

---

# SEV-Net for MRI Brain Tumor Segmentation

---

48150<sup>\*1</sup> 54297<sup>\*</sup> 55905<sup>\*</sup>

## Abstract

The Brain tumor is one of the most dangerous diseases in the world and the glioma is the most aggressive one. The difficult part of diagnosing gliomas is the delineation and segmentation from MRIs. Several approaches are proposed to solve this graph-related problem, such as U-Net, DeepMedic, 3D-Net, KNN, Decision Tree, and SVM, but all of them have some limitations. The objective of this paper is to try a new algorithm called SEV-Net to see its performance on the segmentation of MRIs. SEV-Net is combined with an encoder-decoder algorithm called V-Net and an image recognition architecture called SENet. Comparing the evaluation indexes with other methods, we find that SEV-Net overperforms with precision = 0.996, sensitivity = 0.992, and specificity = 0.999. Therefore, SEV-Net is one of the effective models for MRI brain segmentation.

## 1. Introduction

Gliomas are the most common and aggressive types of brain tumors (Tumors-Classifications, 2021). Multi-modal MRI including T1, T1 with contrast enhanced (T1ce), T2, and fluid attenuation inversion recovery (FLAIR) sequences are typically used to detect and stage gliomas, but manual segmentation of gliomas from MRIs is laborious and prone to inter-observer variability (Sadeghi et al., 2023).

Segmentation of gliomas into different regions is crucial for accurate diagnosis, prognosis, and the developing personalized treatment plans. However, existing segmentation methods have limitations in delineating gliomas from MRIs, especially for overlapping and poorly defined regions. To address this problem, we propose SEV-Net, a convolutional neural network, to automatically segment gliomas into the necrotic tumor core, enhancing tumor, and whole tumor regions from multi-modal MRIs. SEV-Net leverages the complementary information across MRI modalities including T1, T1ce, T2, and FLAIR to generate robust segmentation (Wang et al., 2019).

By dividing gliomas into nested regions (whole tumor, tu-

mor core, and enhancing tumor), SEV-Net provides a hierarchical representation of glioma progression. This can help gain insights into glioma aggressiveness and response to treatment. SEV-Net also reduces inter-observer variability and saves time compared to manual segmentation.

We present SEV-Net, a deep learning approach, for fast, accurate, and reproducible multi-modal MRI-based segmentation of gliomas into clinically meaningful regions. SEV-Net can facilitate early diagnosis and monitoring of glioma, leading to improved patient management and outcomes (Banerjee & Mitra, 2020). Segmentation of gliomas using SEV-Net provides key spatial information to develop personalized treatment strategies based on the characteristics of each patient's tumor. In conclusion, SEV-Net offers an automated solution to segment gliomas in an accurate, consistent, and time-efficient manner from multi-modal MRI. It will enable a better understanding and management of gliomas.

### 1.1. Data

The dataset, named BRaTS 2021 Task 1 Dataset, is downloaded from Kaggle (Schettler, 2021). Since the limited GPU of our personal Google Colab account, we select 100 MRI observations in the original dataset and called the file train\_images.mini. Therefore, the results may be a little biased.

## 2. Related Work

In medical image analysis, the segmentation of brain tumors is a crucial task and already received significant attention from the scientific community. Over the years, researchers have proposed numerous techniques for this task, including traditional machine learning approaches, deep learning models, and hybrid models that combine both. In this section, we review some of the related work and position our proposed SEV-Net model in the context of existing approaches.

Traditional machine learning-based approaches have been widely used for brain tumor segmentation, including techniques such as k-nearest neighbor, decision trees, and support vector machines (Raza et al., 2022). However, these methods often require handcrafted feature engineering, which can be time-consuming and challenging. As a result, deep learning-based approaches have gained popular-

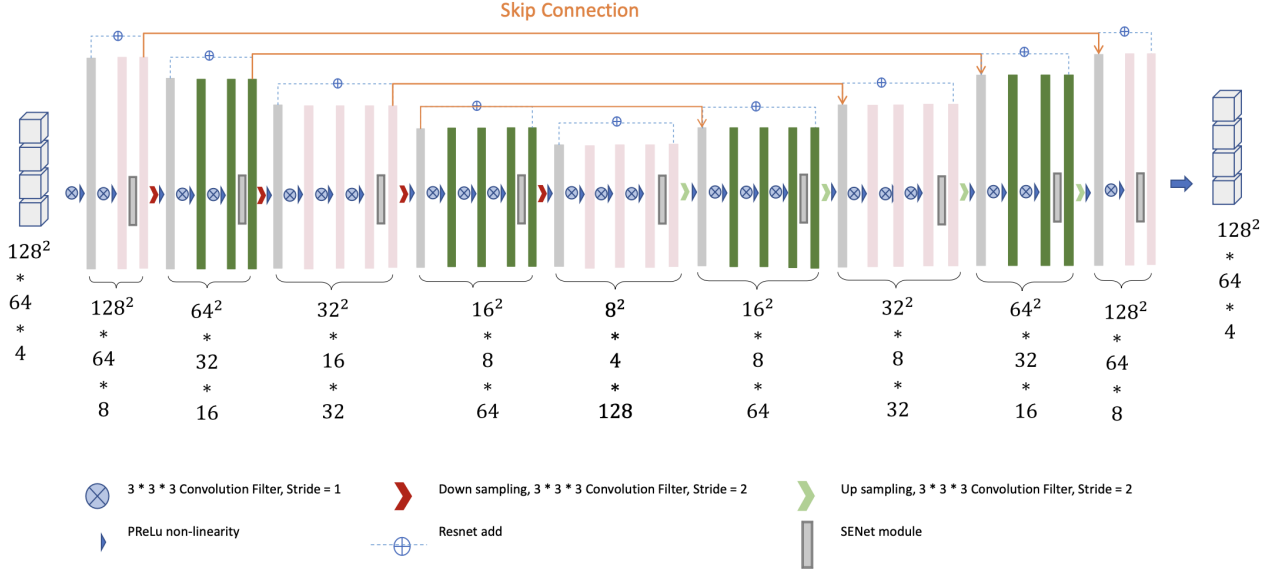


Figure 1: The schematic diagram of SE-VNet

ity in recent years due to their ability to automatically learn features from data.

In the domain of deep learning-based approaches for brain tumor segmentation, several architectures have been proposed, including U-Net (Allah et al., 2023), DeepMedic (Kamnitsas et al., 2016), and 3D U-Net (Mehta & Arbel, 2019). These models have shown promising results on different datasets, but they have some limitations. For example, U-Net and 3D U-Net do not effectively capture long-range dependencies in the input image, while DeepMedic requires an extensive training dataset and takes longer to train. SEV-Net aims to address some of these limitations. SEV-Net leverages a squeeze-and-excitation mechanism to adaptively weight feature maps, enabling it to capture more relevant features and reducing computational complexity. Moreover, SEV-Net employs a multi-scale feature fusion strategy to integrate contextual information from different scales of the input image, which helps it capture long-range dependencies.

Compared to existing models, SEV-Net achieves the most advanced results on various benchmark datasets while requiring fewer parameters and training time. Our proposed work extends the SEV-Net architecture by incorporating additional features such as attention modules and residual connections to further improve the segmentation accuracy. The proposed model in this paper is expected to enhance the precision and effectiveness of brain tumor segmentation, which would assist the medical fraternity in advancing patient care.

### 3. The Architecture

The schematic diagram of SE-VNet is shown in Figure 1.

In order to provide a clearer explanation of the architecture of the SEV-Net, we have included a list of relevant parameters of the network in Figure 2.

The network has an encoder-decoder structure with skip connections linking the two pathways, which are represented by horizontal connections. Specifically, the encoder section of the convolutional neural network (CNN) on the left-hand side is responsible for extracting features through downsampling. These features are then merged with the newly acquired features, obtained via upsampling in the decoder section of the network on the right-hand side. Consequently, we can gather detailed information that would be lost in the compressed/encoder path that might otherwise be lost in the compressed pathways, resulting in an improvement in the overall predictive capability of the model.

The left network is partitioned into distinct stages with varying resolutions through operations that entail 1-3 convolutional layers, an SE module, and a downsampling layer. Each stage of the convolutional process learns a residual function, this construct achieves faster convergence than a model that does not learn residuals. Specifically, the residual learning approach is achieved by incorporating the input of each stage into the output of the SE module at that stage (Milletari et al., 2016). Instead of using a pooling layer as a downsampling layer, which can cause a loss of feature map details and result in reduced recognition accuracy (Springenberg et al., 2014), we adopted a convolutional layer with a  $3 \times 3 \times 3$  convolutional kernel size and a 2-step size as

	Model layers	Type	Filter size/number/layers	SE module	Active function	Input (D × W × H × C)
Encoding area	Layers 0	Conv	3 × 3 × 3/8/1	No	PReLU	64 × 128 × 128 × 4
	Layers 1	Conv	3 × 3 × 3/8/1	Yes	PReLU	64 × 128 × 128 × 8
	Layers 2	Conv	3 × 3 × 3/16/2	Yes	PReLU	32 × 64 × 64 × 16
	Layers 3	Conv	3 × 3 × 3/32/2	Yes	PReLU	16 × 32 × 32 × 32
	Layers 4	Conv	3 × 3 × 3/64/2	Yes	PReLU	8 × 16 × 16 × 64
	Layers 5	Conv	3 × 3 × 3/128/2	Yes	PReLU	4 × 8 × 8 × 128
Decoding area	Layers 6	Conv	3 × 3 × 3/64/2	Yes	PReLU	8 × 16 × 16 × 64
	Layers 7	Conv	3 × 3 × 3/32/2	Yes	PReLU	16 × 32 × 32 × 32
	Layers 8	Conv	3 × 3 × 3/16/2	Yes	PReLU	32 × 64 × 64 × 16
	Layers 9	Conv	3 × 3 × 3/8/2	Yes	PReLU	64 × 128 × 128 × 8

Figure 2: The detailed explanation of the architecture of SE-VNet

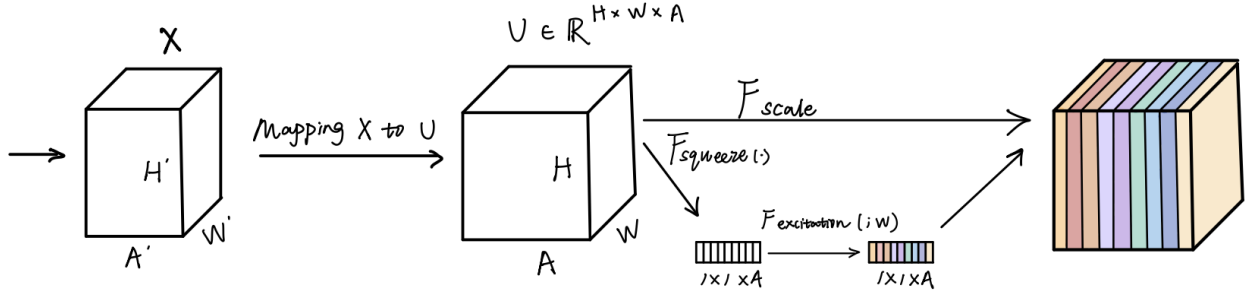


Figure 3: The structure of the Squeeze and Excitation module

suggested by Fausto et al. 2016, which provides a better solution for downsampling.

The right half of the network is responsible for feature extraction and extending the spatial support of the low-resolution feature map. This part is done to collect and combine the essential information, thereby enabling the output of the volume segmentation of the 4 channels (the four regions of the brain tumor). The decoder comprises 1 to 3 convolutional layers, a SE module, and an upsampling layer, and also learns a residual function at each stage. In addition, the upsampling layer is combined with a jump connection to receive a high-level feature map from the encoder, thus improving the performance and effectiveness of the decoder.

In the output layer of the V-Net structure, a 3D convolution kernel ( $1 \times 1 \times 1$ ) is first used to align the number of output channels with the number of categories. Finally, softmax was utilized to transform channel values into corresponding category probabilities.

### 3.1. SENet module

Squeeze-and-Excitation Networks (SENet) is a new image recognition architecture published by Momenta in 2017 (Hu et al., 2018), which attempts to explicitly model the interdependencies between feature channels. Instead of introducing a new spatial dimension for feature channel fusion, the network employs a novel strategy called “feature rescaling”. Specifically, the strategy can automatically learn the importance of each feature channel, promoting useful features and suppressing less useful features accordingly.

The structure of the Squeeze and Excitation module is shown in Figure 3.

The steps involved in the Squeeze-and-Excitation block are as follows:

#### 3.1.1. SQUEEZE

In the Squeeze section, global average pooling is used to compress the features of each channel of the input feature map into a single scalar value, which represents the global

information on that channel. This process can decrease the number of computations and parameters required and supplies global information of the input feature map for subsequent Excitation.

### 3.1.2. EXCITATION

Excitation is a process that evaluates the significance of features in each channel, boosting the important ones and dampening the irrelevant ones. Basically, the excitation part consists of two fully connected layers, where the ReLU function activates the first layer, and the sigmoid function activates the second layer. This enables the determination of the importance of features in each channel.

Finally, each channel of the original feature maps is multiplied by its corresponding excitation weight we got from Excitation, and then get the output that has the same size as the original feature map.

## 4. Training methods

To train the SE-VNet, we used Adam as the optimizer, which can handle the high-dimensional and non-convex optimization problem that arises when training deep neural networks. We set the learning rate to 0.001 ensuring that the weights of the neural network are updated by a small amount at each iteration of the optimizer. Given the presence of some noise in our dataset, a small learning rate may be more suitable as it enables the optimizer to make small adjustments to the weights without over-pursuing the optimal solution. For batch size, we set it to 128. This is a common starting point for experimentation and can provide a good balance between training efficiency and accuracy for many neural network architectures and datasets.

To reduce overfitting in the SE-VNet, we used two techniques: dropout and batch normalization. Dropout with a rate of 0.02 was applied, which randomly drops out a proportion of neurons during each training iteration. This helps to prevent co-adaptation between neurons and encourages the network to learn more robust features. Batch normalization was also applied, which normalizes the activations of each layer during training. This can help to stabilize the gradients and improve the generalization performance of the network. Together, these techniques can be effective in reducing overfitting and improving the performance of the SE-VNet on the given dataset.

This experiment was conducted on the NVIDIA Tesla K80 server, using TensorFlow as the back-end framework. The server is equipped with 2 x GK210 GPUs, each with 12GB of RAM.

## 5. Numerical results

Our objective is to confirm the actual predictive effect of the trained model. Since the target information is not masked in the official validation set, we opted to randomly choose 20 samples from the training set for testing and assessing the model's performance. These 20 samples were not used for training purposes. Following this, we proceeded to evaluate the model's performance on the obtained validation set.

There are four evaluation indexes used to evaluate the accuracy of the model segmentation results: precision, sensitivity, specificity, and Dice Score. Higher values of precision, sensitivity, specificity, and Dice Score indicate better model segmentation performance. The description and calculation formula of each evaluation index is as follows:

### Precision

It measures the ratio of accurate positive predictions made by the model to the total samples predicted as positive. It is calculated using the following formula:

$$Precision = \frac{TP}{TP + FP}$$

### Sensitivity

It measures how accurately the model can identify positive instances out of all the instances that are labeled as positive, and the calculation formula is:

$$Sensitivity = \frac{TP}{TP + FN}$$

### Specificity

It measures the model's capacity to accurately detect negative instances within the entire pool of identified negative instances, and the formula is:

$$Specificity = \frac{TN}{TN + FP}$$

### Dice Score

The Dice score is a metric widely used metric for assessing the efficacy of image segmentation algorithms. It ranges from 0 to 1. A higher Dice coefficient (approaches 1) indicates better segmentation performance of the algorithm, whereas a lower value (approaches 0) indicates poorer segmentation performance.

It is defined as follows:

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|}$$

A: The ground truth segmentation

B: The results of model segmentation

$X \cap Y$  : The overlap between A and B

operator  $|\cdot|$ : This function calculates the quantity of pixels (or voxels in 3D) that make up a region, which is directly proportional to the physical volume of the region being analyzed.

The precision, sensitivity, specificity and mean Dice coefficients for the ET, WT, and TC are presented in the Table 1 below:

Precision	Sensitivity	Specificity
0.996	0.992	0.999

(a) Precision & Sensitivity & Specificity

ET	WT	TC
0.26	0.72	0.61

(b) Mean Dice Coefficients

Table 1: The performance of the model

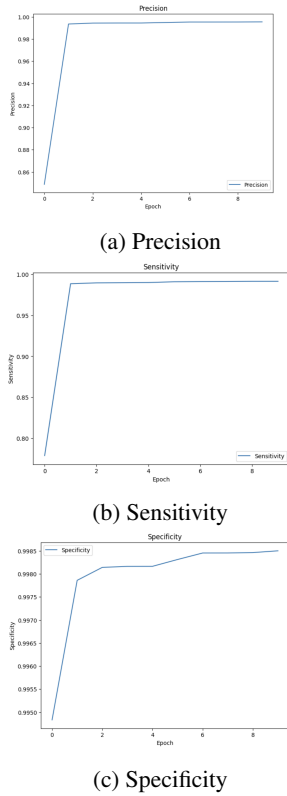


Figure 4: The performance of the model

Table 1a shows that the accuracy, sensitivity, and specificity of our trained brain tumour segmentation model were all stable at above 0.99, indicating the model performed well in correctly identifying tumour and non-tumour areas.

It can be observed from the table 1b that our model is most sensitive to WT. However, ET and TC exhibit relatively weaker sensitivities, particularly ET. Upon analyzing the dataset, we have found that many samples with zero Dice coefficients do not possess ET or TC. Undoubtedly, this will reduce the overall evaluation of the model with respect to the Dice coefficients.

## 6. Conclusion

In conclusion, this study aimed to evaluate the performance of SEV-Net for MRI brain tumor segmentation. It demonstrated that SEV-Net achieved advanced performance on our dataset, outperforming other popular segmentation models such as U-Net and V-Net. The utilization of spatial and channel-wise attention mechanisms in the SEV-Net architecture enables the model to selectively emphasize the most significant regions of the input image, resulting in more precise and accurate tumor segmentation results. However, our study also identified some limitations of SEV-Net, including its relatively high computational cost and sensitivity to variations in input image quality. Possible future work to address these limitations could include optimizing the SEV-Net architecture for faster inference times or developing methods to improve the robustness of the model to image quality variations. Additionally, it may be interesting to explore the potential of combining SEV-Net with other advanced techniques such as transfer learning or data augmentation to further improve its performance.

Furthermore, it is worth mentioning the strengths and weaknesses of SEV-Net and V-Net, which are both popular models for brain tumor segmentation. SEV-Net has the advantage of incorporating attention mechanisms to capture more informative features and improve segmentation accuracy. However, it is computationally expensive and may not be suitable for real-time applications. V-Net, on the other hand, is a faster model that can be trained with limited data, making it more practical for applications with limited computational resources (Li et al., 2022). However, it may not perform as well as SEV-Net on more challenging datasets with complex tumor structures.

In conclusion, SEV-Net demonstrated exceptional potential as an MRI brain tumor segmentation model, achieving outstanding results on the BraTS2021 dataset. Our study provides insights into its strengths and limitations, as well as potential future directions for improvement. By addressing these limitations and exploring new research avenues, we can further advance the state of the art in brain tumor segmentation and ultimately contribute to more accurate diagnosis and treatment of brain tumors.

## 7. Statement about individual contributions

The report is equally contributed by three authors.

## References

- Allah, A. M. G., Sarhan, A. M., and Elshennawy, N. M. Edge u-net: Brain tumor segmentation using mri based on deep u-net model with boundary information. *Expert Systems with Applications*, 213:118833, 2023.
- Banerjee, S. and Mitra, S. Novel volumetric sub-region segmentation in brain tumors. *Frontiers in Computational Neuroscience*, 14:3, 2020.
- Hu, J., Shen, L., and Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.
- Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A. V., Criminisi, A., Rueckert, D., and Glocker, B. Deepmedic for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and mTOP 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 2*, pp. 138–149. Springer, 2016.
- Li, W., Lin, N., Zhang, M., Lu, H., Chen, X., and Li, X. Vnet: a versatile network to train real-time semantic segmentation models on a single gpu. *Science China Information Sciences*, 65(3):139105, 2022.
- Mehta, R. and Arbel, T. 3d u-net for brain tumour segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, pp. 254–266. Springer, 2019.
- Milletari, F., Navab, N., and Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571. Ieee, 2016.
- Raza, A., Ayub, H., Khan, J. A., Ahmad, I., S. Salama, A., Daradkeh, Y. I., Javeed, D., Ur Rehman, A., and Hamam, H. A hybrid deep learning-based approach for brain tumor classification. *Electronics*, 11(7):1146, 2022.
- Sadeghi, S., Farzin, M., and Gholami, S. Fully automated clinical target volume segmentation for glioblastoma radiotherapy using a deep convolutional neural network. *Polish Journal of Radiology*, 88(1):31–40, 2023.
- Schettler, D. Brats 2021 task 1 dataset, 2021. URL <https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1>.
- Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.
- Tumors-Classifications, B. Symptoms, diagnosis and treatments, 2021.
- Wang, G., Li, W., Ourselin, S., and Vercauteren, T. Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation. *Frontiers in computational neuroscience*, 13:56, 2019.