ISSN: 1609-1620
VOLUME 27 ISSUE 06 2025
PAHTEI
REFEREED & REVIEWED JOURNAL
Pp. 95-101 DOİ: https://doi.org/10.32010/CIZM1703

# CASCADED UNET FOR GLIOMA SEGMENTATION

**Elvin Aliyev[1], Latafat Gardashova[2]**
[1,2] Azerbaijan State Oil and Industry University
[1,2] Department of Computer Engineering
[1] Master student, aliyeve669@gmail.com
[2] Vice-rector, professor, l.qardashova@asoiu.edu.az

## ABSTRACT

Magnetic Resonance Imaging (MRI) plays a pivotal role in the diagnosis and treatment of brain tumors, making its accurate assessment critically important. However, the inherent three-dimensional nature of MRI presents several challenges, leading to the common practice of conducting analyses on two-dimensional projections. While this simplification reduces complexity, it also introduces potential biases. Conversely, the more time-intensive three-dimensional evaluations, such as segmentation, can yield precise estimates of various spatial characteristics, enhancing our understanding of disease progression. Recent research focusing on segmentation tasks has demonstrated that Deep Learning techniques outperform traditional computer vision algorithms, although the problem remains complex. In this paper, we introduce a deep cascaded approach for the automatic segmentation of brain tumors. Our method, akin to contemporary object detection techniques, leverages neural networks and includes modifications to the 3D UNet architecture and augmentation strategies to effectively process multimodal MRI data. Additionally, we present a method to improve segmentation quality by incorporating contextual information from models of the same architecture operating on downscaled datasets. We assess our proposed approach using the BraTS 2018 dataset and provide a discussion of the results.
**Keywords**: segmentation, BraTS, UNet, cascaded UNet, deep learning.

## Introduction

Multimodal magnetic resonance imaging (MRI) is a powerful technique extensively used in human brain research. Although it has diverse applications, its main roles lie in disease diagnosis and treatment planning. Accurate interpretation of MRI data is crucial at every stage of this process. Since MRI scans consist of multiple three-dimensional arrays, manual analysis is inherently complex, time-intensive, and demands specialized expertise. Limited availability of professional time and resources can result in less-than-optimal outcomes. Clinicians often review MRI images in two-dimensional slices or projections, which limits the amount of information assessed and can introduce bias in their interpretations. In contrast, precise segmentation and three-dimensional reconstruction provide deeper understanding of disease progression and enhance treatment planning. Nevertheless, these advanced approaches are rarely used due to the significant time required for manual annotation. To address this challenge, the Brain Tumor Segmentation (BraTS) challenge [1, 10] was established as an annual competition that offers a standardized platform for developing and benchmarking cutting-edge segmentation methods. Participants are tasked with producing segmentation maps for different glioma subregions, including the enhancing tumor (ET), tumor core (TC), and whole tumor (WT). The training dataset [2, 3] includes 210 MRI scans of high-grade gliomas and 75 scans of low-grade gliomas, each carefully labeled by expert annotators. The test data is split into two parts: a validation set, which is used for continuous evaluation throughout the challenge, and a final test set, reserved for the conclusive assessment of method performance. Evaluation of the segmentation approaches relies on metrics such as the Dice coefficient, Sensitivity, Specificity, and Hausdorff distance.

$$Y_i = F(X_i, Y_{i-1}, Y_{i-2}, W_i)$$

PAHTEI
REFEREED & REVIEWED JOURNAL
ISSN: 1609-1620
VOLUME 27 ISSUE 06 2025

**Objective**

The main objective of the study is to use a deep cascading approach for automatic segmentation of brain tumors, which is similar to modern object detection methods and uses neural networks. It includes modifications to the 3D UNet architecture and a supplementation strategy for effective processing of multimodal MRI data.

**Methods**

I propose a method for brain tumor segmentation that relies on neural networks. This approach involves a series of classifiers $Ci$ , each having an identical structure $F$, where each classifier progressively refines the segmentation output generated by the previous one. Although these classifiers share the same architectural design, each classifier $Ci$ has its own unique set of parameters $Wi$ that are independently trained. The output at iteration $i$, represented as $Yi$, is computed by applying the function

$Y = F(X_i, Y_{i-1}, Y_{i-2}, W_i)$,

with $Xi$ being the input at iteration $i$. This framework is illustrated in Figure 1. Every classifier $Ci$ is based on a modified UNet architecture, specifically adapted for glioma segmentation. Unlike the original UNet model [11] and its 3D variant [5], our design incorporates multiple distinct encoders to handle various input modalities and includes a specialized fusion mechanism to combine their outputs effectively.

The conventional UNet architecture [11], extended for volumetric data processing [5], consists primarily of an encoder and a decoder. The encoder extracts hierarchical features and contextual information across different scales, while the decoder generates segmentation maps by integrating these features with contextual cues obtained during encoding. To facilitate better feature learning at deeper layers, skip connections link corresponding stages of the encoder and decoder, allowing comparison of features at identical spatial resolutions but with differing receptive fields. Although this architecture is well-suited for processing multimodal MRI data, it treats each modality identically by merging all inputs at the initial stage.

**Multiple encoders Unet.** Our proposed approach addresses this limitation by learning separate feature representations for each modality, then combining them at a later stage in the network.

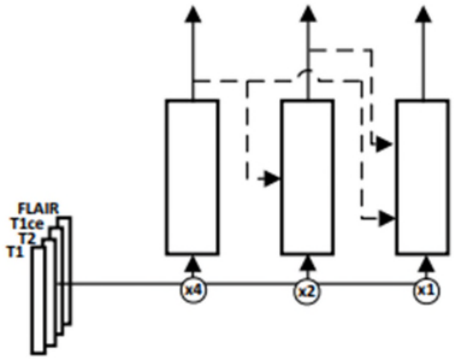


**Figure 1**. Illustrating the method used in this study

The abbreviations T1, T2, T1ce, and FLAIR represent different MRI input modalities. The labels x4 and x2 denote the downsampling rates applied to the network inputs. Dashed arrows show the links between the classifiers $Ci$ , which are depicted as simple blocks.

This is achieved through grouped convolutions within the encoder, where the number of groups matches the count of input modalities—allowing each modality to be processed separately. The feature maps generated by these modality-specific encoders are combined by taking the element-wise maximum across them. To preserve the spatial dimensions after this fusion, a point-wise convolution

ISSN: 1609-1620
VOLUME 27 ISSUE 06 2025
PAHTEI
REFEREED & REVIEWED JOURNAL

(1x1x1) is applied. Following the design principles of the original UNet, our model doubles the number of feature channels with each downsampling stage and halves them during each upsampling step, using ReLU activation functions after every convolutional layer.

The proposed cascaded UNet architecture, illustrated in Figure 1, is composed of three identical blocks based on the UNet model, each equipped with its own loss function at the output. Each block receives the downsampled segmentation output from its predecessor and produces a segmentation map of matching size. Mirroring the DeepMedic architecture [9], this design processes the input at multiple scales simultaneously, retaining features specific to each scale. Additionally, the feature map from just before the final convolution in each block is concatenated with the corresponding feature map from the block at the next finer scale, enabling the exchange of contextual information across networks handling different resolutions.

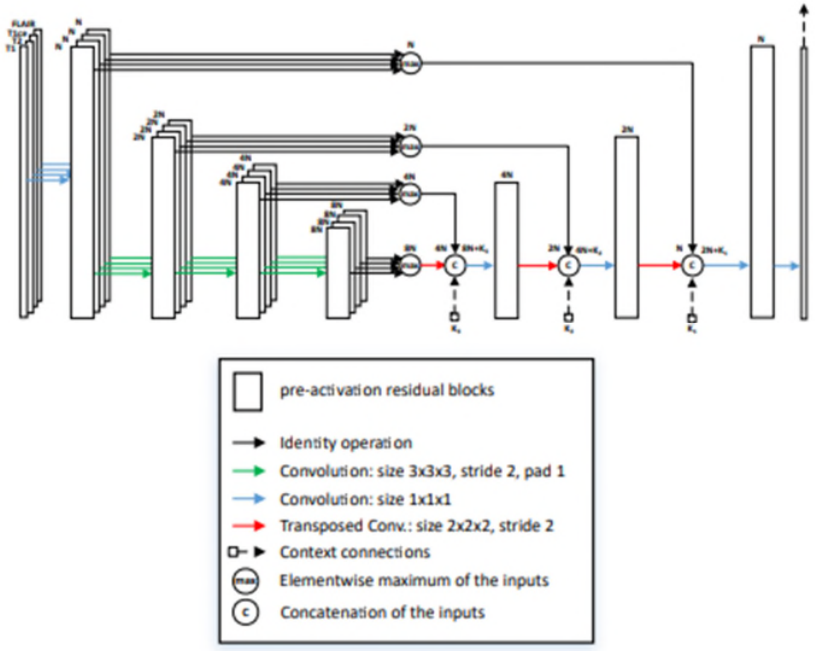


**Figure 2.** Architecture of multiple encoders UNet

T1, T2, T1CE, FLAIR stand for input modalities. N is a base number of filters, K is a number of filters in context feature map obtained from lower scale models.

Within the UNet framework, the decoder's output at a given scale $i$ is influenced by two sources: the encoder's output at the corresponding scale, accessed through skip connections, and the decoder's output from the preceding scale:

$$d_i^t = f(e_i^t, d_{i-1}^t)$$

Here, $d_i^t$ represents the decoder output at scale $e_i^t$ denotes the encoder output at the same scale, and t refers to the network index. By expanding the initial convolution operation of the function $f$, we obtain:

$$d_i^t = g\left(W_{i,e}^t e_i^t + W_{i,d}^t d_{i-1}^t\right)$$

Here, $W$ stands for the learnable parameters. Our approach improves upon this by integrating contextual information from networks operating at lower scales, achieved by concatenating the output of the corresponding network $yt$ (depicted with dotted arrows in Figure 2). This leads to the following updated equation:

$$d_i^t = g\left(W_{i,e}^t e_i^t + W_{i,d}^t d_{i-1}^t + W_{i,y}^t y^{t-i}\right)$$

PAHTEI
REFEREED & REVIEWED JOURNAL
ISSN: 1609-1620
VOLUME 27 ISSUE 06 2025

This technique combines several networks working at various scales, promoting progressive refinement of outputs from earlier stages.

**Preprocessing & Data augmentation.** The preprocessing strategy outlined in [7] proved highly effective. Following the same procedure, we apply z-score normalization specifically to the non-zero voxels corresponding to brain tissue. Afterward, to minimize noise and eliminate outliers, all voxel values are clipped to fall within the interval from -5 to 5.

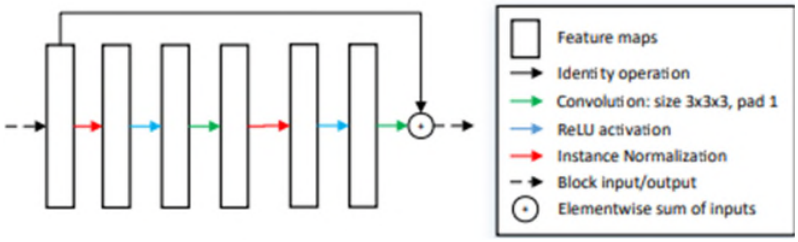


**Figure 3.** Design of the residual block

In the final preprocessing step, brain voxel intensities are scaled to lie within the range [0, 10], while background voxels are set to zero. To augment the training dataset offline, we expand the number of samples by applying b-spline deformations to the original images, utilizing the ITK library [8]. During training, input images are randomly flipped along the sagittal axis, and some input modalities are randomly "muted" with a fixed probability to prevent the network from relying solely on a single modality. To further encourage the model to use all modalities, we introduce Gaussian noise to the input channels with a 10% chance per channel, which results in a 34% probability that at least one of the four modalities is muted during training.

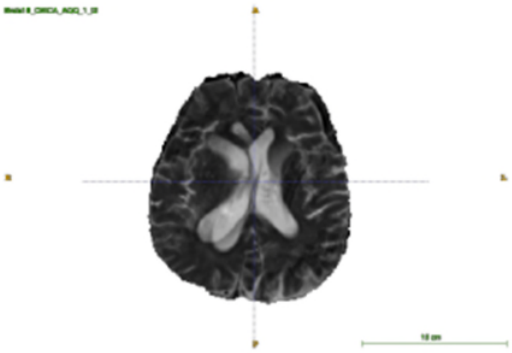

**Figure 4.** Illustration of registration artifacts observed within the training dataset [12]

**Training.** Training is performed on brain regions resampled to a fixed size of 128×128×128 voxels. We opt for down - sampled inputs to preserve important contextual cues, which we consider crucial for achieving reliable segmentation across multimodal MRI scans obtained from various institutions and scanner types. The Mean Dice loss function is employed during training, where $g$ represents the ground truth and $p$ the network's prediction. Optimization is carried out using stochastic gradient descent with an initial learning rate of 0.1, which decays exponentially by 1% every epoch. Additionally, a weight decay of 0.9 is applied, and training is conducted with mini-batches containing 4 samples each. The Mean Dice loss formula is defined as follows:

$$L_m lan_- = 1 - \frac{1}{|c|} \sum_{c \epsilon C} \frac{\Sigma_{i_c} p_c^i q_c^i}{\Sigma_{i_c} p_c^i + q_c^i}$$

Here, $C$ represents the collection of classes. The convolutional neural network was built using the MX - Net library [4] and trained across four GTX 1080TI GPUs, employing a batch size of 4 to facilitate data parallelism. The model was trained over the course of 500 epochs.

**Result**

We report the evaluation outcomes obtained from the online validation platform hosted by the challenge organizers. To prevent the model from over-relying on any single modality, we introduce channel-out augmentation, which randomly replaces input modalities with Gaussian noise, supplementing the standard augmentations such as mirroring and elastic deformations. Results are presented both with this augmentation enabled (see Table 2) and disabled (see Table 1). The challenge validation set [2, 3] includes 66 MRI scans acquired from various scanners and institutions. Validation results on this dataset are summarized in Table 3. A comparison of glioma segmentation performance without channel-out augmentation is provided in Table 1, reporting Dice scores for different tumor regions: WT (whole tumor), ET (enhancing tumor), and TC (tumor core).

| Method | WT | ET | TC |
|---|---|---|---|
| UNet | 0.901 | 0.767 | 0.797 |
| ME UNet | 0.904 | 0.763 | 0.823 |
| C ME UNet | 0.906 | 0.772 | 0.836 |

**Conclusion**

This study introduces an automated segmentation technique aimed at addressing two key challenges in brain tumor segmentation from multimodal imaging: managing complex and diverse input data, and reducing excessive confidence in classifier predictions. Table 2 presents the glioma segmentation outcomes achieved with channel-out augmentation, reporting Dice scores for Whole Tumor (WT), Enhancing Tumor (ET), and Tumor Core (TC).

| Method | WT | ET | TC |
|---|---|---|---|
| UNet | 0.901 | 0.779 | 0.837 |
| ME UNet | 0.907 | 0.784 | 0.827 |
| C ME UNet | 0.908 | 0.784 | 0.844 |

**Table 3.** Evaluation results of the suggested approach on the BraTS-2018 validation_set, showing Dice scores

| | WT | ET | TC |
|---|---|---|---|
| Mean | 0.908 | 0.784 | 0.844 |
| StdDev | 0.065 | 0.237 | 0.161 |
| Median | 0.926 | 0.858 | 0.906 |
| 25quantile | 0.9 | 0.805 | 0.791 |
| 75quantile | 0.943 | 0.897 | 0.947 |

To address the challenge posed by heterogeneous input data, we introduced the use of multiple encoders, with each input modality independently producing its own set of feature maps. Alongside this, we developed a technique to effectively combine these modality-specific feature representations. We also explored the impact of channel-out augmentation on the model's performance, showing that this robust strategy enhances the model's output. By randomly masking individual input modalities during training, channel-out augmentation encourages the model to rely on all available modalities rather than over-depending on a single one. This leads to increased robustness against noisy or corrupted data present in both training and validation sets.

Furthermore, we designed an efficient framework for integrating several models that operate at different spatial resolutions, forming a cascade of classifiers. Each subsequent classifier refines the

segmentation output of the previous stage at its specific resolution, enabling iterative improvement while requiring fewer parameters compared to conventional deep networks. As part of our participation in the BraTS-2018 challenge [10, 1], we implemented and assessed our approach through the challenge's online validation platform. Our method achieved impressive results, with high average scores and outstanding median values, including mean Dice coefficients of 0.908, 0.784, and 0.844 for the Whole Tumor, Enhancing Tumor, and Tumor Core regions respectively on the validation dataset.

**REFERENCES**

1. Bakas, S., Akbari, H., Sotiras, A. et al. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Sci Data 4, 170117 (Sep 2017). https://doi.org/10.1038/sdata.2017.117, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5685212/, 28872634[pmid]
2. Bakas, S., Akbari, H., Sotiras, A. et al. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The Cancer Imaging Archive (2017). https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q
3. Bakas, S., Akbari, H., Sotiras, A. et al. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive (2017). https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF
4. Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., Zhang, Z.: Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. CoRR abs/1512.01274 (2015), http://arxiv.org/abs/1512.01274
5. Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O. 3d u-¨net: Learning dense volumetric segmentation from sparse annotation. CoRR abs/1606.06650 (2016), http://arxiv.org/abs/1606.06650
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR abs/1512.03385 (2015), http://arxiv.org/abs/1512.03385
7. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge. CoRR abs/1802.10508 (2018), http:// arxiv.org/abs/1802.10508
8. Johnson, H.J., McCormick, M., Ib´a˜nez, L., Consortium, T.I.S.: The ITK Software Guide. Kitware, Inc., third edn. (2013), http://www.itk.org/ ItkSoftwareGuide.pdf, In press
9. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. CoRR abs/1603.05959 (2016), http://arxiv.org/abs/1603.05959
10. Menze, B.H., Jakab, A., Bauer, S. et al. The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging 34(10), 1993–2024 (Oct 2015). https://doi.org/10.1109/TMI.2014.2377694
11. Ronneberger, O., Fischer, P., Brox, T. U-net: Convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015), http:// arxiv.org/abs/1505.04597
12. Yushkevich, P.A., Piven, J., Cody Hazlett, H. et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage 31(3), 1116–1128 (2006)

## GLIOMA SEQMENTASIYASI ÜÇÜN KASKAD UNET

**Elvin Əliyev[1], Lətafət Qardaşova[2]**
[1,2] Azərbaycan Dövlət Neft və Sənaye Universiteti
[1,2] Kompüter mühəndisliyi kafedrası
[1] Magistr tələbəsi, aliyeve669@gmail.com
[2] Prorektor, professor, l.qardashova@asoiu.edu.az

ISSN: 1609-1620
VOLUME 27 ISSUE 06 2025
PAHTEI
REFEREED & REVIEWED JOURNAL

**XÜLASƏ**

Maqnetik Rezonans Görüntüləmə (MRI), beyindəki şişlərin diaqnostikası və müalicəsində mühüm rol oynayır, buna görə də onun dəqiq qiymətləndirilməsi çox vacibdir. Lakin, MRI-nin təbii üçölçülü təbiəti bir sıra çətinliklər yaradır və buna görə də analizlərin ikiölçülü proyeksiyalar üzərində aparılması geniş yayılmış bir təcrübədir. Bu sadələşdirmə mürəkkəbliyi azaltsa da, eyni zamanda potensial təhriflərə səbəb ola bilər. Digər tərəfdən, daha çox vaxt tələb edən üçölçülü qiymətləndirmələr, məsələn, seqmentasiya, müxtəlif məkan xüsusiyyətlərinin dəqiq təxminlərini təmin edərək xəstəliyin irəliləməsini daha yaxşı başa düşməyimizə kömək edir. Son tədqiqatlar seqmentasiya tapşırıqlarına yönəlmiş və göstərib ki, Dərin Öyrənmə texnikaları ənənəvi kompüter görmə alqoritmlərini üstələyir, baxmayaraq ki, problem hələ də mürəkkəbdir. Bu məqalədə biz, beyin şişlərinin avtomatik seqmentasiyası üçün dərin kaskadlı yanaşma təqdim edirik. Yöntəmimiz, müasir obyekt aşkarlama texnikalarına bənzəyərək neyron şəbəkələrindən istifadə edir və 3D UNet arxitekturasına və artırma strategiyalarına dəyişikliklər daxil edərək multimodal MRI məlumatlarını effektiv şəkildə işləyir. Əlavə olaraq, biz, eyni arxitektura ilə işləyən və kiçildilmiş verilənlər dəstləri üzərində işləyən modellərdən kontekstual məlumat daxil edərək seqmentasiya keyfiyyətini yaxşılaşdırmaq üçün bir metod təqdim edirik. Təklif etdiyimiz yanaşmanı BraTS 2018 verilənlər dəstində qiymətləndiririk və nəticələrin müzakirəsini təqdim edirik.
**Açar sözlər:** seqmentləşdirmə, BraTS, UNet, kaskadlı UNet, dərin öyrənmə.

# КАСКАДНЫЙ UNET ДЛЯ СЕГМЕНТАЦИИ ГЛИОМЫ

**Эльвин Алиев[1], Латафат Гардашова[2]**
Азербайджанский государственный университет нефти и промышленности
Кафедра «Компьютерная инженерия»
Проректор, профессор, l.qardashova@asoiu.edu.az
Студент-магистр, aliyeve669@gmail.com

**РЕЗЮМЕ**

Магнитно-резонансная томография (МРТ) играет ключевую роль в диагностике и лечении опухолей головного мозга, что делает ее точную оценку критически важной. Однако трехмерная природа МРТ создает ряд проблем, что приводит к распространенной практике проведения анализов на двумерных проекциях. Хотя это упрощение снижает сложность, оно также вносит потенциальные смещения. И наоборот, более трудоемкие трехмерные оценки, такие как сегментация, могут дать точные оценки различных пространственных характеристик, улучшая наше понимание прогрессирования заболевания. Недавние исследования, сосредоточенные на задачах сегментации, продемонстрировали, что методы глубокого обучения превосходят традиционные алгоритмы компьютерного зрения, хотя проблема остается сложной. В этой статье мы представляем глубокий каскадный подход для автоматической сегментации опухолей головного мозга. Наш метод, родственный современным методам обнаружения объектов, использует нейронные сети и включает модификации архитектуры 3D UNet и стратегии дополнения для эффективной обработки мультимодальных данных МРТ. Кроме того, мы представляем метод улучшения качества сегментации путем включения контекстной информации из моделей той же архитектуры, работающих на уменьшенных наборах данных. Мы оцениваем наш предлагаемый подход с использованием набора данных BraTS 2018 и приводим обсуждение результатов.
**Ключевые слова:** сегментация, BraTS, UNet, каскадный UNet, глубокое обучение.

Publishing history: 30.06.2025
Article received: 22.05.2025
Article accepted: 10.06.2025