STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, February 2026, pp 1236–1247.
Published online in International Academic Press (www.IAPress.org)

An Efficient Deep Learning Approach based on 3D Res-UNet for Multimodal Brain Tumor Segmentation

Khaoula ECHINE*, Aziz DAROUICHI

Laboratory of Systems and Computer Engineering, University of Cadi Ayyad, Marrakech, Morocco

Abstract Accurate segmentation of brain tumors in MRI is an important aspect of accurate clinical diagnostics and sound surgical planning. Tumor boundary accuracy is fundamental to informing assessment of a patient's condition, acquired through continuous expansion of Deep Learning based segmentation approaches. This study introduced an effective deep learning method to perform 3D segmentation of multimodal MRI images by enhancing the Res-UNet architecture. Our proposed model, 3D ASPP-ResUNet, incorporates an ASPP (Atrous Spatial Pyramid Pooling) module to better exploit multi-spatial scale features. The BraTS 2020 dataset has been used for training and evaluation. This model performs well according to the dice metric for different tumor regions attaining Dice scores of 0.7578 for TC (tumor core), 0.7025 for ET (enhancing tumor) and 0.8273 for WT (whole tumor). Furthermore, we observed that the 3D ASPP-ResUNet was better than currently leading models with respect to segmentation performance metrics that we defined as the Dice coefficients.

Keywords Brain tumor, MRI, Deep Learning, Res-UNet, ASPP, 3D ASPP-ResUNet, BraTS 2020.

AMS 2020 subject classifications 68T45, 92C55, 68T07

DOI: 10.19139/soic-2310-5070-2786

1. Introduction

Brain tumors represent a serious intracranial pathology that can pose significant health risks.These abnormal clumps of cells disrupt the brain, our body's complex control center. Unlike healthy brain tissue, brain tumors grow uncontrollably, causing harm to their surroundings [1]. This wreaks havoc on the brain's ability to coordinate our body, making brain tumors a serious threat to human health.

In order to diagnose brain tumors, medical imaging is essential. This includes techniques like CT scans, ultrasounds, and X-rays, but MRI reigns is widely regarded as the most effective modality [2]. Its high resolution and ability to capture detailed brain anatomy make it the go-to method for tumor detection. Accurate tumor segmentation, identifying the tumor's boundaries within an MRI scan, is crucial for safe and effective treatment, especially surgery where preserving healthy tissue is paramount. While this segmentation can be done by radiologists manually, it takes a lot of time and is prone to errors. This is where automated computer-aided segmentation comes in a much needed development to streamline the process and ensure consistent results.

Brain tumor segmentation (BTS) is tricky. Tumor cells come in all shapes, sizes, and appearances, and can pop up anywhere in the brain. Researchers are constantly looking for better ways to tackle this challenge, and recently, one useful artifact that has risen is deep learning. Contrary to conventional machine learning, deep learning is able to automatically identify and select the most significant features in an image. For complex jobs like brain tumor segmentation, this is especially beneficial.

*Correspondence to: Khaoula ECHINE (Email: k.echine.ced@uca.ac.ma). Laboratory of Computer and Systems Engineering (L2IS), Department of Computer Science, Faculty of Sciences and Techniques, University of Cadi Ayyad, Marrakech, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

This paper introduces a new deep learning method called 3D ASPP-ResUNet. It builds upon the existing 3D Res-UNet architecture for segmenting 3D MRI scans from multimodalities. We've added a unique module called ASPP (Atrous Spatial Pyramid Pooling) to increase accuracy. With the help of this module, the model may more accurately separate tumors by capturing data from the image at various scales.

The paper has the following structure: Deep learning models related to segmentation of brain tumor are reviewed in Section 2. Sections 3 & 4, unveiling our approach analyze the core of our research: 3D ASPP-ResUNet architecture and its components. Section 5 tests our innovation by examining the outcomes of our suggested method and evaluating its efficacy through comparison studies. Section 6 wraps up the work by providing a summary of our results and suggesting possible avenues for further investigation.

2. Related Works

Due to its unique ability to automatically diagnose the size, type and location of tumors, brain tumor segmentation has been a popular task in recent times. Many researchers have developed some compelling techniques or methods for segmenting brain tumors. There are many deep learning approaches used in challenge of medical imaging segmentation, including fully convolutional networks, cascaded architectures, approaches with dilated convolutions, and the top-down approaches that are very popular.

Long et al. (2015) [3] actually initiated the concept of the fully convolutional network (FCN) by replacing fully connected layers typically found in neural networks with 1×1 convolutions. This is problematic for segmentation because all global image semantic context is lost, and the FCN learned to only exploit local representations. To remedy this situation, the FCN leveraged its basic model VGG-16 and incorporated the lower and upper layers into its segmentation layer using the skip connections.

Havaei et al. (2017) [4] proposed a cascading structure of a convolutional neural network (CNN) that accurately estimates local dependencies. The network leverages both more global contextual features and local feature leveraging an extra source of data from the output of a simple CNN as the input for the overall CNN. Two phases of training is performed over the BraTS 2013 dataset to mitigate tumor label imbalance. Despite the architecture demonstrating a respectable overall performance (Having an overall tumor segmentation Dice score of 0.84 and a specificity of 0.88), the network performance in terms of tumor core segmentation, and tumor highlight (region for tumor) had limitations with Dice scores of 0.71 and 0.57.

Ranjbarzadeh et al. (2021) [5] developed a novel architecture that focuses solely on a tiny part of brain image rather than the entire image to reduce computation time and overcome overfitting issues associated with a cascaded model. They proposed the C-ConvNet/C-CNN model, a straightforward and effective cascaded convolutional network. Both global and local features are used in two different ways by this C-CNN model. Additionally, to increase the precision of tumor segmentation, they integrated a Distance Wise Attention mechanism. Even if the suggested method produced remarkable outcomes on the BraTS 2018 dataset, limitations remain when it is faced with a tumor volume that exceeds one-third of the entire brain.

In order to improve segmentation whilst maintaining resolution, Yu and Koltun (2015) [6] implemented a modification of standard CNNs, dilated convolution (or atrous convolution). In these processes, the benefits of dilated convolution help by providing linear rate increases for parameters and exponential rate increases for receptive fields, which provides an efficient way to extract features. However, since dilated convolution may reduce disassociation, or unlink pixels to their global context, it may also cause errors in classification.

Yang et al. (2020) [7] presented a modified U-Net known as DCU-Net (Dilated Convolution U-Net) for use in segmentation of brain tumor. They recommended that use of multi-scale spatial pyramid pooling and dilated convolutions could increase the receptive field efficiently, and therefore, DCU-Net did not include a max pooling option in the down-sampling path as was done in some other dilated convolution networks, but instead used up-sampling. To optimize the performance of segmentation, the authors used dilated residual blocks and skip

connections. Assessment of the BraTS 2018 dataset indicated that overall, the model performed quite well, achieving strong Dice similarity coefficients in several areas of tumor.

Cahall et al. (2021) [8] introduced DIU-Net (Dilated Inception U-Net) approach for BTS, utilizing U-Net as a base. Both local structural information and global contextual information may be extracted due to this architecture's combination of dilated convolutions and inception modules in the contraction and expansion paths. The DIU-Net architecture demonstrates superior computational efficiency compared to U-Net, while having fewer parameters, according to the promising outcomes on the dataset of BraTS 2018 .

The asymmetric 3D U-Net design presented by Belaid et al. (2020) [9], however, incorporates the 2D EfficientNet classification model in the encoding stage. There are two steps that the encoder and the decoder go through. The input data is three-dimensional, but the early stages of the encoder begin lowering the third dimension according to the EfficientNet 2D network input share. Furthermore, there are two stages to the decoding process: a 2D and a 3D phase. The BraTS 2020 dataset was used by the authors to illustrate the efficacy of their models.

Using four similar encoder structures, Zhang et al. (2021) [10] developed the ME-Net model to enhance the V-Net design, it has a single path for resolution reduction. Each encoder is designed to segment one of the four modalities in parallel. They also implemented multiple skip connections in order to fuse the feature maps that were produced by encoders. Then the combined feature maps are provided to the decoder as input.

More recent work, such as nnU-Net by Isensee et al. (2021) [11], is a self-configuring deep learning framework that automatically configures neural network architecture, preprocessing, post-processing and training to each dataset in biomedical image segmentation. The proprietary framework, nnU-Net, therefore relieves you of the burden of design choices and still achieves competitive performance over a variety of segmentation tasks. nnU-Net gained acceptance given it outperformed or matched specialized methods on a wide range of open datasets so you can get state-of-the-art segmentation without even having to be an expert in hyperparameter tuning.

Methods that are based on transformers have also found their way into volumetric medical imaging in multiple ways. Hybrid encoders (e.g., TransUNet [12] and TransBTS [13]) use CNNs for local detail, but also Transformers for context and have validation BraTS accuracy that are similar. Fully 3D Transformers (e.g., UNETR [14] and hierarchical Swin encoders such as SwinUNETR [15] and SwinBTS [16] closely align to this approach using volumetric input and tend to be the most often seen in high-scoring BraTS submission or validation metrics.

3. Research methodology

In this section, we have briefly outlined what dataset to use and discussed all pre-processing steps associated with that data. We then highlighted all aspects of the suggested model, including the multi-scale feature extractor module, encoder and decoder.

3.1. Dataset

We used the BRATS2020 dataset [17] for our experiments, which included 3D brain MRI scans. Each scan comes in four flavors: T1, T1ce, T2, and FLAIR, providing a comprehensive picture of the brain. To ensure consistency, all scans are formatted as .nii.gz files and resized to $240 \times 240 \times 155$ voxels. Each scan has been manually segmented by one to four experts, providing ground truth annotations for:

- Necrosis and non-enhancing tumor (NCR/NET)
- Enhancing tumor (ET)
- Edema (ED)
- Non-tumor region (Background)

We trained the 3D ASPP-ResUNet model on the BraTS 2020 training dataset, available online to the public through Kaggle [17]. Because the validation dataset lacked ground truth labels, the training dataset was split into 75% for train, 15% for validation, and 10% for test. This dataset consists of 344 scans taken from 34 patients for testing, 52 for validation and 258 for training.

3.2. Preprocessing

While the BRATS 2020 organizers prepped the data, MRI scans can be finicky. Variations in scanner strength can cause intensity inconsistencies. To address this and improve segmentation results, we performed some additional preprocessing steps on the training data:

- **Normalization:** Each MRI volume (T1ce, T2, and FLAIR) was thus normalized by MinMax scaling to the [0, 1] range independently. There is no outlier clipping. The min and max for the scaling was determined by the minimum and maximum intensity of each volume.
- **Resizing:** Due to memory constraints, we resize all images from $240 \times 240 \times 155$ to $128 \times 128 \times 128$.
- **Multimodal Magic:** Since each MRI scan offers valuable information, we combined every modalities (T1, T1ce, T2 and FLAIR) into a single image. During training, the model sees these combined images ($128 \times 128 \times 128$ with a depth of 4 representing the modalities) to get the most comprehensive picture possible.

Figure 1 illustrates a MRI image before and after preprocessing.

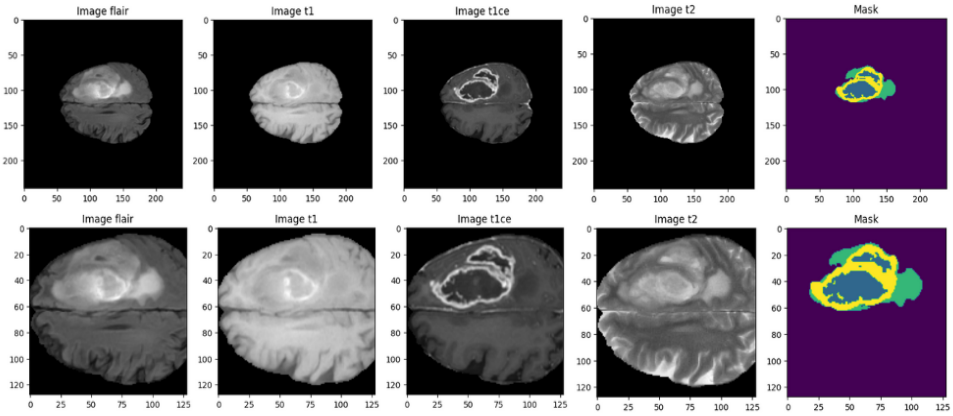

Figure 1. MRI sample image and the ground truth for each of the four modalities. Bottom figures: following preprocessing; top figures: before preprocessing.

4. Network architecture

We provided a comprehensive description of the architecture of the proposed model with special regard to its core component: the Atrous Spatial Pyramid Pooling (ASPP), detailed in the next subsection.

4.1. ASPP module

CNNs use a general technique called Atrous Spatial Pyramid Pooling (ASPP) to gather information at different scales in an image. It can be applied to various tasks that need precise object feature analysis, such as semantic segmentation and image classification.

The ASPP model uses parallel dilated / atrous convolutions at various rates to extract features of objects at multiple scales. Five parallel sub-networks make up the ASPP structure, as seen in figure 2. To gather global information about the image, the first sub-network employs a global average pooling procedure. Following batch

normalization, the generated features are put into a $3 \times 3 \times 3$ convolution with 256 filters. The second sub-network uses a $1 \times 1 \times 1$ convolution, while the other three sub-networks use $3 \times 3 \times 3$ convolutions with different dilation rates of (6,12,18). These dilation rates are in accordance with those used in the 2D ASPP module of DeepLabV3 [18], enabling the network to capture multiscale receptive fields in volumetric MRI data while keeping spatial consistency and avoiding sparse sampling in 3D feature maps. All these sub-networks are equipped with 256 filters and Batch Normalization. Following their extraction from each branch, the features are concatenated and put through another $1 \times 1 \times 1$ convolution using 256 filters and Batch Normalization.

The general form of the 3D atrous convolution is mathematically represented as:

$$y(p) = \sum_{k} x(p + r.k).w(k) \tag{1}$$

where x is the input feature map, p the current voxel position, r the dilation rate, and $w(k)$ denotes the scalar weight in the convolution kernel.

The output size is computed as:

$$output\_size = \frac{input\_size + 2 \times padding - r \times (kernel\_size - 1) - 1)}{stride} + 1 \tag{2}$$

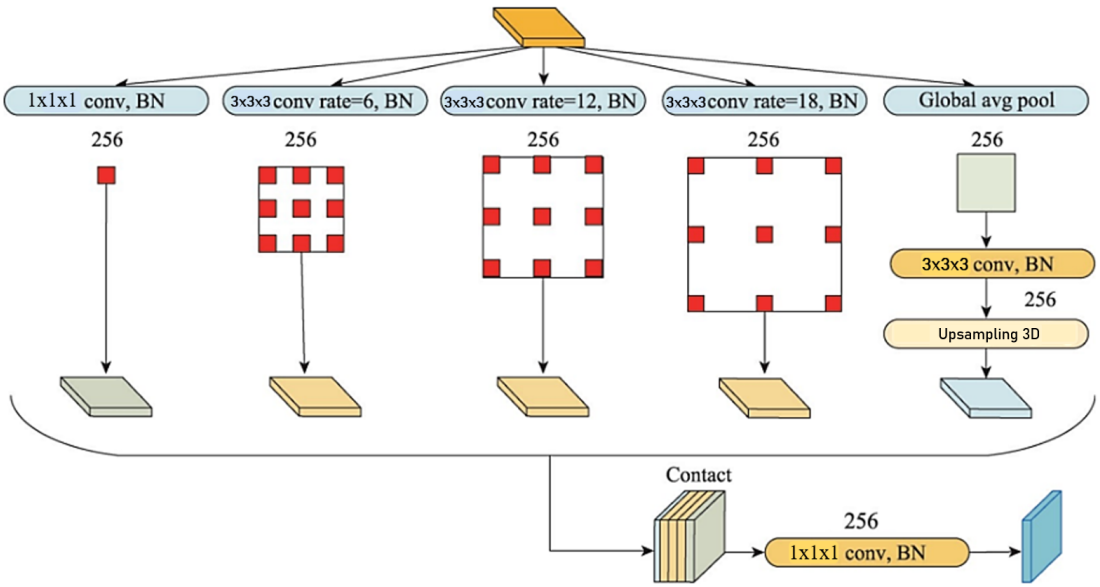

Figure 2. Structure of the ASPP module [19]

4.2. Proposed model Architecture

Our proposed model architecture, 3D ASPP-ResUNet, has a typical Res-UNet architecture consisting of encoder and decoder sides linked by skip connections. We have fused the Res-UNet backbone with a ASPP module (see figure 3). The encoder and decoder feature channel depths of 64, 128, 256, and 512, respectively, and the downsampling step after the final encoder block outputs the bridge features of 1024 channels and spatial dimensions of $8 \times 8 \times 8$. The model's input has the size of $128 \times 128 \times 128 \times 4$ representing the 4 MRI modalities (T1, T1ce, T2, and FLAIR). The output has the same feature spatial dimensions of $128 \times 128 \times 128 \times 4$ representing the 4 segmentation classes: non-enhancing/necrotic tumor core (NCR/NET), enhancing tumor (ET), edema (ED), and background. Ground truth labels were one-hot encoded for use with categorical cross-entropy loss.

Figure 3 illustrates how the encoder path is composed of four residual blocks, each of which has two convolutional layers combined with Batch Normalization and activation functions ReLU. A skip connection is incorporated to preserve the identity mapping, when there is a mismatch in input and output dimension. Formally, for an input tensor x, the corresponding residual block produces:

$$F(x) = BN(W_2 \times ReLU(BN(W_1 \times x))) \tag{3}$$

and output:

$$y = ReLU(F(x) + W_s \times x) \tag{4}$$

Where W_1, W_2 are $3 \times 3 \times 3$ convolution kernels and W_s is $1 \times 1 \times 1$ convolution for channel alignment. The use of this residual architecture allows for improvement in gradient flow and stability of model training as the 3D feature domain grows deeper.

There are 64 feature maps in the first residual block, each measuring $128 \times 128 \times 128$ pixels. The quantity of feature maps increases by double and their dimension is cut in half at each subsampling level. As a result, there are 512 feature maps for the fourth-level residual block, each measuring $16 \times 16 \times 16$.

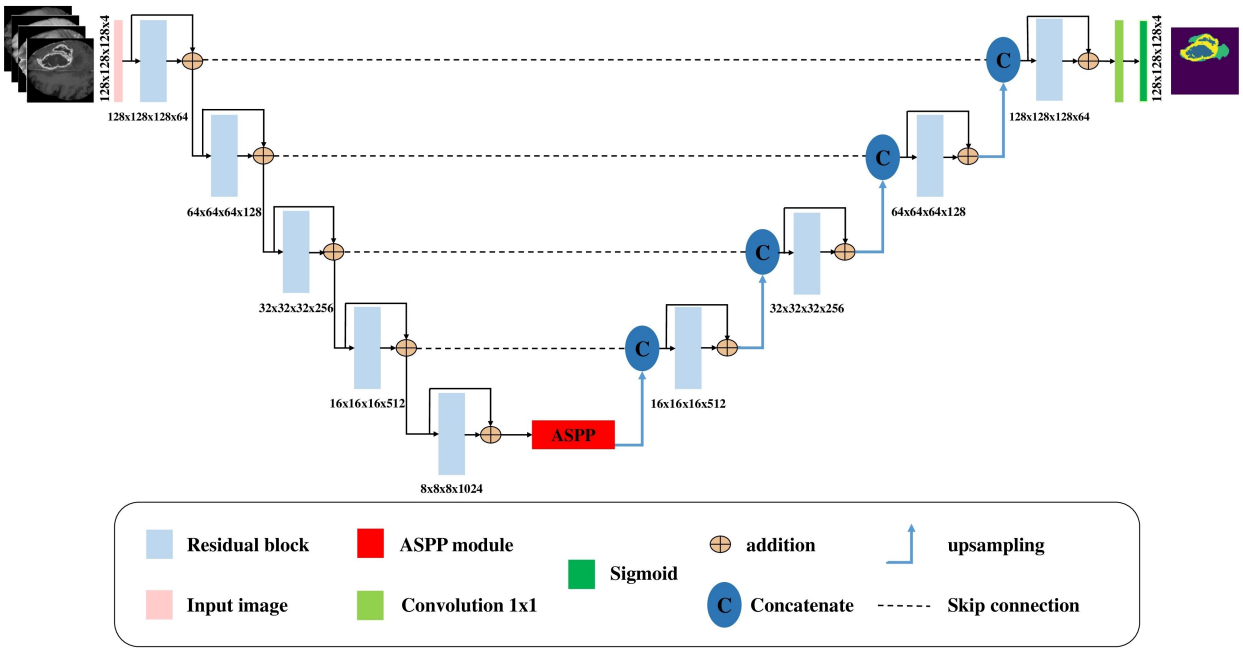

Figure 3. Architecture of 3D ASPP-ResUNet model.

Another downsampling operation (3D max pooling with a stride of 2) is performed after the fourth residual block to derive a bridge representation of 1024 feature maps with size of $8 \times 8 \times 8$. The bridge block serves as an intermediate representation between the encoder and decoder, and ASPP takes to the output of the bridge block before upsampling and passing it to the decoder path.

The decoder path reflects the encoder with four levels, having 64, 128, 256, and 512 features, respectively. Each level begins with a 2×2 upsampling layer and is followed by a residual block. Skip connections are used to connect each decoder block to its corresponding encoder block. The feature maps are doubled in size and cut in half at each decoder block. After passing through a softmax activation function and a classification layer, predictions generated by the final residual block are compared with the ground truth.

5. Results and discussion

The outcomes of our methodology are presented in this part along with an explanation of the implementation process and the different measures utilized to assess the effectiveness of our model.

5.1. Evaluation measure

Tumor Regions Used for Validation : To train their algorithms, participants were provided multi-class labels, and the image data consisted of visual representations of tumor formations. In order to evaluate how well the segmentation techniques worked, we divided the tumor structures into the three overlapping zones depicted in figure 4, which corresponds to clinical uses like tumor volume measurement. These areas are:

1. The "whole tumor" (WT), including all regions of tumor;
2. The "tumor core" (TC), encompassing all tumor structures with the exception of edema;
3. The "enhancing tumor" (ET), representing only the enhancing core components, typically associated with high-grade cases. [20]

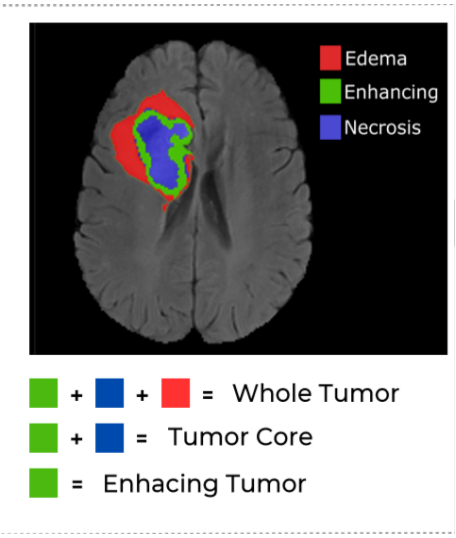

Figure 4. Different tumor regions

The performance of the approach is assessed using the most widely-used metric for BTS tasks: Coefficient Dice. It is a conventional metric for segmenting brain tumor that measures how similar between the ground truth and expected segmentation in order to provide an objective calculation of the accuracy of segmentation decisions.

Performance was evaluated for all tumor regions, overall segmentation (Dice Coef), but also by subregions, whole tumor (Dice WT), tumor core (Dice TC) and enhancing tumor (Dice ET). Dice coefficients are calculated using the equation as below on premise of predicted segmentation and ground truth data.

$$DSC = 2 \times \frac{|X \cap Y|}{|X| + |Y|} \tag{5}$$

5.2. Implementation detail

We developed 3D ASPP-ResUNet model in Jupyter Notebook using necessary Python libraries like Keras, TensorFlow, and NiBabel which helps in carrying on multiple experiments in medical image processing. The

training was conducted on a Geforce RTX 4090 GPU (24GB RAM), through the online platform Vast.ai. Due to the limitations of the GPU memory, batch size = 1 was necessary for training with a total of 120 epochs completed. The Adam optimizer (learning rate = 0.0001) and categorical cross-entropy loss, which is defined by equation 6 below, were employed for training the network.

$$L = -\sum_{c=1}^{C} y_c log(\hat{y}_c) \tag{6}$$

where y_c and $\hat{y}_c$ refer to the ground truth and predicted probability of class c, respectively, and $C = 4$ denotes the number of segmented classes (whole tumor, enhancing tumor, tumor core, and background).

Even though datasets for brain tumor segmentation are fundamentally class-imbalance, we avoided using class weighting explicitly. All MRI volumes were consistently resized from $240 \times 240 \times 240$ to $128 \times 128 \times 128$ to maintain equal spatial dimensions. The use of residual connections and ASPP module helped with multiscale feature representation which partially reduced the impact of class imbalance.

5.3. Quantitative results

In order to intuitively and fairly evaluate the segmentation results obtained with these different enhancement strategies, ablation experiments were conducted to demonstrate that ASPP can strengthen the network's segmentation. To guarantee the credibility of comparing results, two networks were trained separately with identical experimental conditions: the optimizer, learning rate, batch size and training schedule were the same for Res-UNet baseline and the ASPP-ResUNet. Random seeds were not fixed either; therefore, the result could also differ based on a random weight initialization. The metrics that were measured on the validation data (Dice WT, Dice ET and Dice TC) are listed in table 1.

Table 1. Comparison of tumor region dice coefficients in ablation studies using the evaluation set after 120 epochs.

Method	**Dice WT**	**Dice ET**	**Dice TC**
Res-UNet	0.7758	0.6348	0.6534
Res-UNet+ASPP	0.8242	0.6462	0.6846

In table 2, the evaluation measures (Dice WT, Dice ET and Dice TC) calculated on the test set are reported. To analyze the statistical significance of the observed differences in Dice scores per-patient between the baseline and proposed model, we performed the Wilcoxon signed-rank test.

Table 2. Comparison of tumor region dice coefficients in ablation studies using the test set after 120 epochs.

Method	**Dice WT**	**Dice ET**	**Dice TC**
Res-UNet	0.8068	0.6963	0.7551
Res-UNet + ASPP	**0.8273**	**0.7025**	**0.7578**
Wilcoxon (p)	**0.0166**	0.1453	0.5228

The tables above illustrate that, the 3D ASPP-ResUNet model's metrics have significantly improved. After 120 epochs, the Dice coefficients were 0.7758, 0.6348 and 0.6534, respectively in validation, and they increased to 0.8242, 0.6462 and 0.6846 after adding the ASPP module. Similarly, for the testing phase, the Dice coefficients were 0.8068, 0.6963 and 0.7551, and after integration of ASPP, they increased to 0.8273, 0.7025 and 0.7578, for WT, ET and TC, respectively.

From the statistical analyses, we conclude that the proposed 3D Res-UNet with ASPP obtained a statistically significant value for the Whole Tumor (WT) region of interest (Wilcoxon $p = 0.0166$) while values for Enhancing

Tumor (ET) and Tumor Core (TC) regions were not statistically significant ($p > 0.05$). We theorize this is due to the fact that the ASPP module was designed in such a manner as to increase the capacity of the model to capture contextual data on multiple information and global spatial relationships. These traits are of great importance for delineating the extent of the tumor (WT) which has more extensive and heterogeneous structures. The ET and TC subregions on the other hand, are conditioned in part by the fact that they are more localized smaller and less homogeneous and subject to range of intensity differentials, which renders them more difficult to segment accurately and less prone to improvement by global context. The uniformity of numerical improvement seen in all three subregions suggests that the ASPP module improves the overall robustness of the model in terms of output even when absolute statistical significance is not present for every class.

Figure 5 represents the training and validation loss per epoch as well as the dice coefficient curve. The training Dice coefficient gradually increases and plateaus at around 0.99, whereas the validation Dice coefficient increases sharply for the first 30 epochs and then plateaus at around 0.95 for the remainder of the training. The validation loss follows the same trend as the validation Dice, showing a sharp drop early in the training before plateauing at a low stable value with some minor fluctuations, as expected.

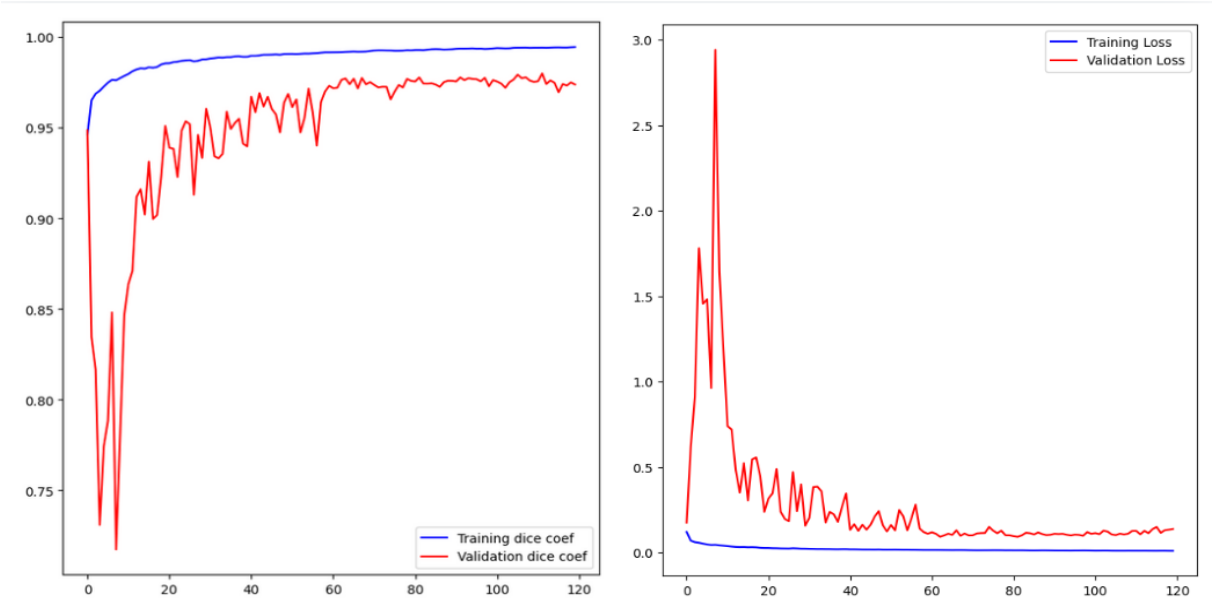

Figure 5. Dice coef and loss of training and validation over 120 epochs

The difference in training and validation curves is a normal occurrence when training high-capacity 3D convolutional architectures on multimodal MRI data. The plateau in the validation Dice and the stabilization of the validation loss indicate that the model is achieving good generalization and not undergoing serious overfitting. The model also achieves promising results on the BraTS2020 test dataset, as outlined in tables 1 and 2, which is another indication of robustness of the proposed model.

5.4. Qualitative Results

The 3D ASPP-ResUNet model yields qualitative results, presented in figure 6, for segmenting brain tumors from test set. The outcomes indicate that the model segments the tumor regions effectively, which is very coherent with the truth.

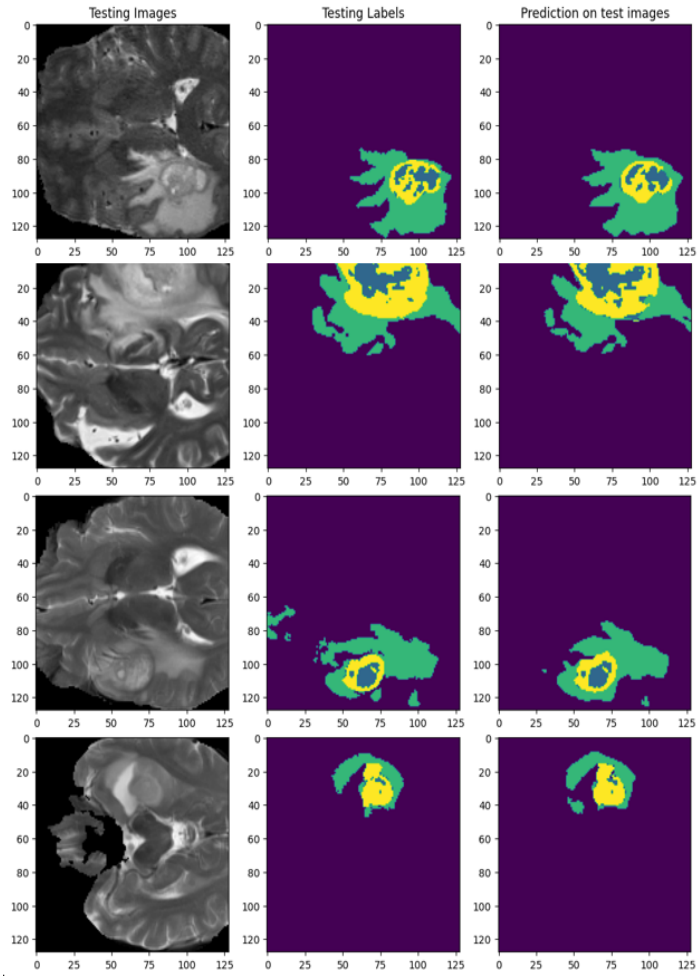

Figure 6. Qualitative results obtained on samples from the test set.

5.5. Comparison with related works

For demonstrating the efficacy of the 3D ASPP-ResUNet method, we compared it against a number of recent methods for tumors segmentation automatically on the BraTS datasets.The Dice scores for WT, TC and ET segmentation using various techniques are compiled in table 3.

Our model (0.8273) outperforms Belaid et al. (0.8068) but lags behind the rest of the comparison methods in WT segmentation. Our model's competitive score (0.7578) in TC segmentation is lower than Isensee et al. (0.8506) or Jiang et al. (0.8030), but performs better than other methods. Our approach (0.7025) is close to the performance of Zhang et al. (0.702) and higher than Belaid et al. (0.6959) for ET segmentation, although it is still less effective than other approaches.

The proposed model shows increased performance over the ET and TC segmentation regions but does not outperform best in class architectures for WT. This effect may relate to the ASPP module's design for improved aggregation of contextual features across multiple scales, which is advantageous for the outlying, complex or heterogeneous tumor subregions such as Enhancing Tumor (ET) and Tumor Core (TC). Thus these characteristics improve the boundary delineations and local detail but may marginally be adverse to the overall volumetric overlap characteristics, which is why the WT Dice yields a marginally lower optimal result.

Table 3. Comparison of our suggested model with other existing methods.

Reference	Method	Dataset	Dice WT	Dice TC	Dice ET
Fang et al. [21]	Three pathways U-Net	BraTS2018	0.8560	0.7220	0.7260
Chen et al. [22]	S3D-UNET	BraTS2018	0.8360	0.6890	0.7830
Belaid et al. [9]	Efficient embedding network	BraTS2020	0.8068	0.7520	0.6959
Zhang et al. [10]	ME-Net	BraTS2020	0.883	0.739	0.702
Isensee et al. [11]	nnU-Net	BraTS2020	0.8895	0.8506	0.8203
Jiang et al. [16]	SwinBTS	BraTS2020	0.8906	0.8030	0.7736
Proposed Model	**Res-UNet+ ASPP**	**BraTS2020**	**0.8273**	**0.7578**	**0.7025**

6. Conclusion and Future work

In medical imaging, the procedure of segmenting brain tumors is highly important. Diagnosing brain tumors early can help improve treatment options and outcome of patient survival. Using the BraTS 2020 DataSet as input, this research proposes a deep learning approach built upon the classical Res-UNet architecture for segmenting brain tumors in 3D multimodal MRI scans. The capacity of the model to delineate global and local contextual features was improved through the effective multi-scale feature extraction technique of ASPP integration. Both experimental evaluation and statistical evaluation showed a statistically significant improvement in Whole Tumor (WT) while still being competitive in performance for Enhancing Tumor (ET) and Tumor Core (TC). The model showed good reliable performance and a balanced performance compared to the modern best practices with low computational expensive performance.

Future work will encompass the integration of enhanced forms of regularization into the approach, particularly dropout, weight decay and advanced data augmentation techniques to ensure stronger generalization of the model. The application of this framework in an exploratory capacity to other complex areas of medical imaging such as lung and breast tumour segmentation show both a further as well as fruitful avenue of future work.

Code and Data Availability

The trained model weights and the source code will be accessible to the general public on GitHub upon publication, to enable reproducibility. The dataset utilized in our work is the BraTS 2020 dataset, which can be accessed on Kaggle.

REFERENCES

1. Raza R, Bajwa UI, Mehmood Y, Anwar MW, Jamal MH. *dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI* Biomed. Signal Process. Control, 2022.
2. Maji D, Sigedar P, Singh M. *Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors.*, Biomed Signal Process Control. 2022.
3. Long J, Shelhamer E, Darrell T. *Fully Convolutional Networks for Semantic Segmentation.* Proc IEEE Conf Comput Vis Pattern Recognit (CVPR). 2015.
4. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain Tumor Segmentation with Deep Neural Networks. Med Image Anal. 2017. *Brain Tumor Segmentation with Deep Neural Networks.* Med Image Anal. 2017.
5. Ranjbarzadeh R, Kasgari AB, Ghoushchi SJ, Anari S, Naseri M, Bendechache M. *Brain Tumor Segmentation based on deep learning and an attention mechanism using mri multi-modalities brain images.* Sci Rep. 2021.
6. Yu F, Koltun V. *Multi-scale context aggregation by dilated convolutions.* Proc Int Conf Learn Represent (ICLR). 2016.
7. Yang T, Zhou Y, Li L, Zhu C. *DCU-Net : Multi-scale U-Net for brain tumor segmentation.* J X-Ray Sci Technol. 2020.

8. Cahall Daniel E, Rasool G, Bouaynaya NC, Fathallah-Shaykh HM. *Dilated Inception U-Net (DIU-Net) for Brain Tumor Segmentation.* arXiv preprint. 2021.
9. Messaoudi H, Belaid A, Allaoui ML, Zetout A, Allili MS, Tliba S, et al. *Efficient embedding network for 3D brain tumor segmentation.* arXiv preprint. 2020.
10. Zhang W, Yang G, Huang H, Yang W, Xu X, Liu Y, et al. *ME-Net : Multi-encoder net framework for brain tumor segmentation.* Int J Imaging Syst Technol. 2021.
11. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.* Nat Methods. 2021; 18(2):203-11.
12. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. *TransUNet: Transformers make strong encoders for medical image segmentation.* arXiv preprint. 2021.
13. Wang W, Chen C, Ding M, Yu H, Zha S, Li J. *TransBTS: Multimodal brain tumor segmentation using transformer.* Med Image Anal. 2021; 71:102036.
14. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. *UNETR: Transformers for 3D medical image segmentation.* Proc IEEE Winter Conf Appl Comput Vis (WACV). 2022; 574-84.
15. Tang Y, Yang D, Li W, Roth HR, Landman BA, Xu D et al. *Swin UNETR: Swin Transformers for semantic segmentation of brain tumors in MRI images.* Proc Int MICCAI Brainlesion Workshop. 2022; 272-84.
16. Jiang Y, Chen H, Liu Y, et al. *SwinBTS: A method for 3D multimodal brain tumor segmentation using Swin Transformer.* Front Oncol. 2022; 12:810030.
17. Kaggle. BRaTS2020 training data. Available from: https://www.kaggle.com/datasets/awsaf49/brats2020-training-data.
18. Chen L-C, Papandreou G, Schroff F, Adam H. *Rethinking Atrous Convolution for Semantic Image Segmentation.*, arXiv preprint. 2017.
19. Qiu X. *U-Net-ASPP : U-Net based on atrous spatial pyramid pooling model for medical image segmentation in COVID-19*, J Appl Sci Eng. 2022.
20. Menze BH, et al. *The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).* IEEE Trans Med Imaging. 2015; 34(10):1993–2024.
21. He H, and Fang L. *Three pathways U-Net for brain tumor segmentation.*, In: MICCAI BraTS Challenge Proceedings. Granada, Spain; 2018. p. 119–26.
22. Chen W, Liu B, Peng S, Sun J. *S3D-UNET: Separable 3D U-Net for brain tumor segmentation*, In: BrainLes 2018, MICCAI Workshop, Granada, Spain; 2019. (Lecture Notes in Computer Science; vol. 11384). Springer.