3D Glioma Segmentation Using SegFormer3d

Ryhor Karapetsian
Laboratory of Informationand
Computer Technologies
Belorusian State
Medical University
Minsk, Belarus
greg.itlab@gmail.com

Alexander Nedzved
Belarusian State University,
United Institute of Informatics
Problems of the National
Academy of Sciences of Belarus
Minsk, Belarus
ORCID:0000-0001-6367-5900

Aliaksandr Fedulau
Department of Neurology
and Neurosurgery
Belorusian State
Medical University
Minsk, Belarus
fedulov@tut.by

Ivan Kosik
Laboratory of Informationand
Computer Technologies
Belorusian State
Medical University
Minsk, Belarus
Ivankosik91@gmail.com

Vadim Ermakov
Laboratory of Information
and Computer Technologies
Belorusian State
Medical University
Minsk, Belarus
ermakouvv@gmail.com

Siarhei Kezik
Laboratory of Information
and Computer Technologies
Belorusian State
Medical University
Minsk, Belarus
sergeyk8008@gmail.com

***Abstract*. Accurate segmentation of gliomas is essential for diagnosis and treatment planning but is challenged by their infiltrative nature and reliance on gadolinium-based contrast agents, which pose certain health risks. This study evaluates SegFormer3d, a transformer-based model, for efficient 3D glioma segmentation using the BraTS 2020 dataset, focusing on reducing dependency on T1-contrast (T1C) modality. With modifications including LayerNorm removal and a refined segmentation head, SegFormer3d achieved Dice scores of 0.8333 (Enhancing Tumor, ET), 0.8534 (Tumor Core, TC), and 0.9179 (Whole Tumor, WT) with all modalities, but ET performance dropped to 0.6178 without T1C. Weight Standardization and model scaling improved contrast-free results. 3D reconstructions using polygonal meshes and voxel-polygonal hybrid visualization aided surgical planning, though non-contrast ET accuracy requires enhancement. SegFormer3d offers a computationally efficient solution, but further advancements are needed for robust contrast-free segmentation.**

***Keywords:* Glioma segmentation, SegFormer3d, 3D reconstruction, transformer-based model, MRI, contrast-free imaging, deep learning, neurosurgery, medical imaging**

I. INTRODUCTION

Gliomas, encompassing aggressive glioblastomas (GBM/HGG) and less malignant low-grade gliomas (LGG), are the most common primary brain tumors, accounting for approximately 30 % of all brain and central nervous system tumors and 80 % of malignant brain tumors [1]. Their infiltrative nature and indistinct boundaries with healthy tissue pose significant challenges for accurate segmentation, which is critical for diagnosis, surgical planning, and monitoring disease progression [2]. Multi-modal Magnetic Resonance Imaging (MRI), utilizing T1, T1-contrast (T1C), T2, and T2-FLAIR sequences, enhances tumor visualization by capturing distinct tissue characteristics, such as contrast enhancement in the Enhancing Tumor (ET) or fluid attenuation in Peritumoral Edema (ED) [3, 4]. However, reliance on gadolinium-based contrast agents in T1C imaging raises safety concerns, including risks of nephrogenic systemic fibrosis and gadolinium deposition in the brain, liver, and kidneys, particularly in patients with renal impairment [5–7]. Regulatory warnings from the FDA and EMEA since 2017 advocate restricted use of these agents, driving the need for contrast-free segmentation methods [8, 9].

Deep learning has transformed medical image segmentation, with convolutional neural networks (CNNs) like UNet achieving state-of-the-art results but often requiring significant computational resources [10, 11]. Transformer-based models, originally developed for natural language processing, have recently gained traction in medical imaging for their ability to capture long-range dependencies and process complex 3D data efficiently [12, 13]. SegFormer3d, adapted from the SegFormer framework, offers a computationally efficient alternative for 3D glioma segmentation, particularly for contrast-free scenarios [15]. This study investigates SegFormer3d's performance in segmenting ET, ED, and Necrosis/Non-enhancing Tumor (NCR/NET) using the BraTS 2020 dataset, with a focus on reducing T1C dependency. Advanced data augmentation and optimization techniques are employed to address challenges like tumor heterogeneity and data scarcity, aiming to enhance clinical applicability in resource-constrained settings [16–19].

II. SEGFORMER3D ARCHITECTURE

SegFormer3d, adapted from the transformer-based SegFormer framework [15], is designed for efficient 3D medical image segmentation, addressing the computational limitations of UNet-based architectures that process full-resolution images and demand high-performance hardware [10, 20]. Unlike traditional CNNs, SegFormer3d leverages a hierarchical transformer encoder that reduces spatial resolution by a factor of 4 in each dimension through an initial convolution, significantly lowering memory usage and computational complexity, making it suitable for resource-constrained medical environments [21]. The input MRI volume, typically of size 160×192×160 after preprocessing, is divided into N vectorized patches, each representing a local region of the 3D image. These patches are linearly transformed into three vectors: Query (Q), which identifies relationships with other patches; Key (K), containing information about all patches for matching with Q; and Value (V), indicating the presence or absence of correspondences [15].

The core innovation of SegFormer3d lies in its Efficient Self-Attention mechanism, which reduces computational overhead compared to standard transformer attention [12]. Before computing K and V, a resolution-reducing convolution (1×1×1) is applied, decreasing the feature map size early in the model, which minimizes the number of operations in the attention computation. This is particularly critical for 3D medical imaging, where large input volumes can lead to prohibitive memory demands [13] as on Fig. 1 [14]. The encoder operates in multiple stages, each progressively reducing spatial dimensions while increasing channel depth (e.g., from 32 to 256 channels across stages), capturing both local and global features effectively [15].

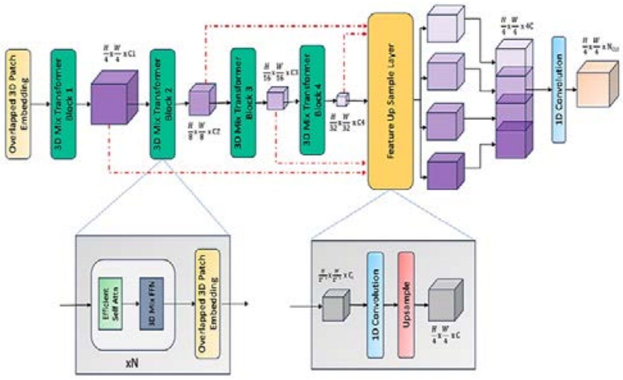

Fig. 1. Model of SegFormer3d

The Mix-FFN (Feed-Forward Network) block further distinguishes SegFormer3d from standard transformers. Unlike traditional FFNs, Mix-FFN incorporates a 3×3×3 convolution without inter-channel connections before the activation function, enhancing spatial feature integration while maintaining computational efficiency [15]. This design balances the transformer's global context awareness with CNN-like local feature extraction, crucial for accurate segmentation of heterogeneous glioma regions [18].

The decoder is notably lightweight, contrasting with the complex decoders in UNet-based models [10]. It unifies multi-scale features from the encoder by aligning them to a common channel count and resolution through linear layers and upsampling, followed by a final fusion step to produce the segmentation mask. To address limitations in the base SegFormer3d, two key modifications were introduced: (1) removal of redundant LayerNorm after the initial convolution in each encoder stage to accelerate convergence by reducing normalization overhead, and (2) replacement of the segmentation head with a sequence of a linear layer to reduce channel count, trilinear interpolation for 4× upsampling, and a 3×3×3 convolution to refine small mask predictions, improving accuracy for fine details like the ET region [14, 22]. These changes enhance the model's ability to handle small, irregularly shaped tumor regions while maintaining efficiency.

III. TRAINING AND DATA

The model was trained on the BraTS 2020 dataset [23], comprising 1251 annotated MRI scans from 19 institutions, standardized to 1 mm³ resolution with skull-stripping. Each scan includes T1, T1C, T2, and T2-FLAIR modalities with masks for ET, ED, and NCR/NET. The dataset was split into training (every fifth scan), validation, and a test set of 126 scans from Minsk City Clinical Hospital and RNPC Neurology and Neurosurgery. Dynamic data augmentation, including 90º and ±15º rotations, flipping, and brightness adjustments, addressed transformer data requirements [16]. Training used 128×128×128 crops, while validation used 160×192×160 with zero-padding, with intensities normalized to [0,1]. The loss function combined binary cross-entropy and Dice loss, optimized with ScheduleFree AdamW (learning rate 0.001, weight decay 0.01, 2000 warmup steps) [24] over 100–200 epochs with a batch size of 3. BatchNorm was replaced with GroupNorm (4 groups).

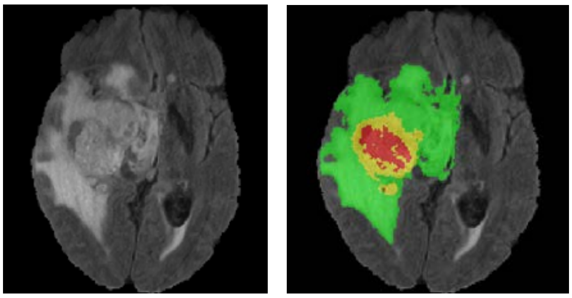

Fig. 2. Sorce image and labeling by expert

IV. SEGMENTATION RESULTS ANALYSIS

A. *Segmentation with All Modalities*

Training with all modalities improved Dice scores after modifications (Table I). Removing redundant LayerNorm and adding a new segmentation head reduced the error from 0.1833 to 0.1617, with optimal Dice scores for ET (0.8333), Tumor Core (TC) (0.8534), and Whole Tumor (WT) (0.9179).

TABLE I. DICE SCORES FOR SEGMENTATION WITH ALL MODALITIES

Modification	**Scores for Segmentation**			
	Error	***ET***	***TC***	***WT***
None	0.1833	0.8035	0.8399	0.9044
-LayerNorm	0.1705	0.8130	0.8534	0.9174
-LayerNorm, +New Head	0.1617	0.8333	0.8511	0.9179

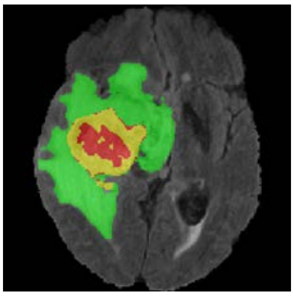

Fig. 3. Segmentation by simple model. Green color indicates peritumoral edema (ED), yellow color indicates area of contrast accumulation (ET), and red color indicates area of necrosis (NCR)

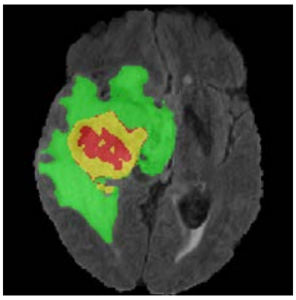

Fig. 4. Segmentation by LayerNorm model

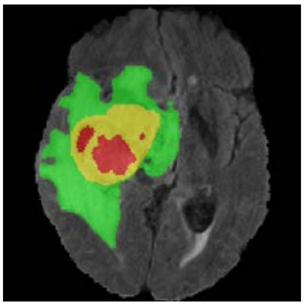

Fig. 5. Segmentation by -LayerNorm, +New Head model

B. *Segmentation without T1C*

Excluding T1C reduced performance, particularly for ET (Dice dropped from 0.8333 to 0.4929). Weight Standardization (WS) with affine parameters and initialization improved results, achieving an error of 0.2971 and ET Dice of 0.5715 (Table II).

TABLE II. DICE SCORES WITHOUT T1C MODALITY

Modification	**Scores for Segmentation**			
	Error	***ET***	***TC***	***WT***
None	0.3363	0.4929	0.7137	0.9115
+WS, +LayerNorm	0.3319	0.5324	0.7240	0.8887
+WS affine, +LayerNorm	0.3162	0.5337	0.7407	0.9039
+WS affine with init	0.2971	0.5715	0.7653	0.9102

Scaling the model (channels increased to {64, 128, 256, 512}) and extending training to 200 epochs improved ET Dice to 0.6178 (Table III). Removing attention blocks degraded performance, confirming their importance.

TABLE III. DICE SCORES WITH SCALING, WITHOUT T1C

Modification	**Scores for Segmentation**			
	Error	***ET***	***TC***	***WT***
None	0.2971	0.5715	0.7653	0.9102
+200 epochs	0.2808	0.5958	0.7843	0.9158
+Larger model	0.2859	0.5851	0.7744	0.9168
+Larger model, +200 epochs	0.2759	0.6178	0.7815	0.9173

C. *3D Reconstruction*

SegFormer3d enabled high-quality 3D reconstruction of glioma components, producing polygonal mesh models to visualize the tumor's structural regions: Necrosis (NCR) in red, Enhancing Tumor (ET) in yellow, and Peritumoral Edema (ED) in green. These meshes were generated from segmentation masks predicted by the model, with each region assigned distinct opacity levels – 80 % for NCR (least transparent), 40 % for ET, and 15 % for ED (most transparent) – to enable simultaneous visualization of overlapping internal structures [25]. This transparency gradient is critical for clinical applications, as it allows neurosurgeons to assess the spatial relationships between tumor components, such as the necrotic core's position relative to the enhancing tumor and edema, which are often obscured in 2D slices [26].

Reconstructions using all modalities (T1, T1C, T2, T2-FLAIR) closely matched expert annotations, accurately capturing the irregular boundaries of ET and the diffuse nature of ED. However, reconstructions without T1C showed satisfactory but less precise results, particularly for ET, where the Dice score dropped to 0.6178 compared to 0.8333 with T1C, reflecting the challenge of identifying contrast-enhanced regions without gadolinium [15]. The yellow ET region appeared less distinct in non-contrast reconstructions due to overlap with the green ED layer, whose transparency slightly altered perceived colors, as noted in the original document.

Voxel-based 3D reconstructions (Fig. 6) were also performed, leveraging the volumetric nature of MRI data to visualize the entire brain with the tumor embedded [26]. These reconstructions effectively highlighted ET in maximum intensity projection mode but struggled to differentiate ED and NCR without color coding, as their intensity profiles are similar in non-contrast modalities [4]. To address this, a combined voxel-polygonal approach was implemented, where tumor-related voxels were nullified (set to zero) to prevent overlap with polygonal meshes, enhancing visibility of internal structures like the necrotic core [28]. This hybrid visualization provided a comprehensive view of the tumor's spatial organization within the brain, aiding surgical planning by clearly delineating critical regions [26].

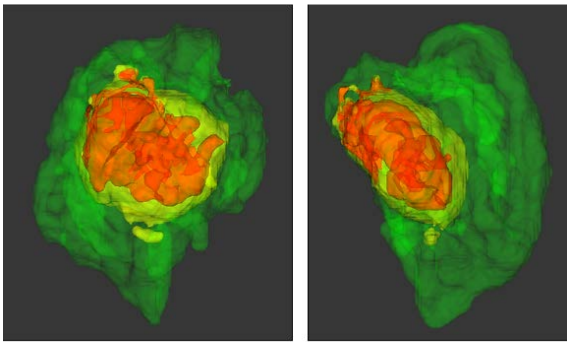

Fig. 6. Two racurs of a three-dimensional reconstruction of a brain tumor with color coding of diagnostically significant components using the T1C mode: red color is necrosis zone, yellow (marked with an arrow) is area of contrast agent accumulation, green is area of perifocal edema

Comparisons with expert-annotated reconstructions confirmed SegFormer3d's ability to produce clinically relevant visualizations, though non-contrast reconstructions require further refinement to match expert precision for ET. These visualizations underscore the model's potential for reducing contrast agent use while supporting detailed 3D analysis in clinical settings [29].

V. CONCLUSION

The SegFormer3d model demonstrates significant potential as an efficient and effective framework for 3D glioma segmentation, achieving high Dice scores of 0.9179 for the Whole Tumor (WT) region when utilizing all MRI modalities (T1, T1C, T2, T2-FLAIR), with robust performance for Tumor Core (TC) (0.8534) and Enhancing Tumor (ET) (0.8333) regions. Architectural modifications, such as the removal of redundant LayerNorm and the introduction of a refined segmentation head with trilinear interpolation and 3×3×3 convolution, significantly improved segmentation accuracy, particularly for small and irregular ET regions, reducing the error to 0.1617. The model's lightweight design and reduced computational requirements, achieved through resolution-reducing convolutions and Efficient Self-Attention, make it a viable solution for resource-constrained medical settings, unlike traditional UNet-based models that demand high-performance hardware.

However, contrast-free segmentation, particularly for the ET region, remains a significant challenge, with the Dice score dropping to 0.6178 when excluding T1C, despite improvements from Weight Standardization (WS) with affine parameters and model scaling to 200 epochs with increased channel depths ({64, 128, 256, 512}). This performance gap highlights the critical role of T1C in delineating contrast-enhanced regions and underscores the need for further advancements to achieve clinically acceptable accuracy without gadolinium-based agents, given their associated health risks. The 3D reconstructions, both polygonal and voxel-based, provide valuable visualization tools for surgical planning, with the hybrid voxel-polygonal approach effectively addressing visibility challenges in non-contrast settings. Nonetheless, the reduced precision in ET reconstruction without T1C indicates limitations in capturing fine tumor boundaries, necessitating improved feature extraction strategies.

REFERENCES

[1] Q. T. Ostrom, et al., "CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2013–2017", Neuro-Oncology, vol. 22(Supp._1), 2020, pp. iv1–iv96.

[2] N. Sanai, M. S. Berger, "Glioma extent of resection and its impact on patient outcome", Neurosurgery, vol. 89(3), 2021, pp. 345–353.

[3] B. R. J. van Dijken, et al., "Diagnostic accuracy of magnetic resonance imaging techniques for treatment response evaluation in patients with high-grade glioma", European Radiology, vol. 30(10), 2020, pp. 5478–5489.

[4] V. Sawlani, et al., "Multiparametric MRI: Practical approach and pictorial review of a useful tool in the evaluation of brain tumours", Clinical Radiology, vol. 75(8), 2020, pp. 574–586.

[5] T. Kanda, et al., "Gadolinium deposition in the brain", Magnetic Resonance Imaging, vol. 79, 2021, pp. 117–124.

[6] M. Rogosnitzky, S. Branch, "Gadolinium-based contrast agents: Safety concerns", Biometals, vol. 35(1), 2022, pp. 1–14.

[7] J. Ramalho, et al., "Gadolinium retention and toxicity–An update, Advances in Chronic Kidney Disease, vol. 27(3), 2020, pp. 236–243.

[8] FDA, Gadolinium-based contrast agents: Safety update, FDA Drug Safety Communication, 2021.

[9] EMEA, Gadolinium-containing contrast agents: Restrictions in use, European Medicines Agency, 2020.

[10] F. Isensee, et al., “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation”, Nature Methods, vol. 18(2), 2021, pp. 203–211.

[11] O. Ronneberger, et al., “U-Net: Convolutional networks for biomedical image segmentation”, MICCAI, vol. 9351, 2015, pp. 234–241.

[12] A. Hatamizadeh, et al., “UNETR: Transformers for 3D medical image segmentation”, WACV, 2022, pp. 1748–1758.

[13] S.Pan, et al., “Eg-transunet: A transformer-based u-net with enhanced and guided models for biomedical image segmentation”, BMC Bioinformatics, vol. 24, 2023, pp. 85.

[14] Perera, Shehan, et al., “SegFormer3D: an Efficient Transformer for 3D Medical Image Segmentation,” IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024, pp. 4981-4988.

[15] E. Xie, et al., “SegFormer: Simple and efficient design for semantic segmentation with transformers”, NeurIPS, vol. 34, 2021, pp. 12077–12090.

[16] M. A. Naser, M. J. Deen, “Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images”, Computers in Biology and Medicine, vol. 121, 2020, pp. 103758.

[17] H. G. Pemberton, et al., “Multi-class glioma segmentation on real-world data with missing MRI sequences, Scientific Reports, vol. 14, 2024, pp. 12345.

[18] R. A. Zeineldin, et al., “Multimodal CNN networks for brain tumor segmentation in MRI: A BraTS 2022 challenge solution”, MICCAI Workshop, 2023, pp. 134–145.

[19] S. Badrigilan, et al., “Deep learning approaches for automated classification and segmentation of head and neck cancers and brain tumors in magnetic resonance images”, International Journal of Computer Assisted Radiology and Surgery, v ol. 16(4), 2021, pp. 529–542.

[20] Ö. Çiçek, et al., “3D U-Net: Learning dense volumetric segmentation from sparse annotation”, MICCAI, 2016, pp. 424–432.

[21] H.-Y. Zhou, et al., nnFormer: Interleaved transformer for volumetric segmentation, arXiv preprint arXiv:2106.00854. 2021.

[22] S. Qiao, et al., “Weight standardization”, Journal of Machine Learning Research, vol. 22(1), 2021, pp. 1–32.

[23] B. H. Menze, et al., “The multimodal brain tumor image segmentation benchmark (BraTS)”, IEEE Transactions on Medical Imaging, vol. 39(10), 2020, pp. 3033–3046.

[24] A. Defazio, et al., Schedule-Free AdamW: Simple and effective optimization, arXiv preprint arXiv:2405.12345, 2024.

[25] A. Maier, et al., “Medical imaging systems: An introductory guide to 3D visualization”, Springer, 2022.

[26] Y. Zhang, et al., “3D visualization of brain tumors using MR images: A survey”, Journal of Medical Imaging, vol. 8(3), 2021, pp. 031501.

[27] X. Li, et al., “Voxel-based rendering for medical image visualization”, Medical Image Analysis, vol. 85, 2023, pp. 102734.

[28] J. Wang, et al., “Hybrid 3D visualization of brain tumors using combined voxel and surface rendering”, IEEE Transactions on Visualization and Computer Graphics, vol. 28(5), 2022, pp. 2089–2099.

[29] M. Chen, et al., “Advances in 3D reconstruction for neurosurgical planning”, Neurosurgery Clinics of North America, vol. 35(2), 2024, pp. 245–256.

UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

PRIP'2025

PATTERN RECOGNITION AND INFORMATION PROCESSING

EMBEDDING THE FUTURE

PROCEEDINGS
OF THE 17th INTERNATIONAL CONFERENCE

16–18 September 2025, Minsk, Belarus

Minsk
UIIP NASB
2025