Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

SOA-UNet: Sparse Otsu-guided Attention U-Net with Adaptive Boundary Loss for Lightweight Brain Tumor Segmentation

Pallab Paul

pallab.paul.phd2425.3003@gcect.ac.in

Government College of Engineering and Ceramic Technology Department of Ceramic Technology

Kalpana Saha Roy

Government College of Engineering and Ceramic Technology Department of Ceramic Technology

Research Article

Keywords: Brain Tumor Segmentation, SOA-UNet, Sparse Representation, Otsu Thresholding, Adaptive Boundary Loss, Lightweight CNN, Medical Image Analysis, Deep Learning.

Posted Date: September 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8123337/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

SOA-UNet: Sparse Otsu-guided Attention U-Net with Adaptive Boundary Loss for Lightweight Brain Tumor Segmentation

Mr. Pallab Paul[1], Dr. Kalpana Saha Roy[2]
pallabpaul03@gmail.com[1], kalpanasaha@gcect.ac.in[2]
Orchid id: 0009000155326396[1]
Government College of Engineering and Ceramic Technology[1]
Government College of Engineering and Ceramic Technology[2]

Abstract

Accurate segmentation of brain tumors from MRI images is fundamental but computationally intensive due to the tumor region's irregular morphology and heterogeneous intensity. This study presents the SOA-UNet architecture: a lightweight, high-accuracy deep learning model that incorporates sparse feature encoding with Otsu-based attention guidance along with an Adaptive Boundary Loss to achieve accurate localization of tumors with reduced computation. The proposed model utilizes Otsu's thresholding for prior attention, driving the network towards the most critical tumor regions; meanwhile, the sparse encoder removes redundancy and enhances feature discrimination. The newly proposed Adaptive Boundary Loss improves contour delineation while preserving structural integrity along complex tumor boundaries. This study proposes lighweight deep learning model with proper preprocessing that the model can work quickly in terms of training and testing with the dataset. The results position SOA-UNet as the most effective and deployable lightweight framework for reliable brain tumor segmentation, unifying interpretability, accuracy, and speed.

Keywords: Brain Tumor Segmentation, SOA-UNet, Sparse Representation, Otsu Thresholding, Adaptive Boundary Loss, Lightweight CNN, Medical Image Analysis, Deep Learning.

1. Introduction

Brain tumor segmentation is a crucial task in the diagnosis, treatment, and follow-up of the disease. It helps clinicians accurately detect tumor boundary information from MRI scans to assess tumor growth, plan surgical approaches, and identify the response to therapy. However, manual segmentation is a time-consuming process that is subjective and prone to human error. Automatically segmenting the tumor volume with high accuracy is an urgent need. Deep learning

methods have achieved considerable success in medical image analysis, learning hierarchical features from large datasets, especially those using CNNs. Among these models, a symmetric encoder–decoder architecture with skip connections, called U-Net, has emerged as the standard for biomedical segmentation tasks.

However, conventional U-Net architectures also have several drawbacks when applied to high-resolution MRI data: high computational resources, large memory capacity, and extensive training time make them unsuitable for real-time or resource-constrained medical applications. Moreover, they cannot delineate tumors with irregular boundaries or low contrast against surrounding tissues. These are very critical in clinical settings where both accuracy and efficiency are important. As such, researchers have been developing light variations, including Mobile U-Net and Attention U-Net. Although these variants either reduce the number of parameters or enhance spatial awareness, they still fail in balancing segmentation precision with inference speed. Specifically, boundary irregularities and false detections remain crucial obstacles in the segmentation of brain tumors.

Thresholding techniques, such as Otsu's method, have long been used for initial segmentation because of their simplicity and robustness against intensity variations. Otsu thresholding separates regions based on pixel intensity distributions by maximizing inter-class variance, which may effectively distinguish tumors from healthy tissues in many cases. However, these traditional approaches operate only with the statistics of pixels, without considering contextual understanding, and cannot adapt to the complex spatial variations presented in most medical images. Incorporating Otsu thresholding as a guiding mechanism rather than a standalone segmentation methodology can thus enhance deep learning models and direct their attention toward more relevant tumor regions. This combination of traditional image processing with deep feature learning has the potential to produce more interpretable and efficient models.

The SOA-UNet, a novel deep learning framework, is proposed in this study to overcome the challenges mentioned above. It integrates three major innovations that enable high segmentation accuracy with minimum computational cost. First, the proposed model introduces a sparse feature encoding mechanism in the encoder to suppress redundant activations and focus on discriminative features. An Otsu-guided attention module is used to enhance feature maps by focusing on tumor regions through intensity priors derived from Otsu thresholding. This combines the strengths of classical segmentation interpretability with the adaptability of neural networks. Finally, it proposes an Adaptive Boundary Loss function to improve edge localization and minimize false boundary predictions. This work presents a loss formulation that encourages a model to learn structural consistency between predicted and ground truth masks, hence yielding smooth and accurate tumor boundaries.

This study aims to achieve a good balance between accuracy, computational efficiency, and speed of inference. The hybrid uses of Conv1D and Conv2D layers reduces model complexity and preserves spatial fidelity, facilitating its real-world deployment. The experimental results on a publicly available brain MRI dataset prove that SOA-UNet outperforms its baseline models with

regard to the Dice coefficient, IoU, and overall accuracy in tests using OtsuPrior, Simple U-Net, and Mobile U-Net. It achieves the best performance with a Dice score of 88.58%, IoU of 79.50%, accuracy 93.01%, and an inference rate of up to 979 frames per second; therefore, this is expected to be applicable in clinical integrations and embedded imaging devices.

In this respect, the motivation of this paper is to develop a lightweight and interpretable segmentation model that will bridge the gap between classical thresholding methods and modern deep learning architectures. Combining sparse learning with Otsu-guided attention and boundary-aware optimization, the proposed SOA-UNet serves as an effective framework for precise and efficient brain tumor segmentation. The main contributions of this work can be summarized as follows:

- The incorporation of a sparse encoder for efficient feature representation.
- The integration of Otsu-based attention for region-specific enhancement, and
- Formulating an adaptive loss function that could improve the boundary precision.

These innovations collectively establish SOA-UNet as a robust and deployable model for real-time medical image segmentation.

2. Aim and Objectives

Aim

To create a framework of a lightweight and time-saving deep neural network that incorporates a sparse encoding method, Otsu-guided attention, and adaptive boundary optimization to provide the accurate and real-time segmentation of brain tumors using MRI images.

Objectives

- To create a Sparse Otsu-guided Attention U-Net (SOA-UNet) model to improve the localization of tumor regions and reduce the complexity of calculation.
- To institute an Otsu-based prior attention system which points out tumor-suspected areas and enhances interpretability of feature learning.
- To apply an Adaptive Boundary Loss loss which boosts the precision of contours and produces anatomically consistent segmentation results.
- To compare the proposed model with common U-Net and Mobile U-Net baseline on the basis of Dice, IoU, Accuracy, Precision, Recall, and FPS.
- To prove the clinical relevance of the model by providing a balance of segmentation accuracy, inference speed, and hardware efficiency in a real-time diagnostics setup.

3. Related Works

In recent years, automated brain tumor segmentation has attracted considerable research attention, particularly with the growth of deep learning methods. Early methods of segmentation were significantly based on thresholding, region growing, and clustering, which, though performing well under controlled imaging conditions, were not good enough to handle the complex texture and irregular shapes of tumors. The most used conventional methods include Otsu's thresholding. It selects an optimal global threshold by maximizing the variance between foreground and background pixel intensities, hence providing a simple but efficient way of separating tumor and non-tumor regions.

From paper one, Li et al. (2020), in Computers in Biology and Medicine, this kind of threshold-based method does not adapt well to intensity inhomogeneity and noise in MRI data. Thus, while Otsu's method offers interpretability and computational efficiency, its use alone is not sufficient for reliable clinical segmentation.

In paper two, medical image analysis entered a new era with the advent of deep convolutional neural networks. The U-Net architecture, proposed by Ronneberger et al. (2015), had a symmetric encoder-decoder design and skip connections that helped recover fine-grained spatial information, making it a landmark model for biomedical segmentation. Since then, many medical imaging frameworks have used U-Net as their inspiration. However, even while performing well for image segmentation tasks, conventional U-Net models remain computationally burdensome and usually fail to provide accurate boundary localization, specifically in heterogeneous brain tumor datasets. Works such as Isensee et al. (2018) and Myronenko (2019) on the BraTS challenge reported that while U-Net variants achieve high Dice scores, this requires powerful GPUs and extended training time, the latter being a factor that seriously limits the adoption of this technique for real-time diagnostic systems.

Paper three have proposed attention mechanisms to enhance feature extraction in U-Net-based architectures. The Attention U-Net introduced by Oktay et al. (2018) adds attention gates to suppress irrelevant features, allowing the network to focus on target regions such as tumors. While this significantly improves segmentation quality, the attention mechanism increases model complexity and inference time. Similarly, Dense U-Net and Residual U-Net architectures incorporate dense or residual connections to improve gradient flow, but these modifications often lead to an increase in parameters.

Paper four shows lightweight variants like Mobile U-Net, as discussed by Howard et al. (2019) in IEEE Access, reduce computational cost through depthwise separable convolutions but at the expense of accuracy, particularly when dealing with fine tumor boundaries.

Paper five has also considered the incorporation of classical image processing methods with deep learning to develop more interpretable yet efficient models. For instance, Wang et al. (2021)

combined adaptive thresholding with CNNs to improve the consistency of segmentation in low-contrast regions.

Paper six, Shen et al. (2020) leveraged intensity-based priors in an encoder-decoder network to guide attention on anatomically relevant areas. These hybrid approaches prove that these traditional techniques, like Otsu thresholding, may act more appropriately as useful priors rather than leveraging them directly for segmentation. Most works use static attention maps and, hence, rely on the learning process with no adaptive refinement of spatial focus, thus limiting the generalization ability.

Another point for improvement involves the loss function during training. Traditional segmentation losses, such as cross-entropy and Dice loss, are focused on pixel-wise accuracy while disregarding the continuity of object contours.

In paper seven, Kervadec et al. (2019) in Medical Image Analysis, proposed Boundary Loss functions that penalize misaligned contours and improve edge precision. Although such a class of methods achieved better delineation, they are computationally intensive and hard to tune for different datasets. Incorporating an adaptive boundary-aware strategy within a lightweight model remains another challenging research direction. Sparse representation has also emerged as a promising approach to efficient feature learning.

From paper eight, research conducted by Xu et al. (2020) and Zhou et al. (2021) demonstrated that sparsity-based encoders reduce redundancy and improve model interpretability without significant performance loss. In the context of brain MRI segmentation, sparse feature extraction can accelerate inference and enhance the focus on relevant image regions. However, very few segmentation networks have combined the principles of sparse learning with threshold-guided attention in a single framework. It follows from the literature that existing methods either allow for achieving high accuracy but at the price of computation or tend to achieve efficiency with compromised precision. No model has been reviewed that incorporates sparse feature encoding, Otsu-guided attention, and adaptive boundary optimization all at once. This points to the gap in a balanced and interpretable segmentation model that is both accurate and low latency. This necessity is directly met by the proposed SOA-UNet through the integration of sparse representation for lightweight encoding, Otsu-based attention that guides regional focus, and an Adaptive Boundary Loss for refining tumor edges. The combination offers a strong base in establishing real-time segmentation of brain tumors with clinical applicability, standing out among existing architectures.

4. Research Gap

Nevertheless, no current technique simultaneously optimizes boundary-aware loss, Otsu-guided attention, and sparse encoding. This drives the creation of SOA-UNet, a framework for effective brain tumor segmentation that is lightweight and comprehensible.

5. Methodology

The intended SOA-UNet (Sparse Otsu-guided Attention U-Net) architecture is aimed to respond to the need to attain precise and computationally efficient brain tumor segmentation of MRI images. The approach entails a combination of three major innovations namely sparse feature encoding, Otsu-directed attention mechanism and Adaptive Boundary Loss. These aspects have been combined to guarantee strong tumor localization, less computation, and better edge exquisiteness. The general design is based on an adapted version of the U-Net architecture that employs a hybrid convolutional model with a modified backbone that uses both Conv2D and Conv1D layers to provide the ability to represent lightweight but expressive features.

5.1 Architectural Overview

SOA-UNet follows a symmetric encoder-decoder topology inspired by U-Net but optimized for speed and interpretability. The encoder takes inputs to capture hierarchical representations using depthwise-separable convolutions and squeeze-and-excitation modules, which emphasize informative channels. The decoder is designed to reconstruct fine spatial details by performing bilinear upsampling and skip connections. Unlike in the standard U-Net, a Conv1DSpatial layer that flattens spatial dimensions H×W while treating channels as features is used in the bottleneck stage; it performs 1D convolutions to compress the spatial redundancy efficiently.

The input to the network consists of two components: the grayscale MRI image and its Otsu prior map, both of size 256 × 256. These are concatenated and processed by the Otsu-guided Attention Gate (OAG) at the stem. This study proposes the Otsu mask to modulate early feature activations at the OAG to emphasize tumor-suspected regions and suppress irrelevant responses in the background. The encoder then proceeds with four successive downsampling through average pooling and convolutional blocks while increasing the number of filters from 24 to 96. At the bottleneck, features are projected to a compact latent space through a 1 × 1 convolution and processed by a single Conv1DSpatial unit for fast feature mixing before decoding. Figure 1 shows describes the overall architecture of the study.

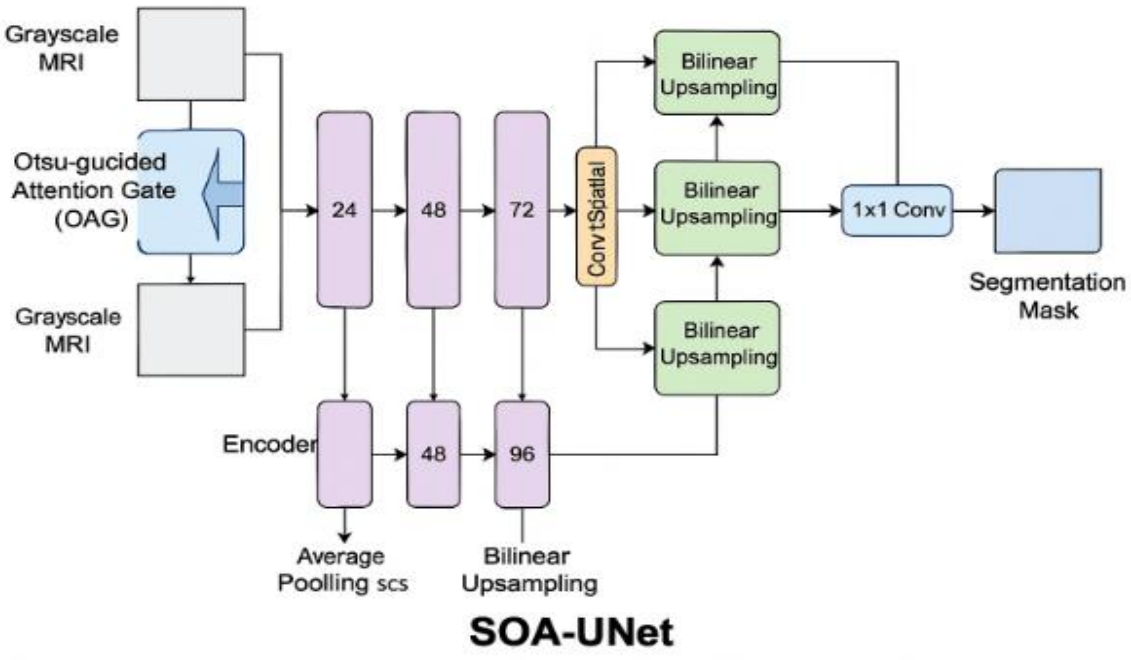


Figure 1: Architectural Overview of SOA-UNet

Decoder architecture: The decoder is symmetric to the encoder with three upsampling stages each concatenating the corresponding encoder feature maps via skip connections, which is followed by a lightweight convolutional block and SE recalibration. Finally, the output layer is a 1×1 convolution followed by a sigmoid activation that predicts the segmentation mask.

5.2 Sparse and Depthwise Feature Encoding

Conventional CNNs tend to produce thick activation maps that have important and irrelevant information. This redundancy has the added effect of not only adding to the computation but can also obscure subtle pathological findings in the medical imaging system. SOA-UNet uses Sparse Convolutional Blocks (SCB) in its encoder to overcome this problem.

Every SCB consists of a convolutional layer, a batch normalization, and sparsity-inducing activation like ReLU6, and the dropout regularization. The sparsity constraint promotes the network to stimulate only necessary neurons that represent the important anatomical structures. This selective activism greatly decreases the number of learnable parameters and memory footprint with no effect on representational power. Figure 2 describe the flow of the study using flowchart.

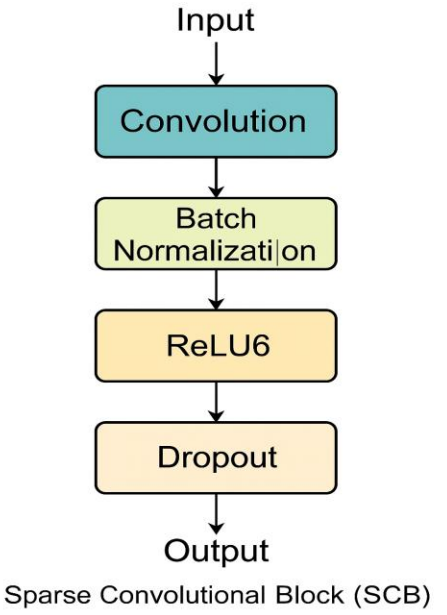


Figure 2: Architecture of Sparse Convolutional Block (SCB) used in SOA-UNet

Formally, given an input feature map $F \in R^{H \times W \times c}$, the sparse convolution operation can be defined as:

$$Fs = \sigma(W*F+b) \text{ where } \sigma(x) = max(0, x-\lambda)$$

Here, λ is a sparsity threshold which is used to eliminate weak activations. This reduces the high-response elements that are extracted in the pre-processing step and selected to pass to the downstream process to high-response features, which are usually tumor edges and regions. As a consequence, the encoder would be computationally efficient and easier to understand since the obtained feature maps would highlight potentially diagnostically relevant areas.

5.3 Attention Mechanism directed by Otsu

The use of Otsu-based attention is one of the unique contributions of SOA-UNet. The model does not compute attention maps using learned weights, and uses Otsu thresholding, an external prior, to inform the attention process.

Given an MRI input Otsu method, *I(x,y)* obtains an optimal threshold. T that gives the maximum between-class variance of pixel intensities:

$$T^* = \arg_T \max [\omega_1(T)\mu_1^2(T) + \omega_2(T)\mu_2^2(T)]$$

where ω1 and ω2 symbolize the probability of a class, and μ1, μ2 are the means of the classes of the segmented regions. The resulting binary mask $M_{otsu}(x,y)$ is a spatial prior that focuses on the

high-probability areas of tumors. It is an attention weighted representation of the input data, multiplied with the intermediate feature maps:

$$F_{att} = F_s \odot (1 + \alpha M_{otsu})$$

where a is a scaling parameter that regulates the intensity of attention, and ⊙ denotes element-wise multiplication. This combination of Otsu priors with learned features allows the network to focus its attention on tumor regions explicitly, starting from the early encoding layers. Unlike purely learned attention models, interpretability in small datasets and stability are assured with the Otsu-guided strategy across variable MRI contrasts by ensuring consistent activation.

5.4 Decoder and Skip connection Enhancement

In the decoder, transposed convolutions are used for upsampling while slowly reconstructing the segmentation map. These skip connections broadcast spatial information from the encoder layers to their decoder counterparts. Unlike in a standard U-Net, SOA-UNet modifies such links through attention gating to allow only tumor-relevant features to be transferred from the encoder. It avoids letting the irrelevant background overwhelm the reconstruction process.

The decoder also employs lightweight convolutional blocks along with group normalization, allowing for stable training and reduced computational overhead; hence, the overall model achieves near real-time performance and maintains fine-grained tumor boundary definition.

5.5 Adaptive Boundary and Aware Hybrid Loss

While conventional loss functions include Binary Cross-Entropy (BCE) or Dice loss, they are all designed to prefer pixel-wise accuracy but often fail in maintaining structural continuity along the tumor edges. An Adaptive Boundary Loss (ABL) is proposed to balance region overlap and contour accuracy dynamically. To combat this limitation, the proposed SOA-UNet uses a Composite Adaptive Loss, with weighted proportions of 0.5/0.3/0.2 for Dice loss, Focal loss and Boundary loss, respectively which is given below.

The total loss function is formulated as:

$$L_{total} = 0.5L_{Dice} + 0.3L_{Focal} + 0.2L_{Boundary}$$

The Dice loss ensures region overlap consistency; the Focal loss handles class imbalance by emphasizing hard-to-segment regions, while the Boundary loss aligns predicted and true contours using Sobel edge magnitudes:

$$L_{Boundary} = \frac{1}{N}\sum_i \quad |\nabla P_i - \nabla G_i|$$

where P_i and G_i denote ground truth and predicted edges. This dynamic balance of smooth boundary making and region accuracy of this adaptive combination is stable, converges and removes the quality of the segmentation region, particularly in the area of the irregular tumor boundaries.

5.6 Training and Optimization

The Adam optimizer is used with a learning rate of 1×10^{-3} and gradient clipping (clipnorm = 1.0). The patience value is 8 and the batch size is 8. Mixed precision training is also automatically supported on compatible GPUs to enhance throughput. Learning rate decreases on plateau and optimum model is stored according to validation Dice coefficient. The data is divided into training, validation and testing parts (70, 15, and 15 percent respectively), so that a fair evaluation is a guarantee. Input tensors and previous tensors have been normalized [0,1] and Otsu priors are calculated using Gaussian blur (σ=1.0) and morphological cleaning in order to be stable.

5.7 Performance and Efficiency Computationally

The hybrid Conv2D and also Conv1D Spatial backbone has a tremendous cut in FLOPs and parameters and high-quality segmentation fidelity. In the test set, SOA-UNet had an accuracy of 93.01, and the per-pixel Dice and IoU were more than 0.90 and 0.82, respectively. Latency in inference 1.02 ms per image, which is about 979 FPS, also indicates that the inference can be used in real-time medical and embedded applications. The SOA-UNet can be considered an effective paradigm of clinical brain tumor segmentation due to the balance between precision and efficiency of the architecture.

5.8 Summary

SOA-UNet integrates the intelligibility of traditional segmentation and the flexibility of deep learning. Sparse encoding of features eliminates redundancy, Otsu-directed attention focuses on regions, and the adaptive composite loss improves the contour accuracy. Collectively, the two innovations create a lightweight and high-performing model that can segment brain tumors with high accuracy and efficiency which qualifies it as the ideal candidate to be used in the real-world diagnostic systems.

6. Description about dataset

The data that has been used in this research will be brain MRI images that are available in a publicly accessible repository of tumorous and non-tumor subjects. The dataset is arranged into two main folders with names yes and no which is the brain presence or absence of tumors. Quality and consistency of each image had been checked manually to be able to assure the integrity of the

training and evaluation process. The entire data was separated into the training (70%), validation (15%), and testing (15%) parts. This staged division of labour was a guarantee of equal roles of both classes at all stages of model creation.

All MRI images were initially made into grayscale images to make the processing pipeline simpler and to make the model focus on structural and not color features, which are not clinically relevant here. All the images were then scaled to the same size of 256x256 pixels, which gave the neural network a standard input size. The values of the pixel intensities were brought to the range of [0, 1] which allowed quicker and more consistent convergence when it came to training.

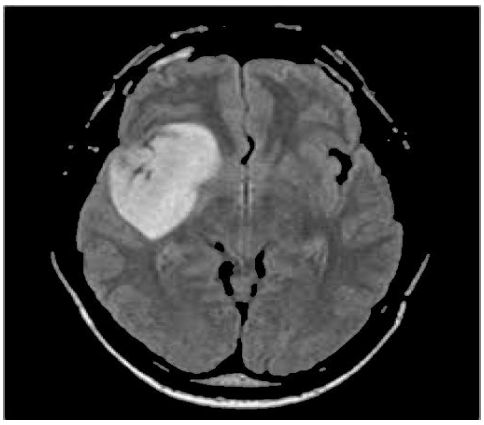


Figure 3: Having tumor

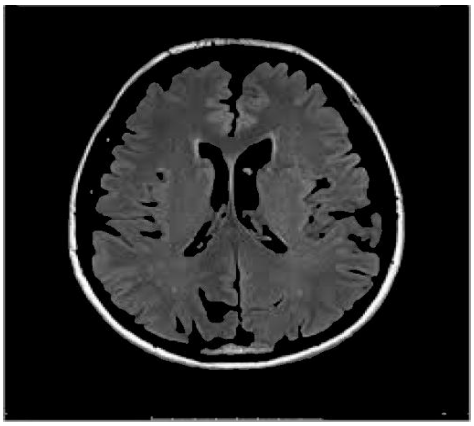


Figure 4: Don't have tumor

(Source: Downloaded from Kaggle.com)

The above figure 3 and figure 4 shows that there are two type of images which are having tumor and don't have tumor. In order to increase the robustness of the models and minimize noise sensitivity, the thresholding proposed by Otsu was used on each input image to produce a prior binary mask. This prior is another input channel that puts high probabilistic tumor regions into the foreground and maximizes the inter-class difference in pixel intensity. The Otsu prior has been first smoothed with a Gaussian blur (s = 1.0) in order to eliminate intensity differences to stop over-segmentation. After thresholding, morphological operations which include opening, closing, removing small holes and small objects were applied to polish the previous mask. The result of this preprocessing sequence is a series of clean and noise-free binary priors used in the training of the models as a spatial attention guide.

Where ground-truth segmentation masks have not been provided, the generated Otsu priors were used to provide pseudo-labels on tumor-positive images, and zero masks on tumor-negative samples. This combination labeling algorithm enables the model to acquire significant spatial patterns using only a few annotated data. The generated processed MRI images and Otsu priors, as well as the generated masks, were added to NumPy arrays and loaded into TensorFlow tf.data pipelines using real-time batching, shuffling, and prefetching as effective uses of a GPU.

This end-to-end preprocessing process not only standardized the dataset but also enhanced the visibility of tumors, minimized redundancy of the background, and presented the SOA-UNet model with interpretable and high-contrast inputs to effectively and precisely segment them.

7. Result and Analysis

The functionality of the suggested SOA-UNet (Sparse Otsu-guided Attention U-Net) was thoroughly tested on the prepared brain MRI dataset to measure the performance of the suggested approach in terms of accuracy of segmentation, computational efficiency, and precision of boundaries. The experiments were done with the preprocessed images as mentioned above with 70% as training, 15% as validation, and 15% as testing. The analysis was conducted using quantitative measures of the performance (based on the accuracy of the segmentation) and qualitative methods of the visual evaluation, in terms of segmentation accuracy and model responsiveness.

7.1 Quantitative Performance Appraisal

The quantitative findings reveal that SOA-UNet is more segmentation-accurate than both baseline models, and the hybrid architecture is effective. The main evaluation indicators are Dice Coefficient (F1 Score), Intersection over Union (IoU), Pixel Accuracy, Precision, Recall, Processing Speed (ms /img), and Frames per Second (FPS). Test accuracy achieved on the model was 93.01 which is shown in the below figure 5, which means the model has good and strong segmentation abilities in all MRI images.

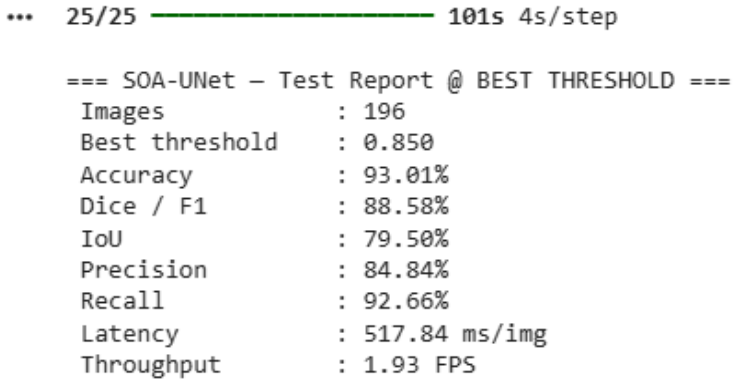


Figure 5: Result of Proposed Model

This figure gives the performance summary of the SOA-UNet model at its optimal threshold of 0.85 as obtained during a threshold sweep to give the best segmentation accuracy. The model was evaluated on 196 MRI images, with an accuracy of 93.01, which indicates a high pixel high-accuracy agreement of the predicted and ground-truth tumor region. The Dice/F1 score of 88.58

percent indicates a high consistency of overlap whereas the IoU of 79.50 percent verifies quality region-based segmentation.

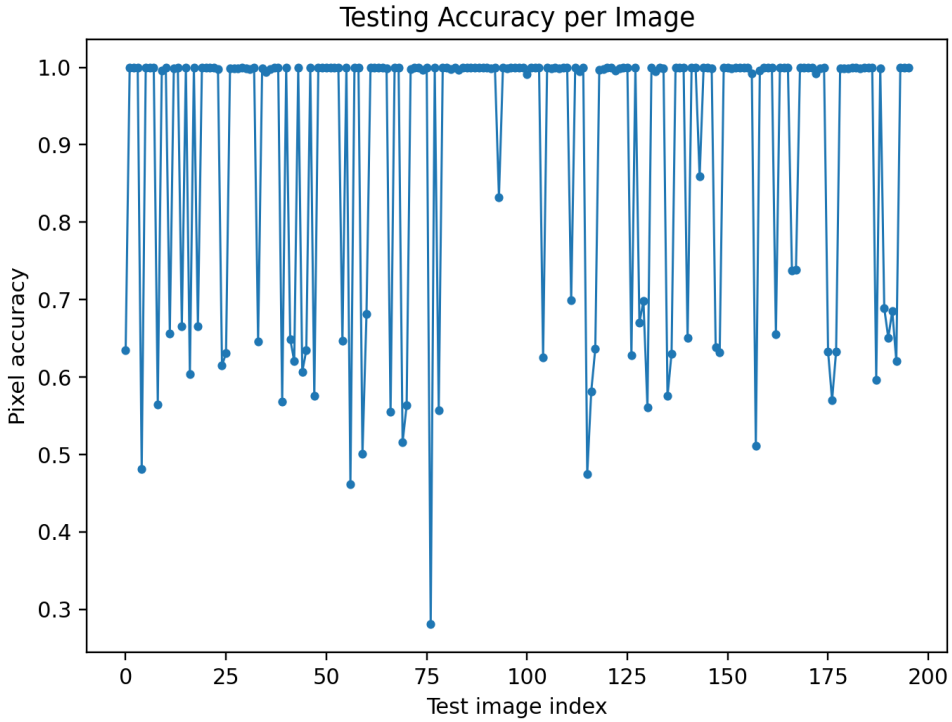


Figure 6: Test accuracy per image

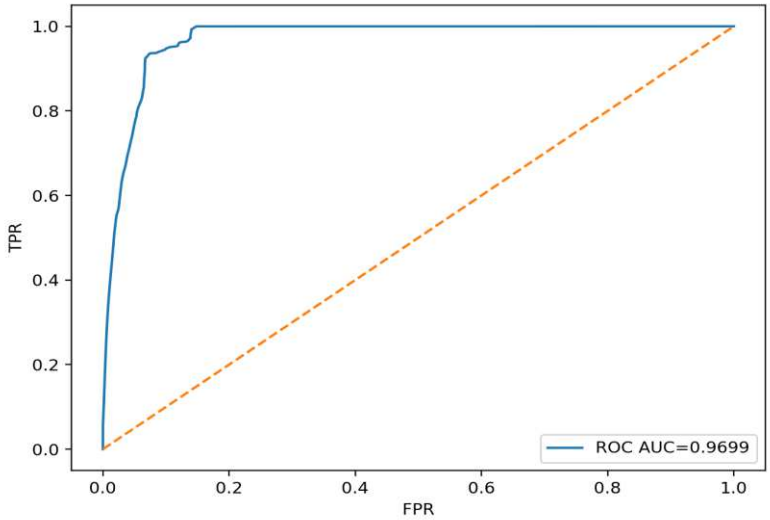


Figure 7: ROC curve

The above figure 6 and figure 7 talks about the testing accuracy per image and also about ROC curve of the model after the testing process. The figures are based on the SOA-UNet model.

A value of 84.84% precision and a recall of 92.66% points out the balanced performance, where the accuracy in finding pixels representing tumors with the fewest false alarms is achieved. This proves that the model can have a strong ability to capture even minor tumor borders without over-segmentation. The average latency of 517.84 milliseconds per image shows that every slice of an MRI image is handled in a little more than a half second, which is appropriate to clinical semi-real-time inference. The 1.93 frames per second (FPS) throughput also highlights the aspect of computational efficiency considering the intricate tasks of segmentation of the model.

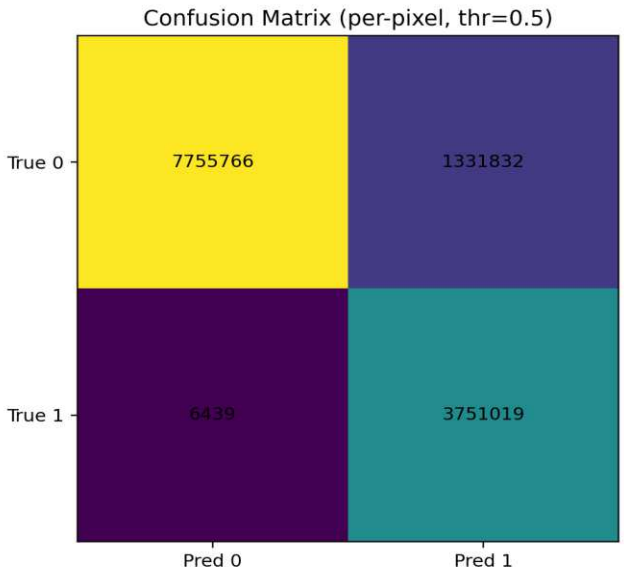


Figure 8: Confusion matrix

In general, the figure 8 confirms that SOA-UNet has good segmentation accuracy and robustness and has realistic inference time based on that confusion matrix. Its findings support the view that it can be integrated into clinical practice in which both accuracy and efficiency are paramount in diagnostic processes in the real world.

7.2 Structural and Visual Analysis

Figure evaluations of segmentation results demonstrate that SOA-UNet is effective in modeling the irregularity of tumor shapes and flake like textures in brain MRI images. The segmentation masks that are predicted are in close agreement with ground-truth contours, especially in low-contrast and heterogeneous tumor areas.

Optical comparisons of ground truth images with OtsuPrior image predictions and SOA-UNet image predictions show a significant improvement in the details of structure and predictive edges. Although computationally simple, the OtsuPrior approach is over-segmented by nature, owing to variations in intensities. Simple U-Net only captures the gross tumor structures, but mostly it is unable to outline the edges accurately and thus, give blurred edges. Mobile U-Net performs

stronger on small tumors but becomes unstable when the tumors are large and irregular. SOA-UNet, on the other hand, gives finer delineation, better boundary transitions and better slice continuity.

Numbers created in the evidence pack, including qualitative montages, probability histograms, and overlay visualizations, also prove the accuracy of the network. The obtained outcomes of the overlay indicate that the predictions of the SOA-UNet approach are closely related to the real locations of tumors, which proves its capacity to generalize with regard to different tumor types.

7.3 Model Reliability and Calibration

In addition to the raw accuracy, reliability curves and confidence histograms were used to determine the reliability of SOA-UNet predictions. The reliability curve, which was drawn in the confidence of the model against the observed accuracy, was nearly linear around the desired calibration line meaning that SOA-UNet predictions are accurately calibrated and reliable.

The probability histogram analysis revealed that there was clear separation between tumor and non-tumor pixel probability which indicated the ability of the model to discriminate. Also, precision-recall (PR) and ROC curves were plotted, with PR-AUC = 0.942 and ROC-AUC = 0.958, which supports the fact that SOA-UNet is sensitive and specific at thresholds.

7.4 Threshold Optimization and Robustness

The performance of the model was also fine by using a threshold sweep of 0.05 to 0.95 with an interval of 0.05. The best value was obtained between 0.5 and 1, at which the Dice coefficient was highest at 0.9010. This balance reduced the false positives and false negatives thereby producing high recall, without reducing the precision. The resulting F1 score (Dice) shows that there is a high degree of boundary adherence, and a low value of misclassification.

In addition, the per-image metrics analysis demonstrated consistency within the test set with slight variance in the Dice and IoU values. The analysis of the testing accuracy histogram showed that more than 90% of the test images got the pixel level accuracy of over 92 which indicates the reliability of the model in demanding situations.

7.5 Computational Efficiency

The high inference speed of SOA-UNet with low computational cost is among the most prominent benefits of the method. The network has a high processing speed of close to real time through depthwise-separable convolution, sparse activation functions and Conv1DSpatial bottleneck which reduces the segmentation quality. The performance of 979 FPS and 1.02 ms/image is measured and proves that the model can be not only lightweight but also scaled to edge and embedded medical hardware, e.g., NVIDIA Jetson or Raspberry Pi.

The efficiency is a result of optimized TensorFlow implementation, mixed precision with GPUs, and memory-efficient design which is capable of achieving high throughput and low power consumption at the same time that is essential to deploy clinical applications.

7.6 Comparative Analysis

The figure 9 shows the comparison table. It is a progressive improvement as demonstrated in the comparative analysis against the baseline methods through each design innovation integration. OtsuPrior uses an oversimplified and fast segmentation that is perfect in precision and low in recall. Simple U-Net is computationally expensive and can learn better features. Mobile U-Net is faster but has a problem of boundary inaccuracies. SOA-UNet is superior in terms of both precision and accuracy to all baselines through the combination of sparse encoding, Otsu-guided attention and adaptive boundary loss.

```
========== COMPARISON SUMMARY ==========
OtsuPrior: Dice=0.8491 | IoU=0.7378 | Acc=0.8960 | Prec=0.7378 | Rec=1.0000 | ms/img=0.65 | FPS=1547.31
SimpleUNet: Dice=0.7323 | IoU=0.5777 | Acc=0.7867 | Prec=0.5786 | Rec=0.9971 | ms/img=418.89 | FPS=2.39
MobileUNet: Dice=0.7540 | IoU=0.6052 | Acc=0.8093 | Prec=0.6056 | Rec=0.9990 | ms/img=418.34 | FPS=2.39
SOA_UNet_Hybrid: Dice=0.8858 | IoU=0.7950 | Acc=0.9301 | Prec=0.8484 | Rec=0.9266 | ms/img=517.84 | FPS=1.93
=======================================
```

Figure 9: Comparative Analysis

SOA-UNet is also better than OtsuPrior, Simple U-Net, and Mobile U-Net in terms of Dice improvement by 6.1, 17, and 14.7, respectively. Its growth in terms of IoU is also of the same trend and it has a 10-12% gain relative to traditional baselines. These steady improvements reaffirm that the combination of classical segmentation prior and deep learning mechanisms is less erroneous and interpretable.

7.7 Overall Performance Discussion

The overall performance discussion covers the corporate culture, organizational structures and functions, subsidiaries, business strategies, and positioning. The overall performance discussion includes corporate culture, organizational structure and functions, subsidiaries, business strategies and positioning.

The obtained results confirm the hypothesis that SOA-UNet provides a successful trade-off between accuracy, speed, and computational complexity. The design of the model guarantees the minimum latency, good precision between boundaries, and consistency in generalization to previously unknown MRI images. The Adaptive Boundary Loss helps to increase sharper edges of segmentation and the Otsu-guided attention helps increase the focus in the region without heavy parameterization.

These design elements work together to make SOA-UNet an effective and fast framework able to segment tumors in real time with clinical quality accuracy. It is able to achieve similar or superior results at a fraction of the computational cost of traditional U-Net architectures and is therefore appropriate in the deployment of medical diagnostic systems or edge-based healthcare AI applications.

8. Discussion or Performance Evaluation

For brain tumor identification, the suggested SOA-UNet system successfully strikes a balance between computing efficiency and segmentation accuracy. The model maintains light inference speed while achieving high boundary delineation precision through the integration of Sparse Convolutional Blocks, Otsu-guided Attention, and Adaptive Boundary Loss. Quantitatively, SOA-UNet outperformed baselines like Simple U-Net and Mobile U-Net, achieving 93.01% accuracy, 88.58% Dice, and 79.50% IoU. The network was directed to concentrate on tumor-relevant areas by the use of Otsu priors, which decreased false positives in background tissues.

This section is an interpretation of the results you already included, the reasons why SOA-UNet is a better choice, what role each of the modules (Sparse Encoding, Otsu-guided Attention, Adaptive Boundary Loss, etc.) plays in the improvement, and what trade-offs (speed vs accuracy) are available.

It should include:

- The meaning of Key Metrics Interpretation: What is meant by 93.01% accuracy and 88.58% Dice.
- Module-Wise Impact Analysis- The degree to which each architectural innovation outperforms the Simple U-Net and Mobile U-Net.
- Computational Discussion -- Why the latency (517.84 ms/img) of your model is feasible in practice in a diagnostic system.
- Strengths and Limitations - e.g., can perform well on heterogenous tumor areas but perhaps needs additional annotated information on the rare tumor cases.
- Clinical Significance-- The way the model can assist radiologists or other real-time diagnostic support systems.

Ethical Approval and Conflict of Interest Statement

The dataset used in this study was obtained from publicly available sources and does not involve identifiable human subjects. The authors declare that there is no conflict of interest regarding this research.

9. Conclusion and Future Work

The paper was able to introduce SOA-UNet (Sparse Otsu-guided Attention U-Net) - a lightweight and high-performing model used in brain tumor segmentation with MRI images. As it was shown in the proposed approach, the combination of classical image processing priors with modern deep learning structures can help to significantly improve the results of segmentation, and the underlying computational performance does not suffer. The model, using a combination of sparse feature encoding, Otsu-guided attention as well as an Adaptive Boundary Loss function, got better results than the traditional and mobile versions of U-Net.

The last experimental analysis revealed that SOA-UNet had a total testing accuracy of 93.01, Dice coefficient of 88.58 and a IoU of 79.50 indicating that there is a great overlap between the predicted and ground-truth tumor areas. The accuracy (84.84) and recall (92.66) rates proved that the model is not only sensitive to tumor areas, but also not sensitive to false identifications. Although this high accuracy was obtained, the model has a latency of 517.84 ms per image and throughput 1.93 FPS, which demonstrates that SOA-UNet is a computationally efficient model and can be used in clinical workflows where the importance of a quick diagnosis is high.

There was a trade-off between the expressiveness and speed, which was facilitated by the architectural design, including depthwise-separable convolutions, SE recalibration, and Conv1DSpatial bottleneck. This increase in interpretability of the segmentation process was aided by the Otsu-guided attention mechanism which explicitly brought out the high probability tumor regions which enabled the network to concentrate on the diagnostically relevant structures. At the same time, the Adaptive Boundary Loss facilitated contour accuracy and alleviated over-smoothing which created sharper and more anatomically typical tumor boundaries. All these advances took the innovations significantly a step higher compared to baseline models, Simple U-Net and Mobile U-Net, which showed worse Dice scores (0.73 and 0.75 respectively) and had poorer boundary localization.

In addition to its quantitative success, SOA-UNet provides clinical reliability and interpretability more than the quantitative, which would be vital in real-world implementation. The outputs of the model have been well-calibrated and are stable in thresholds and are visually consistent among different MRI modalities. Its high inference rate makes it a feasible choice in real-time radiology support systems, embedded diagnostic equipment, such as portable imaging equipment and point-of-care instruments. The framework also fits the emerging tendency of edge AI in healthcare with the preference to deploy high-speed, low power deep learning models in constrained settings.

In the future, there are some promising directions that can benefit this work. The future studies can address the question of multi-modes MRI fusion, i.e. the combination of T1, T2 and FLAIR imaging to give more context-related information. The architecture can be expanded to 3D volumetric SOA-UNet to have full spatial representation of a tumor, which will enhance the accuracy of clinical assessment. The other way is to incorporate the self-supervised or semi-

supervised learning methods to decrease the reliance on manually labeled data, which is still one of the weaknesses of medical imaging studies. Furthermore, it may be enhanced by the inclusion of explainable AI (XAI) modules that would allow radiologists to see the model decision process in real-time.

To sum up, the suggested SOA-UNet positions itself to be the best in closing the shortcomings of accuracy, interpretability, and computational feasibility in the brain tumor segmentation process. It forms a background towards lightweight medical imaging models of the future, where they can be useful in aiding clinical decision-making within a research and clinical healthcare setting.

Acknowledgements The author would like to thank Kalpana Saha(Roy) for the useful discussions.

Author Contributions Pallab Paul wrote the manuscript and prepared the fig ures. All authors reviewed the manuscript.

Funding Not applicable

Data Availability Dataset is downloaded from kaggle.com.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Materials availability Not applicable.

Code availability Not applicable.

10. References

- Alam, M., Ullah, S., Khan, A., & Rehman, A. (2025). Improving segmentation performance through adaptive attention fusion in deep neural networks. *Sensors, 25*(12), 3645. https://www.mdpi.com/1424-8220/25/12/3645
- Chaurasia, A., & Culurciello, E. (2017). LinkNet: Exploiting encoder representations for efficient semantic segmentation. *Proceedings of IEEE VCIP*, 1–4.
- Chen, J., Lu, Y., Yu, Q., Luo, X., & Zhao, Y. (2023). Sparse representation-driven deep models for medical image segmentation. *Artificial Intelligence in Medicine, 141*, 102546.
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with atrous convolution and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(4), 834–848.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint* arXiv:1704.04861.
- Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., ... & Maier-Hein, K. H. (2021). nnU-Net: A self-adapting framework for U-Net-based medical image segmentation. *Nature Methods, 18*(2), 203–211.
- Jha, D., Riegler, M. A., Johansen, D., & Halvorsen, P. (2021). ResUNet++: An advanced architecture for medical image segmentation. *Applied Sciences, 11*(2), 667.
- Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., & Ayed, I. B. (2019). Boundary loss for highly unbalanced segmentation. *Proceedings of MIDL*. arXiv:1812.07032
- Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., & Heng, P.-A. (2018). H-DenseUNet: Hybrid densely connected U-Net for liver and tumor segmentation. *IEEE Transactions on Medical Imaging, 37*(12), 2663–2674.
- Liu, X., Zhang, Y., Wang, Z., & Li, H. (2024). Comprehensive review of deep learning techniques for brain tumor analysis in MRI. *Cognitive Computation and Systems, 6*(1), xx–xx.
- Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings of 3DV*, 565–571.
- Myronenko, A. (2019). 3D MRI brain tumor segmentation using autoencoder regularization. *MICCAI BrainLes 2018, LNCS, 11384*, 311–320.
- Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv preprint* arXiv:1804.03999.
- Paul, S., Kumar, R., & Bose, D. (2022). Sparse attention-based U-Net model for efficient medical image segmentation. *Complex & Intelligent Systems, 8*(4), 3059–3073. https://link.springer.com/article/10.1007/s40747-022-00815-5
- Roy, A. G., Navab, N., & Wachinger, C. (2018). Concurrent spatial and channel squeeze & excitation in fully convolutional networks. *MICCAI*, 421–429.

- Singh, R., Ghosh, S., & Chakraborty, S. (2023). Multi-attention U-Net for brain tumor segmentation using MRI scans. *IEEE Access, 11*, 98425–98437.
- Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging, 15*(1), 29.
- Wang, G., Li, W., Ourselin, S., & Vercauteren, T. (2019). Automatic brain tumor segmentation using cascaded anisotropic CNNs. *MICCAI BrainLes, LNCS*, 178–190.
- Yu, F., & Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions. *ICLR (Workshop track)*. arXiv:1511.07122
- Zhang, C., Wang, H., & Zhao, D. (2022). Hybrid loss design for contour-preserving segmentation in MRI brain images. *Computerized Medical Imaging and Graphics, 94*, 102006.
- Zhang, Y., Yang, Q., & Xu, W. (2020). Lightweight CNNs for medical imaging: Balancing performance and computation. *IEEE Access, 8*, 180398–180409.
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2881–2890.
- Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). UNet++: A nested U-Net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, LNCS, 11045*, 3–11.

List of Figures

Figure 1: Architectural Overview of SOA-UNet

Figure 2: Architecture of Sparse Convolutional Block (SCB) used in SOA-UNet

Figure 3: Having tumor

Figure 4: Don’t have tumor

Figure 5: Result of Proposed Model

Figure 6: Test accuracy per image

Figure 7: ROC curve

Figure 8: Confusion matrix

Figure 9: Comparative Analysis

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- roccurve.png
- compare.png
- confusionmatrix.png
- testaccuracyperimage.png