TÜBİTAK

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2023) 31: 908 – 927
© TÜBİTAK
doi:10.55730/1300-0632.4026

Multi-view brain tumor segmentation (MVBTS): An ensemble of planar and triplanar attention UNets

Snehal RAJPUT[1], **Rupal A KAPDI**[2], **Mehul S RAVAL**[3], **Mohendra ROY**[1,*]

[1]School of Technology, Pandit Deendayal Energy University, Gandhinagar, India
[2]Institute of Technology, Nirma University, Ahmedabad, India
[3]School of Engineering and Applied Science, Ahmedabad University, Ahmedabad, India

Received: 24.06.2023 • **Accepted/Published Online:** 07.10.2023 • **Final Version:** 27.10.2023

Abstract: 3D UNet has achieved high brain tumor segmentation performance but requires high computation, large memory, abundant training data, and has limited interpretability. As an alternative, the paper explores using 2D triplanar (2.5D) processing, which allows images to be examined individually along axial, sagittal, and coronal planes or together. The individual plane captures spatial relationships, and combined planes capture contextual (depth) information. The paper proposes and analyzes an ensemble of uniplanar and triplanar UNets combined with channel and spatial attention for brain tumor segmentation. It investigates the significance of each plane and analyzes the impact of uniplanar and triplanar ensembles with attention to segmentation. We tested the performance of these variants on the BraTS2020 training and validation datasets. The best dice similarity coefficients for enhancing tumor, whole tumor, and tumor core over the training set are 0.712, 0.897, and 0.837, while they are 0.699, 0.875, and 0.782, over the validation set, respectively (obtained through BraTS model evaluation platform). The scores are at par with the leading 2D and 3D BraTS models. Therefore, the proposed approach with fewer parameters (almost 3× less) demonstrates comparable performance to that of a 3D model, making it suitable for brain tumor segmentation in resource-limited settings.

Key words: Attention network, gliomas, triplanar ensemble, brain tumor segmentation, UNet

1. Introduction

Qualitative brain tumor segmentation (BTS) is crucial for the prognosis and diagnosis of brain tumor patients. The total survival tenure of brain tumor (glioma) patients is no longer than two years, [1] because of the high irregularity in shape, structure, location of the tumor, and intensity inhomogeneity within and between tumor tissues. However, early diagnosis can be extremely helpful for the oncologist to fortify treatment planning, which can increase the survival chances of tumor patients. Gliomas have been categorized by the World Health Organization (WHO) as Type IV (the most lethal and common type) of brain tumor. In recent years, advancements in deep neural network techniques have played a crucial role in achieving significant milestones in automated medical image segmentation [2–4].

[*]Correspondence: mohendra.roy@ieee.org

Convolution neural network (CNN) is one of the most commonly used deep neural network models, which can learn complex discriminative features directly from the images. CNN networks have reached human-comparable performance accomplishing complex real-time tasks, including in the medical imaging domain. For example, modifications to UNet [5] were proposed for BTS through the inclusion of dense module, variational autoencoder (VAE), residual module, attention mechanism, convolutional block-attention module (CBAM), and self-attention transformer [3, 6–9].

The brain tumor segmentation (BraTS) challenge [10, 11] comprises delineating tumor tissues from healthy tissues. The task is to label each voxel/pixel of the images as either necrosis (NCR/NET), active/enhancing tumor (ET), peritumoral edema (ED), or background. Again, these labels are combined to form active/enhancing-tumor (ET), tumor-core (TC), and whole tumor (WT) regions. Here, TC includes the ET region, and WT includes TC regions. This challenge assesses the tumor regions by categorizing them into ET, TC, and WT regions. Since gliomas are the most aggressive tumors, their boundaries are frequently uncertain and challenging to distinguish from normal tissues. Additionally, pixels of tumors are very few compared to normal tissues, causing a highly imbalanced dataset, making it very challenging to delineate tumor tissues. Multiple Magnetic Resonance Imaging (MRI) technologies are frequently used to solve this problem, including T1-contrasting (T1C), fluid attenuation inversion recovery (FLAIR), diffusion-MRI (dMRI), proton-density (PD) imaging, T2 (relaxation), and T1 (spin-lattice relaxation). The contrast among these MRI modalities allows each tissue to have a distinctive signature [12, 13].

1.1. 3D, 2D, and 2.5D segmentation models

Some of the contemporary methodologies are: Isensee et al. [8] proposed an ensemble of 3D UNet models where the authors put more emphasis on region-based training, extensive data augmentation, increasing batch size and postprocessing yielding top-ranking segmentation. The Dice Similarity Coefficient (DSC) obtained are 0.798 (ET), 0.912 (WT), and 0.857 (TC). Similarly, Haozhe et al. [7] employed an ensemble of single and cascaded networks to predict segmentation, where the cascade network focuses on segmenting tumor regions from coarser to finer. The obtained DSC are 0.787 (ET), 0.913 (WT), and 0.855 (TC). Similarly, Yuan et al. [14] proposed an ensemble of lightweight attention networks that integrates both low-level and high-level features across various scales. The DSC obtained are 0.793 (ET), 0.911 (WT), and 0.853 (TC). Ma et al. [15] proposed a five-layered single 2D UNet, based on residual connection for segmentation. Further, postprocessing is used to improve segmentation. The DSC obtained are 0.704 (ET), 0.879 (WT), and 0.773 (TC). McKinley et al. [13] proposed an ensemble of triplanar models trained on multiple planar views. However, the initial layers were built on a 3D convolution to obtain complete information from a 3^{rd} dimension. The model was one of the top-ranking models of the BraTS2018 challenge. The obtained DSC are 0.77 (ET), 0.91 (WT), and 0.83 (TC). Similarly, Ali et al. [16] introduced an ensemble of 2D and 3D networks, where the 2D is trained on multi-planar views and the 3D is a lightweight network built using MultiFiber (MF) and dilated convolutions. The obtained DSC are 0.748 (ET), 0.871 (WT), and 0.748 (TC).

A 3D model can correlate spatial and depth information of pixels, enhancing the model's discriminating capability. Nevertheless, massive computations and memory are required to process spatial and depth information of images together [17]. Due to the enormous model parameters and propensity for overfitting, its use can be severely constrained [18]. In comparison, the 2D-based model requires less training time, computation cost, and memory consumption and can be easily optimized. But it only has spatial information of tumors, which limits its segmentation performance. Moreover, due to the intriguing slice-based characteristic of MRI in contemporary medical procedures, some authors emphasized that a 2D approach should be considered [17]. Also, large medical datasets are available in the 2D format, and robust classification models are proposed [17, 19] based on them. Therefore, we desire a model with a 3D network's discriminative ability but with lower memory and computational requirements like that of a 2D model.

MRI volumetric data can be visualized by 2D-triplanar processing. It is a method for producing 2D images along axial, sagittal, and coronal planes. Here, triplanar processing consists of models trained on axial (transverse or X-Z), sagittal (lateral or Y-Z plane), or coronal plane (frontal or Y-X). The resulting image can be examined individually or together to gain insights into anatomy. Due to combining 2D images along different planes, this visualization provides a pseudo sense of depth and is known as 2.5D processing. The 2.5D triplanar processing offers reduced computations, better interpretability, easier data acquisition, and reduced memory demands [17, 20, 21]. As a result, 2.5D approaches have been proposed where UNet was trained on different planar images and combined to provide depth information like a 3D network. These triplanar or 2.5D UNet models require less computation and memory than 3D models and use depth information to provide accurate segmentation [20]. The triplanar models, when hyper-tuned and trained effectively, can perform at par or closer to a 3D model [13, 16]. These models performed better than all other competing techniques in the biomedical segmentation domain [13, 22, 23]. Other parallel approaches were proposed called "Efficient net" or "Light-weight network", where optimized 3D UNets were utilized for the BTS problem [24–26].

The attention mechanism in the UNet facilitates focusing on a specific part or certain features of the input and selectively provides them importance. This allows the network to attend only to relevant information, understand the input better, and improve precision. In this paper, channel attention (CA) focuses on allowing UNet to select channels and assign importance to them during feature extraction. It employs global pooling and learnable transformation to correlations on a per-channel basis. In the BTS problem, CA is introduced in each decoding path within skip connections, where global pooling summarizes spatial information across each channel, and then it is passed through dense or convolution layers to learn channel-wise weights as per the contextual information it carries. CA helps to learn "what" information in the feature map the model has to focus on. In contrast, Spatial attention (SA) lets the model focus on important locations or regions in the activation map. Combining relevant information from both the channel and spatial dimension, thereby suppressing nonrelevant information, can improve the representation and discriminating ability of the model [27, 28]. It uses a convolution layer to generate the attention maps, and rescaling adjusts the importance of each spatial location based on them. SA allows the network to attend important locations and provides good segmentation.

In summary, the paper focuses on the following aspects:

- Developing a hybrid ensemble of planar and triplanar models with different combinations of CA and SA. We show that a hybrid ensemble requires lesser training time, memory, and computation, than the ensemble of 3D models on the validation set of the BraTS2020 challenge.

- We study and analyze the impact of CA, concurrent spatial and channel attention (CSCA), and sequential spatial and channel attention (SSCA) on variants of planar and triplanar models. The former allows the network to capture spatial and channel correlations simultaneously, while the latter derives attention maps along space and channel and merges them.

- Analyzing the impact of planar and triplanar model variants on segmentation.

The rest of this paper is arranged in the following manner: Section 2 examines materials and approaches. The experimental results and discussions are addressed in Section 3. Section 4 presents conclusions and future research.

2. Materials and approaches

2.1. Dataset sources and assessment metrics

The BraTS-2020 [10, 11, 29] challenge's training and validation datasets are employed to train and evaluate the proposed work. This dataset is volumetric and multiparametric in nature. The training set comprises MRI data of 369 patients, while the validation set encompasses images of 125 patients. Each sample includes modalities - T1, T1-contrasted (T1C), T2, and FLAIR. Individual modality has a volume measuring $240(W) \times 240(H) \times 155(C)$. All the images are already coregistered to an identical anatomy template, skull-stripping, and are resampled to a $1mm^3$ resolution. The annotated labels or regions of interest (ROIs) for each training patient have the values 0 for background, 1 for nonenhancing tumor and necrosis tumor (NET or NCR), 2 for peritumoral edema (ED), and 4 for enhancing/active tumor (ET). However, groundtruth labels are not provided for the validation set. We submit the segmentation results to the BraTS organizers' online evaluation platform at (https://ipp.cbica.upenn.edu/) for quantitative evaluation of the models on the validation set.

The evaluation metrics of BraTS-2020 consist of the DSC and 95^{th} percentile of Hausdorff-Distance (HD) in millimeters (mm). The DSC calculates the spatial intersection between the predicted labels and truth labels. In contrast, HD is the largest distance between a point in one set (ground-truth segmentation G) and the nearest point in the other set (predicted segmentation P). In other words, it calculates how well the predicted segmentation boundary is aligned with the ground-truth boundary. Mathematically, DSC and HD are defined as:

$$DSC = \frac{2TP}{FP\,+\,2TP\,+\,FN} \tag{1}$$

$$HD = \max\left\{ \underset{g \epsilon G}{S} \quad \underset{p \epsilon P}{I} \quad d\left(g,p\right), \quad \underset{p \epsilon P}{S} \quad \underset{g \epsilon G}{I} \quad d\left(g,p\right) \right\} \tag{2}$$

where FN, FP, TP, S, and I represent false negative, false-positive, true positive predictions supremum, and infimum, respectively, and $d(g,p)$ is the distance between points $g \epsilon G$ and points $p \epsilon P$. Larger DSC values indicate better segmentation, while a smaller value suggests poor segmentation. For HD, smaller values suggest good alignments of ground truth and predicted segmentation, and larger values signify poor alignment. However, for ranking the competing models, the BraTS challenge mainly considers DSC, whereas HD is used to identify potential instances of over-segmentation or under-segmentation within the tumor subregions through the methods contributed by participants.

2.2. Preprocessing and augmentation

Using the N4 algorithm [30], which corrects the inhomogeneity present in MRI images, all training images are bias-field corrected. Further, nonbrain pixels are removed from the modalities while retrieving 2D images

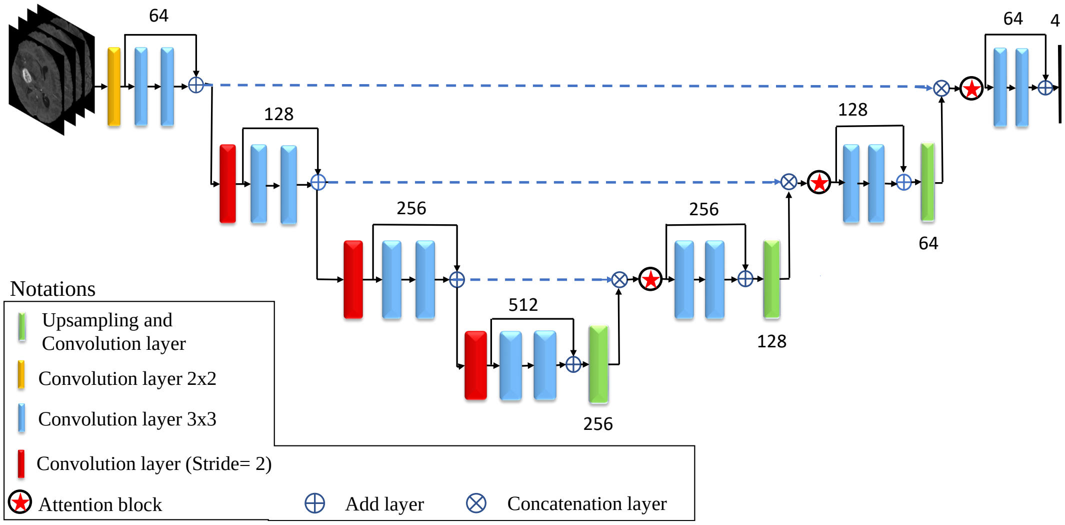


Figure 1. The architecture of the proposed 2D UNet.

because some of the slices do not have brain information, which also helped to overcome the class imbalance. Later, the top and bottom 1% intensities are removed, considering them outliers. Finally, z-score normalization is performed on each slice of the images. The size of the images is $192 \times 152, 152 \times 144$, and 192×144 for axial, coronal, and sagittal planes, respectively. During training, we employ random horizontal and vertical flipping of the images to augment the dataset and alleviate the overfitting problem.

2.3. Proposed ensemble for planar and triplanar UNets with attention

The network structure of the suggested 2D UNet can be viewed in Figure 1. The network has 4 layered encoder-decoder paths; each layer consists of a ResNet [31] like convolution block. The 2D image slices that have been randomly chosen serve as the inputs for the encoder path of every planar model. The dimensions of the input image slices for each individual planar model are specified in the preceding paragraph. In each layer of the encoding path, strided convolution is utilized to halve spatial resolution while simultaneously doubling the count of channels. The initial count of channels is 64; whereas similar to the approach in Noori et al. [32], each ResNet block consists of two convolution blocks with a kernel size of 3×3, along with batch normalization and a Parametric ReLU (PRelu) unit [33]. On the decoder side, each layer decreases the feature maps' numbers by half and doubles their size, using an upsampling layer and a 2×2 convolution layer. Further, each layer feature map on the decoder side is concatenated to the respective encoder layer feature maps. Finally, it is passed to the attention block [34], which recalibrates each channel feature map and forwards it to the subsequent layers. The network uses softmax activation to output 4 channels, each for the NCR/NET, ET, ED, and background. Further, these channels are combined to dissect the tumors into ET, TC, and WT regions. This base network (shown in Figure 1) is trained with input images from the axial plane (AP). We created three models for the AP based on CA, CCSA, and SCSA. The attention modules are implemented on the decoder side (with circular star symbols in Figure 1). Similarly, three models are created for the coronal plane (CP) and three for the sagittal plane (SP). Thus, nine models (three planes × three attention per plane) are trained with the same hyperparameters and network structure.

Different ensembles are created using a suitable combination of the outputs from these nine networks, and they are shown in Table 1. The *planar ensemble* combines outputs for the same plane but with three different attention mechanisms. The outputs of three models from the AP are combined to create an Axial - Ensemble. Similarly, outputs from three CP and three SP models are combined to produce a Coronal - Ensemble and a Sagittal - Ensemble, respectively. The axial planar ensemble is shown in Figure 2a. In the *triplanar ensemble*, we combine outputs from three planes - AP, CP, and SP; each with the same attention mechanism. For example, the model's outputs for AP, SP, and CP with channel attention are combined. Similar ensembles are created using three orthogonal planes with CCSA and SCSA, respectively. The triplanar ensemble is shown in Figure 2b. We also created a *Super-ensemble* by combining planar and triplanar ensembles. It is shown in Figure 3.

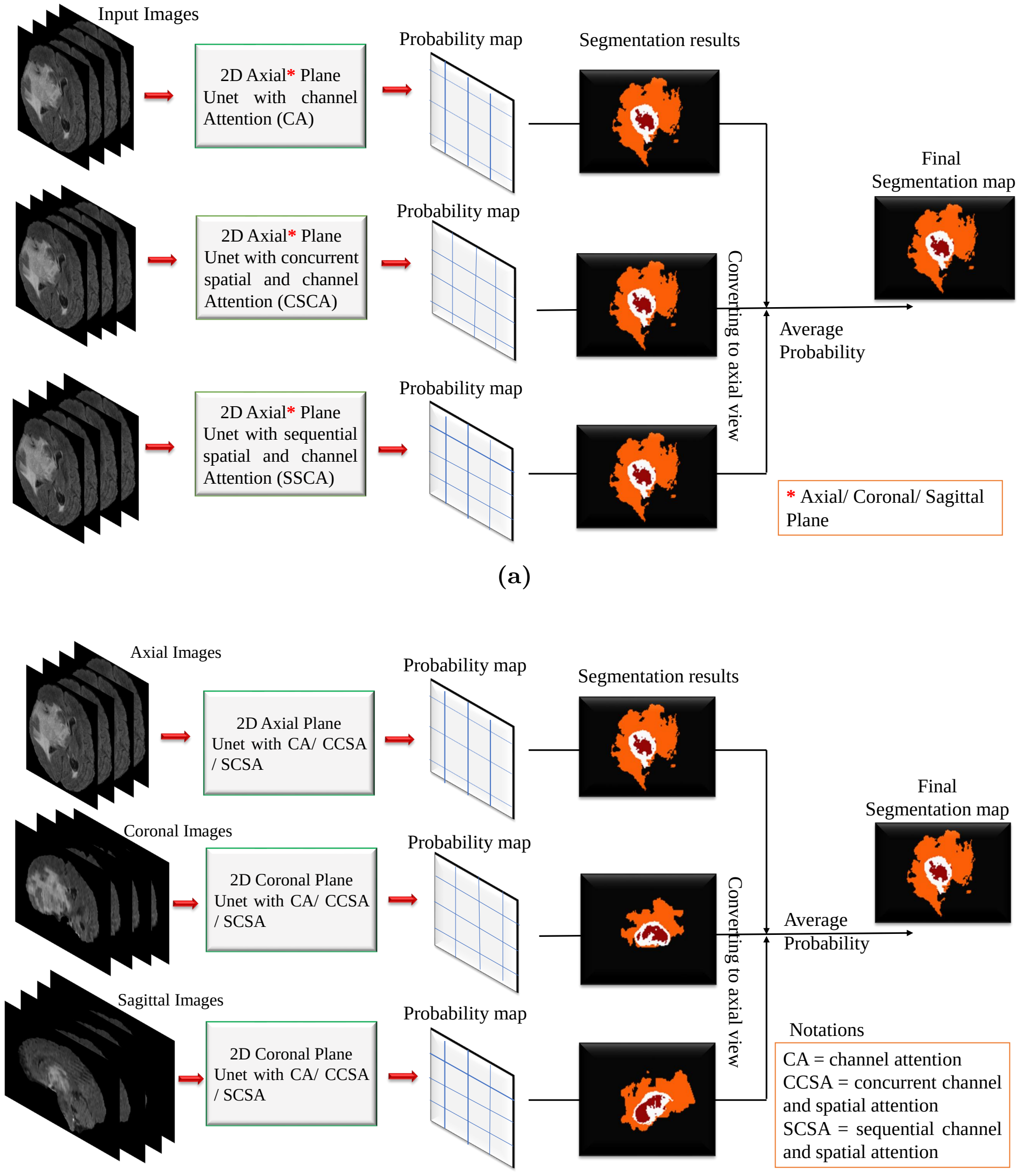
2.4. Attention mechanism

The attention mechanism is [34] was employed to boost the accuracy of the encoder-decoder network. It explores the interdependencies between the channels or spatial locations. The primary purpose of this attention mechanism is to enable the network to use the most pertinent portions of the input feature sequence in an adaptable way, depending on the context it carries. Those input features can be within the feature map (spatial attention) and across the feature map (called channel attention), which can further focus on enhancing performance. The most pertinent input feature vectors receive the highest weights; less informative vectors receive lower weights [34]. All the attention techniques used in this work are simple in structure and marginally impact the model complexity [28, 34].

2.4.1. Channel attention (CA)

Each channel is weighted equally in standard convolution when producing the output feature maps. It produces feature-maps that jointly encode the spatial and channel information by learning filters that capture local spatial patterns along all input channels. The channel attention mechanism weights each channel separately, allowing it to recalibrate semantic attributes per salient features it carries [35, 36]. While significant efforts are invested in refining the combined representation of spatial and channel feature information, there remains a substantial gap in the exploration of encoding spatial-wise and channel-wise information separately.

The contemporary study has made an effort to resolve this problem by formally demonstrating the interdependencies between the feature map channels. Recently, an architecture framework called squeeze and excitation (SE) network [36] was proposed, which has a simple structure and is easy to integrate into the existing network. Therefore, we have also used SE techniques to implement channel attention. SE works on two principles: First, it utilizes global average pooling (GAP) to squeeze the feature maps into a single numeric value (number of channels) to gain global statistics of the channels. Second, the excitation operator captures nonlinear and nonmutual exclusive relations between the channels and a gating mechanism that utilizes a sigmoid activation to assign weights to each channel based on the information it holds. In summary, it squeezes across the spatial dimensions and reweights across the channel dimensions to improve feature representations by exploring interdependencies between feature channels. Mathematically, a squeeze operation on the feature map U can be defined as:


Input Images
2D Axial* Plane Unet with channel Attention (CA)
2D Axial* Plane Unet with concurrent spatial and channel Attention (CSCA)
2D Axial* Plane Unet with sequential spatial and channel Attention (SSCA)
Probability map
Probability map
Probability map
Segmentation results
Converting to axial view
Average Probability
Final Segmentation map
* Axial/ Coronal/ Sagittal Plane
Axial Images
Coronal Images
Sagittal Images
2D Axial Plane Unet with CA/ CCSA / SCSA
2D Coronal Plane Unet with CA/ CCSA / SCSA
2D Coronal Plane Unet with CA/ CCSA / SCSA
Probability map
Probability map
Probability map
Segmentation results
Converting to axial view
Average Probability
Final Segmentation map
Notations
CA = channel attention
CCSA = concurrent channel and spatial attention
SCSA = sequential channel and spatial attention

(a)

(b)

Figure 2. The idea of planar and triplanar networks. (a) The axial planar network, where segmented outcomes from CA, CCSA, and SCSA networks trained on axial images, are combined to generate an outcome. Similarly, we can create a Coronal - Ensemble and a Sagittal - Ensemble. (b) An overview of the triplanar network, where segmented outcomes generated from individual attention networks (e.g., CA network) trained on axial, coronal, and sagittal images are combined to generate an outcome. Similar segmented outcomes are generated from CCSA and SCSA attention networks trained on three orthogonal planes.

$$z = F_{sq}(U) = \sum_{c=1}^{C} \left[\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \right] = GAP \tag{3}$$

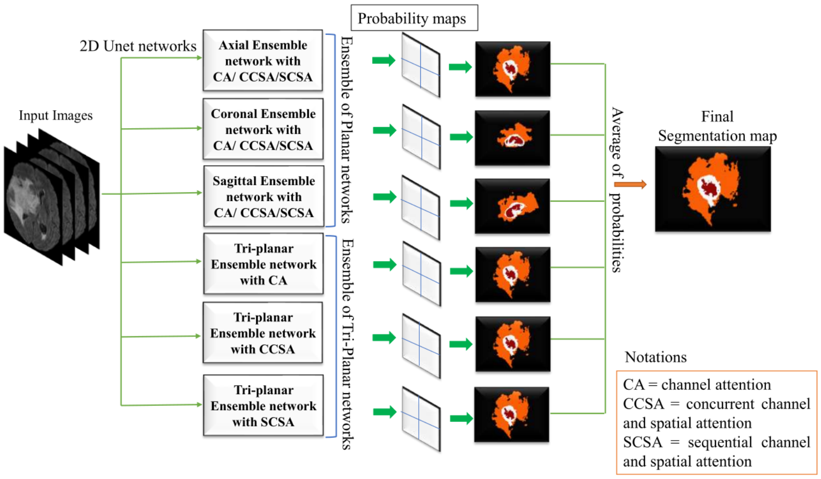

2D Unet networks
Input Images
Axial Ensemble network with CA/ CCSA/SCSA
Coronal Ensemble network with CA/ CCSA/SCSA
Sagittal Ensemble network with CA/ CCSA/SCSA
Tri-planar Ensemble network with CA
Tri-planar Ensemble network with CCSA
Tri-planar Ensemble network with SCSA
Ensemble of Planar networks
Ensemble of Tri-Planar networks
Probability maps
Average of probabilities
Final Segmentation map
Notations
CA = channel attention
CCSA = concurrent channel and spatial attention
SCSA = sequential channel and spatial attention

Figure 3. The architecture of super-ensemble, where planar and triplanar networks' outcomes were combined to generate BTS.

where, $(U) = \{u_1, u_2, ...u_c\}$ is a feature map which consists of multiple filters $\{1, 2, ...c^{th}\}$. The excitation can be defined as:

$$F_e x = F(z, W) = \sigma \left[\overbrace{W_2(\underbrace{\delta(W_1(z))}_{1^{st}\text{FC layer}})}^{2^{st}\text{ FC layer}} \right] \tag{4}$$

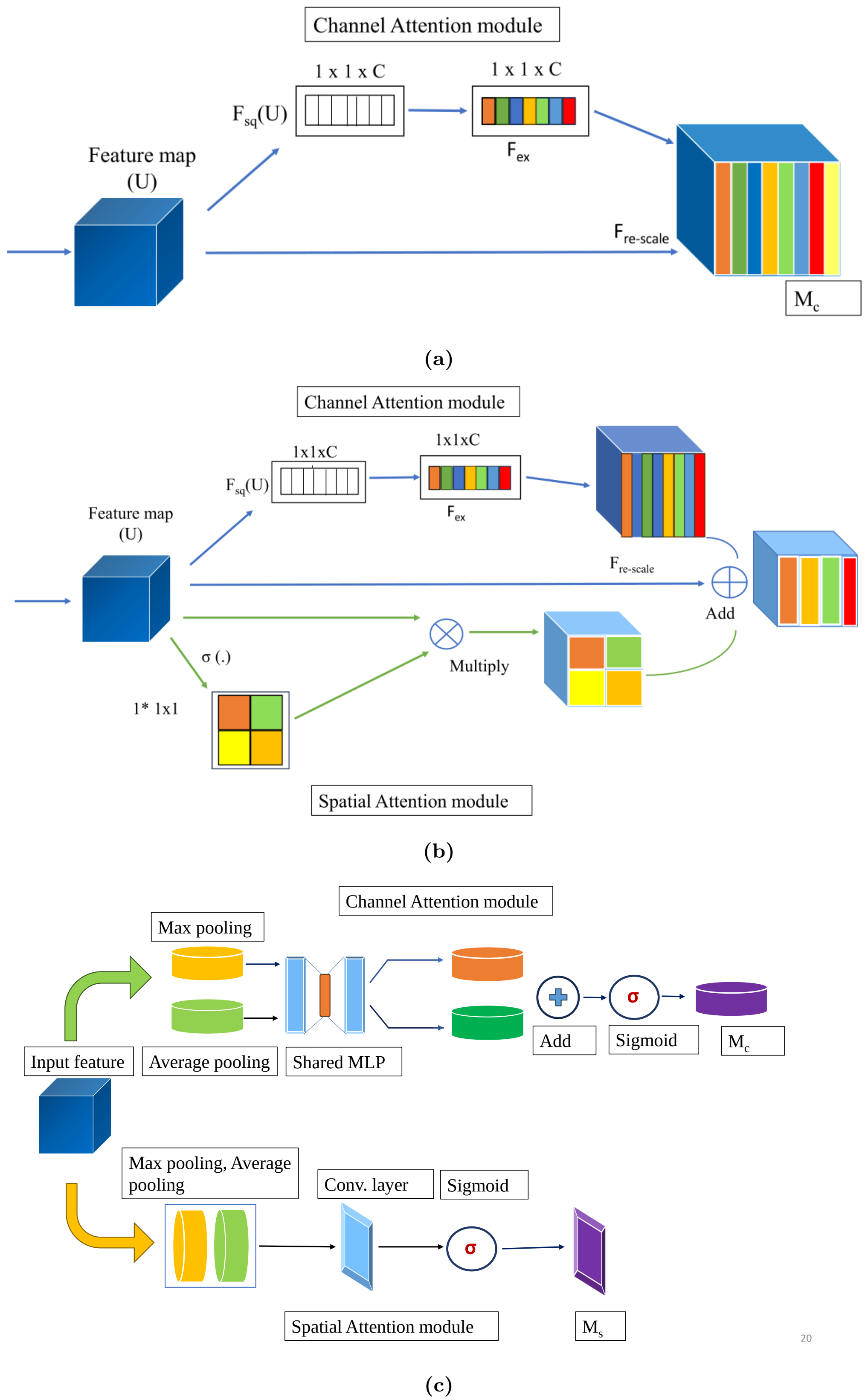
where σ is a sigmoid function and W_1 and W_2 are the weights of 1^{st} and 2^{st} FC layers, respectively, and z is the feature map after the squeeze operation. The CA module is illustrated in Figure 4a.

2.4.2. Concurrent channel and spatial attention

Similarly, taking clues from the study mentioned in [27] suggests that calibrating the feature maps parallelly along spatial and channel, then combining the results can enhance the spatial and channel information. Therefore we experimented with infusing concurrent channel and spatial squeeze-and-excitation blocks into the segmentation network. The CCSA module is illustrated in Figure 4b.

2.4.3. Sequential channel and spatial attention

This module sequentially creates attention maps spanning the channel and spatial dimensions from an intermediary feature map. It subsequently aggregates these attention maps with the input feature map, allowing for an adaptable recalibration of features [28]. It incrementally derives a 1D channel attention map (A_c), such that $A_c \ \epsilon \ \{C \times 1 \times 1\}$ and a 2D spatial attention map (A_s), such that $A_s \ \epsilon \ \{1 \times H \times W\}$. For the channel and spatial module, channel and spatial information aggregation is done using max-pool and average-pool operations that capture effective features across the channel and spatial dimensions. Mathematically, the whole attention



(a)

(b)

(c)

Figure 4. (a) A channel attention(CA) module [36] (b) Concurrent channel and spatial attention (CCSA) [27] (c) Sequential channel and spatial attention (SCSA) [28].

procedure can be described as follows:

$$F' = A_c(F) \oplus F \tag{5}$$

$$F'' = A_s(F') \oplus F' \tag{6}$$

where, $\oplus$ indicates elementwise multiplication, F is a feature map, $A_c(F)$ is the channel attention map, $A_s(F')$ is the spatial attention map and F'' is the final output.
Whereas $A_c(F)$ and $A_s(F')$ can be defined as:

$$A_c(F) = \sigma\left[mlp\{AP(F)\} + mlp\{MP(F)\}\right] \tag{7}$$

$$A_s(F') = \sigma\left[\, f_c^{7\times 7}\{mlp\{AP(F')\} \cdot mlp\{MP(F')\}\}\,\right] \tag{8}$$

where σ denotes the sigmoid function, mlp is multi-layered perceptron, AP, MP are the average and maxpooling operations on the feature map respectively, $\cdot$ is the concatenation operation, and $f_c^{7\times 7}$ indicates a convolution operation with a 7×7 kernel size. The SCSA module is illustrated in Figure 4c.

2.5. Network training and optimization

The individual network extracts 2D input images randomly and applies image augmentation on the fly to train the models. The input sizes for the axial, coronal, and sagittal models are 192×152, 152×144, and 192×144, respectively. The batch size is 15, and models are trained on five cross-validation sets. Based on the outcomes from our literature review (demonstrated in Table 2), we employed a hybrid loss function that integrates the Generalized Dice-Loss (GDL) [37] with categorical cross-entropy to alleviate class imbalance, as shown in Equation 9. GDL recalibrates the DSC to consider the significance of each class and balances their impact based on their prevalence. It helps the loss function to prioritize underrepresented classes, leading to an enhanced capability of the model to segment these specific classes accurately. Further, the initial training learning rate is 8×10^{-3}, which reduces by a factor of 0.5 when the validation-loss does not decrease for 30 epochs. Further, the model's training stops if the validation-loss does not reduce for 50 epochs. The models were trained using a stochastic gradient-descent optimizer. The training hyperparameters are the same for all the models. The deep learning framework is developed using Keras and Tensorflow2.2. The models are trained on Quadro RTX 5000 system having a 16GB GPU and 128GB RAM.

$$\begin{gathered}
Loss = GD_{loss}(G,P) + CE_{loss}(G,P) \\
GD_{loss}(G,P) = 1 - 2\frac{\sum_{l=1}^{C}(W_l \times \sum_{i=1}^{N} g_{li} \times p_{li})}{\sum_{l=1}^{C}(W_l \times \sum_{i=1}^{N} g_{li} \times p_{li})} \\
CE_{loss}(G,P) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{C}(g_{li} \times \log p_{li})
\end{gathered} \tag{9}$$

where G indicates groundtruth label, P is the predicted label, $W_l = \frac{1}{(\sum_{i=1}^{N} g_{li})^2}$ is the adaptive weight for the l^{th} channel, p_{li} is the predicted label l for voxel i, g_{li} is the ground truth label l for voxel i, $g \epsilon G$ and $p \epsilon P$.

3. Experimental results and discussions

As discussed in Section 2.3, we trained nine models (three distinct attention-based architectures along three different anatomical planes in 2D slices) in this work. In addition to submitting and analyzing results for each variant separately, we create ensembles by adding their output probabilities and averaging across them to obtain final segmentation results for training and validation BraTS2020 datasets. Table 1 shows a quantitative performance evaluation of each variant on training and validation datasets. We observe from all the proposed variants, ensemble models give the best segmentation results. Among the 16 variants, Table 1 highlights the five best-performing models on both the validation and training datasets. Their graphical performance representation on the BraTS2020 validation dataset is shown in Figure 5. Considering the performance of *Super-Ensemble* among all variants shown in Table 1, we use it for comparison with leading BraTS2020 methods in subsequent analysis.

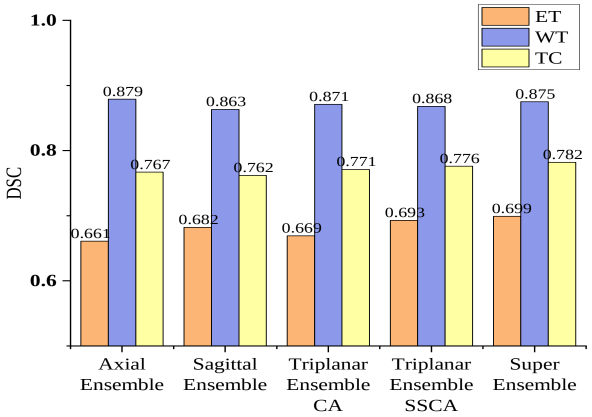


Figure 5. The performance representation of our top five models on the BraTS2020 validation dataset. On Y-axis, DSC is shown, whereas on the X-axis are the model names, and vertical bars represent ET, WT, and TC tumor regions.

3.1. Planar analysis along with parameter counts

Comparing the planar models, we observe that each excels in segmenting different regions. Training models using different planar views allow models to capture distinct properties of lesions. For instance, among AP model types, the CA model is able to dissect WT (DSC: 0.873) and TC (DSC: 0.754) regions well, whereas the SCSA model dissects all the regions well, such as, for ET (DSC: 0.670), WT (DSC: 0.866), TC (DSC:0.760). Similarly, among CP model types, we observe that all the models (CA, CCSA, and SCSA) have moderate DSCs for all the lesions. However, across all model types, the DSC for ET is better than AP model types. Whereas, among SP model types, the CA model has dissected ET (DSC: 0.663) and WT (DSC: 0.850) regions well, the CCSA has dissected TC (DSC: 0.759) well, and SCSA has dissected all the lesions well (DSC: 0.680(ET), 0.853(WT), 0.756(TC)). The parameter counts for all the model types, on average, is $10.27\,M$.

Further, a comparison is also conducted among planar ensembles. We observed that both the *Axial ensemble* and the *Sagittal ensemble* are able to dissect all the lesions well. Where the DSC for *Axial ensemble* are 0.661, 0.879, and 0.767 for ET, WT, and TC, respectively, and for *Sagittal ensemble*, they are 0.682, 0.863, and 0.762 for ET, WT, and TC, respectively. Furthermore, *Coronal ensemble* models trained on coronal view images have lower performance than the other two planar images. The reason can be that some modalities are acquired in 2D axial or sagittal view, and each slice has a specific thickness. So while reconstructing it to coronal views, the reconstruction requires an interpolation technique to fill that thickness area. It causes anisotropy in resolution [38], meaning discrepancy in resolution along different planes, which occurs due to differences in the voxel size and acquisition parameters. It can be explicitly seen in Figure 6, where 6(a) is the axial view of the Flair image and the resolution is intact compared to coronal and sagittal slice views shown in 6(b) and 6(c). The coronal slice has the least detailed structure. In their work on BTS, McHugh et al. [17] also highlighted the issue of data loss caused by the coronal plane view. On average, the parameter counts of our planar ensemble models are $30.8\,M$.

Table 1. Quantitative evaluation of each variant of models on the training and validation datasets of BraTS2020 challenge. Bold text signifies the best five performing models on validation and training sets. Results from models are indicated using the same color for training and validation sets. The first two rows show the top-ranking model of the challenge. Validation results are obtained from the challenge's online assessment platform https://ipp.cbica.upenn.edu/.

				Mean Dice similarity coefficient (DSC)			Mean Haudorff Distance (HD)		
Dataset	Model Type	Model Name	Parameter Numbers in millions (M)	ET	WT	TC	ET	WT	TC
Validation	3D Ensemble of 5 models	Isensee et al. [8]	30.2 × 5 = 151 M	0.798	0.912	0.857	26.410	3.730	5.640
Validation	3D Ensemble of 11 models	Yuan et al. [14]	16.5 × 11 = 181.5 M	0.793	0.911	0.853	18.196	4.097	5.888
Validation	Axial	CA	10.246 M	0.610	0.873	0.754	56.627	9.532	14.038
Validation	Axial	CCSA	10.247 M	0.618	0.854	0.738	52.666	13.759	17.896
Validation	Axial	SCSA	10.333 M	0.670	0.866	0.760	46.227	9.328	15.22
Validation	Coronal	CA	10.246 M	0.633	0.836	0.747	54.469	19.712	22.547
Validation	Coronal	CCSA	10.247 M	0.629	0.827	0.743	58.344	23.447	22.538
Validation	Coronal	SCSA	10.333 M	0.642	0.825	0.738	52.114	25.978	19.806
Validation	Sagittal	CA	10.246 M	0.663	0.850	0.737	48.509	10.414	15.969
Validation	Sagittal	CCSA	10.247 M	0.652	0.835	0.759	49.567	14.381	15.010
Validation	Sagittal	SCSA	10.333 M	0.680	0.853	0.756	46.402	17.448	18.758
Validation	Axial Ensemble	CA-CCSA-SCSA	30.825 M	**0.661**	**0.879**	**0.767**	**41.232**	**6.754**	**12.034**
Validation	Coronal Ensemble	CA-CCSA-SCSA	30.825 M	0.648	0.849	0.762	53.748	16.502	15.713
Validation	Sagittal Ensemble	CA-CCSA-SCSA	30.825 M	**0.682**	**0.863**	**0.762**	**43.714**	**11.391**	**16.799**
Validation	Triplanar Ensemble	CA	30.737 M	**0.669**	**0.871**	**0.771**	**44.512**	**6.929**	**12.558**
Validation	Triplanar Ensemble	CCSA	30.739 M	0.667	0.860	0.773	45.395	9.764	13.708
Validation	Triplanar Ensemble	SCSA	30.999 M	**0.693**	**0.868**	**0.776**	**42.274**	**10.120**	**13.989**
Validation	Super-Ensemble	CA-CCSA-SCSA	92.47 M	**0.699**	**0.875**	**0.782**	**36.752**	**8.037**	**14.846**
Training	Axial Ensemble	CA-CCSA-SCSA	30.825 M	**0.688**	**0.897**	**0.837**	**41.136**	**7.056**	**7.699**
Training	Sagittal Ensemble	CA-CCSA-SCSA	30.825 M	**0.715**	**0.883**	**0.812**	**40.200**	**6.415**	**8.338**
Training	Triplanar Ensemble	CA	30.737 M	**0.692**	**0.893**	**0.827**	**41.489**	**7.130**	**8.523**
Training	Triplanar Ensemble	SCSA	30.999 M	**0.727**	**0.885**	**0.826**	**39.745**	**9.099**	**9.734**
Training	Super-Ensemble	CA-CCSA-SCSA	92.47 M	**0.712**	**0.897**	**0.837**	**40.310**	**6.378**	**7.114**

3.2. Triplanar analysis along with parameter counts

Triplanar is an ensemble of multi-view (axial, coronal, and sagittal) model types. The triplanar ensemble of all the models has been able to dissect all the regions well. Triplanar Ensemble with CA has DSC 0.669 (ET), 0.871 (WT), and 0.771 (TC), respectively, whereas, with CCSA and SCSA, it is 0.667 (ET), 0.860 (WT), 0.773 (TC) and 0.693 (ET) 0.868 (WT), 0.776 (TC) respectively. Likewise, comparing among triplanar models, both CA and SCSA exhibit superior performance compared to the CCSA model. On average, the parameter counts for all the triplanar ensemble models are $30.8\,M$.

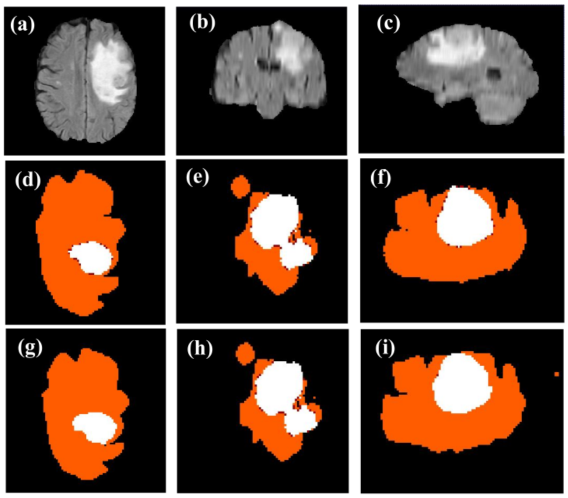


Figure 6. Visual comparisons of a sample from the training set: (a), (b), and (c) are the axial, coronal, and sagittal views of the input FLAIR image whereas (d), (e), and (f) are the ground-truth segmentation for all the anatomical views. Similarly, (g), (h), and (i) are the predicted segmentation for all the anatomical views. Here, the ET region is shown in white color, TC in brown, and WT in orange. Comparing (a), (b), and (c), axial slices contained the most detailed structural information, whereas coronal had the least detailed (high-contrast and high-resolution) structure.

3.3. Planar-triplanar analysis

Further comparing the ensemble models, i.e.planar models with triplanar models, we can observe that the latter (2.5D) have performed better than planar models, which are based on 2D processing. Using information from different planes to segment tumors has been fruitful for triplanar models. The model is able to integrate some 3D information while predicting segmentation labels. Triplanar models with CA and SCSA attention modules have performed better than other ensemble models. Triplanar with CA excels in segmenting WT (DSC: 0.871) and TC (DSC: 0.771), whereas with SCSA modules, it excels in segmenting ET (DSC: 0.693) and TC (DSC: 0.776). This shows triplanar ensemble models equipped with specific attention mechanisms have learned distinct features, significantly enhancing the model's ability to discriminate between multiple lesions.

Additionally, we observe more favorable outcomes in the ensemble than those produced by each individual model. The ensemble techniques reduce the variance from the individual models. This average ensemble technique is more straightforward and faster to estimate the mean over the probability distribution, hence it is widely used for segmentation [39]. The best-performing is the *Super-Ensemble* model, which is the ensemble of all 3 planar and triplanar ensemble models. It has DSC of 0.699 (ET), 0.875 (WT), and 0.782 (TC) with

$92.47\,M$ trainable parameters. It has the least HD for all the lesions, in comparison to all the variants, which signifies that the segmented boundaries are closely aligned with the ground truth boundaries.

3.4. Attention mechanisms analysis

Similarly, observing attention modules - CA, CCSA, and SCSA, we experimentally verify that sequentially refining channel and spatial features helps the model learn distinct features from multiple lesions or regions-of-interests (ROIs), resulting in better segmentation results. Another reason is the combination of max and average pooling helped aggregate finer features useful for learning finer discriminating features of target lesions [28]. Comparing the performance of all three networks, CCSA has the least DSC, and SCSA has the most. Woo et al. also mention that a sequential combination of these attention modules gave satisfactory results compared to a parallel combination [28]. Also, all these attention modules are light in weight, adding only 86016 (CA), 86912 (CCSA), and 173334 (SCSA) parameters to the ordinary model (without the attention module having $10.16\,M$ trainable parameters).

3.5. Parameters analysis and comparisons with BraTS2020 leading models

Further, comparing the number of parameters of the proposed models with the top-ranking BraTS2020 models in Table 1, which is an ensemble of five 3D-UNet [8], and eleven 3D-UNet [14] models. We note that each UNet in [8] and [14] have $30.2\,M$ and $16.5\,M$ parameters, respectively, which require enormous computation resources at a time. In contrast, each UNet in a proposed ensemble (e.g., *Triplanar ensemble*) requires $\approx 10.2M$ trainable parameters at a time. This is because the proposed models are trained individually (one model at a time), and predictions are ensembled to get final segmentation results. Thus we reduce almost $3\times$, $1.6\times$ computation and memory requirement in comparison to that of the top-ranking models, keeping the results comparable for WT and TC regions. The proposed method has achieved comparable segmentation results for WT (DSC: 0.875) and TC (DSC: 0.782) with less trainable parameters ($10.33\,M$), thus requiring limited computation resources. Likewise, the ensemble of UNets [8] and [14] are $1.68\times$ and $2\times$ more than the proposed *Super-ensemble* method.

Further, we have compared the proposed *Super-Ensemble* method with various leading BraTS2020 models shown in Table 2. Rows 2 - 5 show top-ranking models of the BraTS2020 challenge. A comparison of parameters between the proposed method and two top-ranking models is presented in Table 1, which we elaborated upon in the preceding paragraph. Additionally, by comparing our model with other leading BraTS2020 models (shaded in Table 2), we can clearly notice that the proposed *Super-Ensemble* (planar and triplanar together) has outperformed models in segmenting regions shown as shaded in Table 2. Comparing these results, we conclude that an optimized 2D model(s) can compete or perform nearly equally to the 3D model(s).

For qualitative performance evaluation on the training set, random $\frac{1}{3}$ samples from the training set are selected, and segmentation results are obtained from the Super ensemble model. We evaluate segmentation results by submitting them to the BraTS challenge assessment platform. The DSC for the *Super-Ensemble* over the BraTS2020 training set is 0.712, 0.897, and 0.837 for ET, WT, and TC, respectively, as can be seen in Table 1. Also, the training results of the other four best models can be found in Table 1. At the same time, visual comparisons of the proposed model can be observed in Figure 6, which compares an input FLAIR image, ground truth, and segmented map across all the planar views. This justifies the effectiveness of our proposed ensemble model.

3.6. Ablation study

We carried out an ablation to study the impact of convolution layer kernel size k at the convolution block in yellow color and upsample block in green color in Figure 1. We experiment with 3×3 and 2×2 kernel size *(k)* at the preliminary level. A 2×2 kernel size increases the performance metrics and comparatively reduces the number of parameters. For a CP network with a CA attention module with $k = 3$, the DSC is 0.647 (ET), 0.823 (WT), and 0.734 (TC). Whereas for CSCA and SSCA, it is 0.614 (ET), 0.798 (WT), 0.710 (TC) and, 0.548 (ET), 0.748 (WT), 0.622 (TC). These results can be compared with kernel size $k = 2$ as shown in $6-8^{th}$ rows of Table 1, showing a significant increase in scores with $k = 2$.

Table 2. Comparison between the proposed *Super-Ensemble* method and leading methods of BraTS2020 challenge on validation dataset. All the methods were trained on the BraTS2020 training set. The details of the proposed *Super-Ensemble* method are shown in the first row. Rows 2^{nd}-5^{th} show top-ranking models of the BraTS2020 challenge. The proposed *Super-Ensemble* method outperforms the other lead models in accurately segmenting lesions depicted by cells shaded in gray color.

			Mean Dice similarity coefficient (DSC)			Mean Haudorff Distance (HD)		
Model name	UNet Model Type	Loss Type	ET	WT	TC	ET	WT	TC
Proposed method	2D Super-Ensemble	Generalized Dice-loss + Cross Entropy	0.699	0.875	0.782	36.752	8.037	14.846
Isensee et al. [8]	3D Ensemble	Dice-loss + Cross entropy	0.798	0.911	0.857	26.410	3.730	5.640
Haozhe et al. [7]	3D Ensemble	Generalized Dice-loss + Binary Cross Entropy	0.787	0.913	0.855	26.575	4.184	4.972
Yuan et al. [14]	3D Ensemble	Jaccard distance loss + focal loss	0.793	0.911	0.853	18.196	4.097	5.888
Liu et al. [40]	3D	Dice-loss + Cross Entropy	0.764	0.882	0.801	21.390	6.490	6.680
Messaoudi et al. [19]	3D	Dice-loss + Cross Entropy	0.654	0.841	0.680	NA	NA	NA
Asenjo et al. [41]	2D, 3D Ensemble	Cross Entropy + Dice-loss + HD loss	0.886	0.782	0.736	30.468	4.696	18.185
Ballestar et al. [42]	3D	Generalized Dice Loss	0.720	0.840	0.790	37.970	10.930	12.240
Soltaninejad et al. [43]	3D	NA	0.660	0.870	0.800	47.330	6.910	7.800
Ma et al. [15]	2D	Dice loss + Binary Cross Entropy	0.704	0.879	0.773	NA	NA	NA
Ali et al. [16]	2D, 3D Ensemble	NA	0.748	0.871	0.748	3.929	9.428	10.090
Agravat et al. [44]	3D	Dice loss + Focal loss	0.763	0.873	0.753	27.704	7.038	10.873
Xu et al. [45]	2D	Generalized Dice Loss + Cross Entropy	0.673	0.861	0.704	40.608	7.942	15.750
Bommineni et al. [46]	3D	Cross Entropy	0.718	0.884	0.788	30.767	4.834	9.258
Colman et al. [47]	2D	Cross Entropy	0.676	0.886	0.672	47.620	12.110	15.740

Similarly, in the CA network, we experiment with the reduction ratio of 16 and 8 in the dense layer. The segmentation results are better with a reduction ratio of $(r = 8)$. Therefore we also keep $r = 8$ in all the channel attention modules used in CCSA and SCSA networks. Further, in the SCSA block, for implementing spatial attention, we replaced $k = 7 \times 7$ kernel size as shown in Equation 8, with $k = 3 \times 3$ to balance parameter numbers, but the model's performance deteriorates. For an axial planar model with $k = 3 \times 3$ in SCSA attention block, the DSC is 0.670 (ET), 0.866 (WT), and 0.760 (TC). These results can be compared with kernel size $k = 7$ shown in Table 1 (5^{th} rows), showing a significant increase in scores with $k = 7$.

3.7. Limitation of the proposed work

First, the proposed triplanar model does not fully utilize depth dimension. Subsequently, the model struggles to effectively eliminate minor, isolated false positive labels. Furthermore, the current model's parameters pose a challenge for real-time implementation on edge devices. The size of the weight file for the proposed models is large, and it will not fit into the memory of edge devices. Lastly, the presence of anisotropy within image resolution is also a notable consideration, and the method's performance will degrade due to the artificial transformation of image planes, e.g., from axial to coronal.

4. Conclusions and future research

In this work, we explored a 2D triplanar-ensemble network that has segmenting performance of a 3D model. Also, this approach is influenced by the fact that most large publicly available medical datasets consist of 2D images. Therefore, we studied 2D networks and optimized them (2.5D) to improve their performance metrics. The proposed network uses three 2D UNet networks to generate axial, coronal, and sagittal slice predictions. These predictions are subsequently integrated into a final multiple-view prediction, which enables partial capturing of spatial information in the depth dimension. Additionally, infusing attention mechanisms into the network causes the inclusion of relevant information from channel and spatial dimensions, thereby suppressing unnecessary information which in turn improves the discriminating power of the segmentation model.

We can observe from this study that an ensemble of the triplanar network based on UNet provides robust BTS and requires fewer parameters, and thus requires less computational memory. Additionally, we observed that training models across multiple planes enable them to learn and discriminate between different ROIs. Models trained using axial and sagittal planar views can segment ROIs more robustly than those trained on coronal planar view. Combining ensembles of these models further enhances the overall segmentation performance. Likewise, we observe that incorporating channel and spatial attention into the network in a sequential manner enables the model to learn significant features from channel and spatial dimensions effectively. Moreover, incorporating channel attention alone into the network also increases the model's discriminating capabilities. In other words, SCSA and CA attention-based models have shown better segmenting performance than CCSA.

In summary, optimizing 2D models using the attention-based triplanar approach can compete with 3D models with limited complexity and computation requirements. These can be extremely useful when implemented in resource-constrained environments or integrated with legacy systems where datasets are in 2D images. The proposed 2D network has shown comparable results to the top-performing BraTS2020 models. However, the performance and parameter numbers can be further optimized.

Future research can focus on many factors: 1. Investigating postprocessing techniques aimed at reducing false positive regions. Refining network architecture and optimizing hyperparameters (such as exploring variants of both 2D and/or 3D models and methods based on Graphical neural network (GNN) may offer opportunities to optimize resource consumption, improve performance, and extend the scope of clinical applicability. This is particularly pertinent given the abundance of extensive publicly accessible datasets. 2. Incorporating Multimodal Medical Image Fusion (MMIF) techniques, which combine multiple medical modalities into a single image. This fusion can help to combine necessary and valuable information captured by multi-modalities into one, which can improve the model's discriminatory ability. and yet, these methods demand more training time and specialized GPUs, resulting in higher computational expenses than other techniques. 3. Incorporating functional-imaging techniques (e.g., positron emission tomography (PET) image, functional Magnetic-Resonance Imaging (fMRI) image into conventional MRI can help Deep learning (DL) models learn the physiological, metabolic, and biological details of tumor lesions. 4. Integrating explainable AI techniques can assist in understanding and subsequently fine-tuning segmentation decisions, thereby strengthening the models' segmentation capabilities.

5. Acknowledgement

M. Roy acknowledges the seed grant No. $ORSP/R\&D/PDPU/2019/MR/RO051$ of PDEU (for the computing facility), the core research grant No. $CRG/2020/000869$ of the Science and Engineering Research Board (SERB), Government of India and the project grant no $GUJCOST/STI/2021-22/3873ofGUJCOST$, Government of Gujarat, India. M. S. Raval acknowledges the grant number. $GUJCOST/STI/2021-22/3858$ of Gujarat Council of Science and Technology (GUJCOST), Government of Gujarat, India for computing facility to support the work.

References

[1] Chato L, Kachroo P, Latifi S. An automatic overall survival time prediction system for glioma brain tumor patients based on volumetric and shape features. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6. Springer 2021: 352-65.

[2] Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging. 2019;32:582-96.

[3] Jia Q, Shu H. Bitr-unet: a cnn-transformer combined network for mri brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part II. Springer 2022: 3-14.

[4] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data. 2021; 8:1-74.

[5] 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. Springer 2015: 234-41.

[6] Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. 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. Springer; 2019: 311-20.

[7] Jia H, Bai C, Cai W, Huang H, Xia Y. HNF-Netv2 for brain tumor segmentation using multi-modal MR imaging. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part II. Springer; 2022: 106-15.

[8] Isensee F, Jäger PF, Full PM, Vollmuth P, Maier-Hein KH. nnU-Net for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6. Springer; 2021. p. 118-32.

[9] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR et al. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries:

7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part I. Springer; 2022. p. 272-84.

[10] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data. 2017;4 (1):1-13.

[11] Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:181102629. 2018.

[12] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A et al. Brain tumor segmentation with deep neural networks. Medical image analysis. 2017;35:18-31.

[13] McKinley R, Rebsamen M, Meier R, Wiest R. Triplanar ensemble of 3D-to-2D CNNs with label-uncertainty for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5. Springer; 2020:379-87.

[14] Yuan Y. Automatic brain tumor segmentation with scale attention network. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part I 6. Springer; 2021: 285-94.

[15] Ma S, Zhang Z, Ding J, Li X, Tang J et al. A deep supervision CNN network for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021:158-67.

[16] Ali MJ, Akram MT, Saleem H, Raza B, Shahid AR. Glioma segmentation using ensemble of 2D/3D U-Nets and survival prediction using multiple features fusion. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021: 189-99.

[17] McHugh H, Talou GM, Wang A. 2d dense-unet: A clinically valid approach to automated glioma segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021: 69-80.

[18] Zhao C, Zhao Z, Zeng Q, Feng Y. MVP U-Net: multi-view pointwise U-net for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021: 93-103.

[19] Messaoudi H, Belaid A, Allaoui ML, Zetout A, Allili MS et al. Efficient embedding network for 3D brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part I 6. Springer; 2021. p. 252-62.

[20] Sundaresan V, Griffanti L, Jenkinson M. Brain tumour segmentation using a triplanar ensemble of U-Nets on MR images. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part I. Springer; 2021: 340-53.

[21] Sundaresan V, Zamboni G, Rothwell PM, Jenkinson M, Griffanti L. Triplanar ensemble U-Net model for white matter hyperintensities segmentation on MR images. Medical image analysis. 2021;73:102184.

[22] Ottesen JA, Yi D, Tong E, Iv M, Latysheva A et al. 2.5 D and 3D segmentation of brain metastases with deep learning on multinational MRI data. Frontiers in Neuroinformatics. 2023.

[23] Hitziger S, Ling WX, Fritz T, D'Albis T, Lemke A et al. Triplanar U-Net with lesion-wise voting for the segmentation of new lesions on longitudinal MRI studies. Frontiers in Neuroscience. 2022;16.

[24] Cheng X, Jiang Z, Sun Q, Zhang J. Memory-efficient cascade 3D U-Net for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5. Springer; 2020. p. 242-53.

[25] Chen C, Liu X, Ding M, Zheng J, Li J. 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22. Springer; 2019. p. 184-92.

[26] Li J, Zheng J, Ding M, Yu H. Multi-branch sharing network for real-time 3D brain tumor segmentation. Journal of Real-Time Image Processing. 2021:1-11.

[27] Roy AG, Navab N, Wachinger C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. Springer; 2018. p. 421-9.

[28] Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV); 2018. p. 3-19.

[29] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging. 2014;34 (10):1993-2024.

[30] Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A et al. N4ITK: improved N3 bias correction. IEEE transactions on medical imaging. 2010;29 (6):1310-20.

[31] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016: 770-8.

[32] Noori M, Bahri A, Mohammadi K. Attention-guided version of 2D UNet for automatic brain tumor segmentation. In: 2019 9th international conference on computer and knowledge engineering (ICCKE). IEEE; 2019: 269-75.

[33] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision 2015: 1026-34.

[34] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L et al. Attention is all you need. Advances in neural information processing systems. 2017;30.

[35] Chen L, Zhang H, Xiao J, Nie L, Shao J et al. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 5659-67.

[36] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 7132-41.

[37] Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3. Springer; 2017. p. 240-8.

[38] Wu Z, Wei J, Wang J, Li R. Slice imputation: Multiple intermediate slices interpolation for anisotropic 3D medical image segmentation. Computers in Biology and Medicine. 2022;147:105667.

[39] Sawant S, Erick F, Schmidkonz C, Ramming A, Lang E et al. Comparing ensemble methods combined with different aggregating models using micrograph cell segmentation as an initial application example. Journal of Pathology Informatics. 2023;14:100304.

[40] Liu C, Ding W, Li L, Zhang Z, Pei C et al. Brain tumor segmentation network using attention-based fusion and spatial relationship constraint. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part I 6. Springer; 2021. p. 219-29.

[41] Marti Asenjo J, Martinez-Larraz Solís A. MRI Brain Tumor Segmentation Using a 2D-3D U-Net Ensemble. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part I 6. Springer; 2021. p. 354-66.

[42] Ballestar LM, Vilaplana V. MRI brain tumor segmentation and uncertainty estimation using 3D-UNet architectures. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part I 6. Springer; 2021. p. 376-90.

[43] Soltaninejad M, Pridmore T, Pound M. Efficient MRI brain tumor segmentation using multi-resolution encoder-decoder networks. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021. p. 30-9.

[44] Agravat RR, Raval MS. 3D semantic segmentation of brain tumor for overall survival prediction. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021. p. 215-27.

[45] Xu JH, Teng WPK, Wang XJ, Nürnberger A. A deep supervised u-attention net for pixel-wise brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021. p. 278-89.

[46] Bommineni VL. Piecenet: A redundant unet ensemble. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021. p. 331-41.

[47] Colman J, Zhang L, Duan W, Ye X. DR-Unet104 for Multimodal MRI brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 2020, Revised Selected Papers, Part II 6. Springer; 2021. p. 410-9.