely to the characterisation of tissue properties. To safeguard the integrity of the multimodal data throughout the training, validation, and testing cohorts, we ensured that a complete set of modality slices for each individual case was preserved during the randomization process. This deliberate approach guarantees the exposure of the model to a consistent and holistic multimodal dataset, thereby mirroring the intricate and heterogeneous nature of brain tumors as encountered in authentic clinical environments. Such a strategy is pivotal in preserving the rigour of the evaluation of the model and the integrity of the dataset.

During the training process, a total of 335 data points were randomly partitioned into groups. Out of these, 285 data points were assigned as validation sets, while the remaining 50 data points were designated as test sets. To ensure the elimination of randomness, the results were averaged. Table II presents the results of various conventional convolutional neural network segmentation techniques. The network segmentation results of FCN-16s and FCN-32s exhibited suboptimal performance and did not contribute to the contrast analysis. Simultaneously, we conducted a comparative analysis of the impact of four distinct groups of initial convolution layers on the segmentation results of the U-net model. Specifically, U-net (8) denotes the utilization of an initial convolution layer with a size of 8, while U-net (16) signifies the use of an initial convolution layer with a size of 16, and so forth.

We designed a series of ablation experiments to evaluate the impact of CAM, SAM, and their combination on the performance of our modified U-net model. We integrated CAM and SAM into the encoder and decoder parts of the U-net, respectively, and compared the performance of models under the following configurations: only CAM, only SAM, both CAM and SAM, and a baseline model without any attention modules.

Our experimental results indicate that the inclusion of CAM and SAM significantly improves the accuracy of brain tumor image segmentation. Specifically, models utilizing CAM showed an average increase of 0.01% in the Dice coefficient compared to the baseline

Table 2
Traditional convolutional neural network segmentation results

Sub-Tumour	SegNet	U-net(8)	U-net(16)	U-net(32)	U-net(64)
WT	85.67	86.56	86.67	87.43	87.5
TC	82.13	83.69	84.23	85.23	86.02
ET	81.62	80.58	81.43	81.19	81.39

model, while models with SAM exhibited an improvement of 0.01%. Furthermore, when both CAM and SAM were applied, the performance of the model enhanced further, with an average increase of 0.03% in the Dice coefficient, confirming the complementary nature of these two attention mechanisms in improving the ability of the model to capture tumor features.

Based on the data presented in Table 2, it can be inferred that among the classic neural network models, U-net exhibits the greatest dice coefficient and achieves the most effective segmentation. SegNet follows with a moderately effective segmentation, while FCN has comparatively poorer segmentation performance. Furthermore, the precision of the U-net model will be enhanced as the number of initial convolutional layers is augmented. Nevertheless, the expansion of the number of initial convolutional layers is constrained by the finite computing and memory resources available. The subsequent multi-branch TS-U-net architecture is derived from the U-net (64) model.

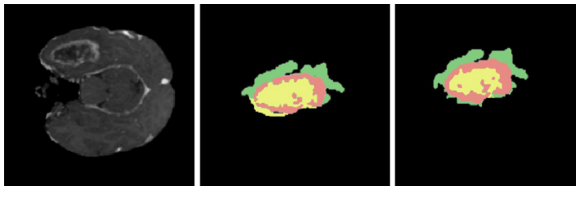
As presented in Table 3, the comparative efficacy of the U-net(64) and the innovative Multi-branch U-net architectures is evaluated through the Dice coefficient metric for segmenting sub-tumoral regions. The U-net(64) achieves a Dice score of 87.5 for Whole Tumor (WT) segmentation, which is marginally outperformed by the Multi-branch U-net, attaining a score of 87.6. The segmentation accuracy for Tumor Core (TC) is notably improved with the Multi-branch U-net, which achieves a Dice score of 86.8, compared to the U-net(64)'s 86.02. For the Enhanced Tumor (ET) category, both models demonstrate comparable proficiency, with the U-net(64) achieving a score of 81.39 and the Multi-branch U-net a slightly superior score of 81.4. These results suggest a consistent, albeit slight, superiority of the Multi-branch U-net in segmenting various tumor subregions, potentially indicative of its superior capability to accurately delineate the complex pathology of gliomas.

Table 3
Dice between U-net(64) and our method

Sub-Tumor	U-net(64)	Multi-branch U-net
WT	87.5	87.6
TC	86.02	86.8
ET	81.39	81.4

Figure 8 illustrates the output results a multi-branch TS-U-net, alongside the standard results of manual annotation. The original brain images of the output results of the multi-branch TS-U-net are presented from left to right. The model is designed to differentiate between various tumor tissues, and the segmentation is depicted in different grayscale intensities. The lighter areas represent the enhancing tumor, while the darker regions indicate non-enhancing tumor or necrotic tissue. Figure 8(c) provides the manual segmentation performed by clinical experts, serving as the ground truth against which the performance of the model is compared. It is evident that while human brain tumor forms vary, there is disparity in the proportion of tumor sub-regions. This section presents the design of a multi-branch TS-U-net network model for the

Figure 8
Detailed segmentation analysis of brain tumors. (a) The original brain images; (b) The output results of the multi-branch TS-U-net; (c)The manually annotated standard segmentation results



segmentation of different tumor sub-regions and the precise delineation of fine structures. The results obtained from the network output are found to be highly consistent with the manually annotated gold standard, indicating the promising potential of this model in the domain of medical image semantic segmentation.

In order to enhance the credibility of the multi-branched TS-U-net model's segmentation accuracy inside the identical dataset, a comparison was conducted with the segmentation results of the top-ranking validation set from the Brats 2019.

Table 4
Dice between the first of Brats2019 and our method

Sub-Tumor	The top of Brats2019	Multi-branch U-net
WT	91	87.6
TC	86.68	86.8
ET	82.33	81.4

5. Discussion

In this paper, we conduct a comprehensive analysis and evaluation of various established traditional fully convolutional neural networks. Our focus is on their performance in the context of a 2D brain tumor segmentation task. The results of the study indicate that U-net demonstrates superior performance compared to other fully convolutional neural networks in the task of semantic segmentation of medical images. Specifically, U-net achieved dice coefficients of 87.6, 86.8, and 81.4 for the WT, TC, and ET regions, respectively.

Moreover, our work introduces a novel approach called multi-branch TS-U-net, which aims to overcome some limitations seen in the conventional U-net architecture. The network addresses the issue of inadequate medical image data by employing a robust data augmentation technique. It tackles the problem of imbalanced brain tumor image classes by utilizing weighted cross entropy as a loss function. Additionally, it incorporates a multi-branch attention mechanism to mitigate the loss of multimodal data and enhance segmentation accuracy. The dice coefficients for the multiple branch TS-U-net model were 0.876, 0.868, and 0.814 in the tumor subregions of WT, TC, and ET, respectively.

Acknowledgement

This research was supported in part by the General Research Project of Education of Zhejiang Province under Grant Y202044894, and supported in part by Basic Scientific Research Project of Wenzhou Science and Technology Bureau under Grant Y2023148, and supported in part by Wenzhou Cyber Security Detection and Protection Technology Research Center.

Author Contributions

The authors confirm contribution to the paper as follows: conceptualization, methodology, writing original draft, review and editing: Yajie Zhang, Hea Choon Ngo; validation and visualization of results: Yifan Zhang, Noor Fazilla Abd Yusof; draft manuscript preparation: Yifan Zhang, Xiaohan Wang; supervision: Hea Choon Ngo. All authors reviewed the results and approved the final version of the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12), 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615
2. Dwivedi, T., Gupta, A. A simplified overview of World Health Organization Classification Update of Central Nervous System Tumors 2016. Journal of Neurosciences in Rural Practice, 2017, 8(4), 629. https://doi.org/10.4103/jnrp.jnrp_168_17
3. Glass, J., Hochberg, F. H., Gruber, M. L., Louis, D. N., Smith, D., Rattner, B. The Treatment of Oligodendrogliomas and Mixed Oligodendroglioma-astrocytomas aith PCV Chemotherapy. Journal of Neurosurgery, 1992, 76(5), 741-745. https://doi.org/10.3171/jns.1992.76.5.0741
4. Holland, E. C. Progenitor Cells and Glioma Formation. Current Opinion in Neurology, 2001, 14(6), 683-688. https://doi.org/10.1097/00019052-200112000-00002
5. Işın, A., Direkoğlu, C., Şah, M. Review of MRI-based Brain Tumor Image Segmentation Using Deep Learn-

ing Methods. Procedia Computer Science, 2016, 102, 317-324. https://doi.org/10.1016/j.procs.2016.09.407

6. Liu, L., Zhang, H., Rekik, I., Chen, X., Wang, Q., Shen, D. Outcome Prediction for Patient with High-Grade Gliomas from Brain Functional and Structural Networks. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016, 2026, 26-34. https://doi.org/10.1007/978-3-319-46723-8_4
7. Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., Ohgaki, H., Wiestler, O. D., Kleihues, P., Ellison, D. W. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A Summary. Acta Neuropathologica, 2016, 131(6), 803-820. https://doi.org/10.1007/s00401-016-1545-1
8. Milletari, F., Navab, N., Ahmadi, S. A. V-net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In 2016 Fourth International Conference on 3D vision (3DV), 2016, 565-571. https://doi.org/10.1109/3DV.2016.79
9. Ostrom, Q. T., Bauchet, L., Davis, F. G., Deltour, I., Fisher, J. L., Langer, C. E., Pekmezci, M., Schwartzbaum, J. A., Turner, M. C., Walsh, K. M., Wrensch, M. R., Barnholtz-Sloan, J. S. The Epidemiology of Glioma in Adults: a "State of the Science" Review. Neuro-Oncology, 2014, 16(7), 896-913. https://doi.org/10.1093/neuonc/nou087
10. Ronneberger, O., Fischer, P., Brox, T. 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, 234-241. Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28