計測自動制御学会東北支部 60 周年記念学術講演会 (2024.12.17)

資料番号 S60-3

マルチモーダルMRIを用いた脳腫瘍セグメンテーションの深層学習モデルに関する評価

Comparative Evaluation of MRI Modalities for Deep Learning-Based Brain Tumor Segmentation

Wentong Zhou[*], Yuwen Zeng[*], 張 暁勇[**], 劉 皓洋[*], 市地 慶[*], 本間 経康[*]

Wentong Zhou[*], Yuwen Zeng[*], Xiaoyong Zhang[*], Haoyang Liu[*], Kei Ichiji[*], Noriyasu Homma[*]

[*]東北大学, [**]仙台高等専門学校

[*]Tohoku University, [**]National Institute of Technology, Sendai College

キーワード： 脳腫瘍のセグメンテーション（brain tumor segmentation），磁気共鳴画像法 (magnetic resonance imaging, MRI)，深層学習(deep learning)，マルチモーダル(multi-modality)

連絡先：〒980-8575 仙台市青葉区 星陵町 2-1 東北大学 大学院医学系研究科 知能システム医工学分野，周雯童 Tel.: 022-717-8190, E-mail: zhou.wentong.p1@dc.tohoku.ac.jp

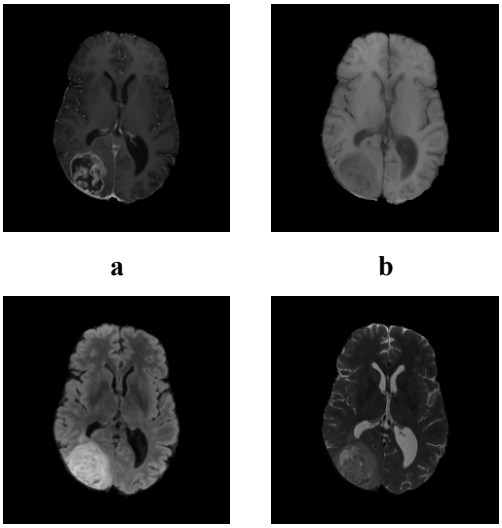
1. Introduction

Brain tumor accounts for 66.5% of brain and malignant central nervous system tumors, and there has been no significant advance in prevention, early detection, and treatment of brain tumor over the past four decades [1)]. Tan *et al.* reported that despite improvements in short-term survival for glioblastoma patients, the five-year survival rate remains just 5.8% [2)].

Brain tumor segmentation is critical in clinical applications such as surgical planning, image-guided interventions, tumor monitoring, and radiation therapy [3)]. However, reading scans is time-consuming for doctors due to the many MRI slices, the need to interpret multiple modalities, and the irregular shapes and heterogeneity of brain tumors. Therefore, deep learning methods have been introduced to automatically segment tumor regions for efficient and accurate analysis.

Previous studies predominantly utilized deep learning methods for brain tumor segmentation in MRIs, generally achieving good results [4, 5)]. However, segmentation performance on the whole tumor region consistently outperforms that of the enhancing tumor and tumor core, with Dice coefficient 3–10% higher. Furthermore, Bjoern *et al*. revealed that different approaches excel in different tumor sub-regions, no single approach consistently ranks the best across all subregions [6)].

Most studies used MRI data but overlooked differences among MRI modalities. As shown in Fig. 1, the same tumor tissue appears differently across modalities, with some features prominent in



a **b**

c **d**

Fig. 1 An example of multimodal MRI for brain tumor: **a.** T1c modality, **b.** T1 modality, **c.** T2 modality, **d.** FLAIR modality.

one but less distinguishable in others.

The purpose of this study is to analyze the impact of various combinations of MRI modalities on segmentation performance across different tumor subregions. By testing different modality combinations, we can observe how modalities interact and affect segmentation outcomes, helping to identify the most effective combinations for more accurate segmentation.

2. Method

2.1 Dataset

The dataset used in this research comes from BraTS2023 [7, 8], which was manually divided into training set (1000 cases), validation set (200 cases), and test set (50 cases). This dataset contains four MRI modalities for each patient: T1, T1c, T2, and FLAIR. Each provides a distinctive contrast of the brain structure and pathology.

In addition, this dataset provides three annotations: enhancing tumor (ET), peritumoral edematous/invaded tissue (ED), and necrotic tumor core (NCR), which are annotated across all image data.

2.2 Models

Three extended models of U-Net [9]: 3D U-Net [10], Attention U-Net [11], and UNETR [5] were trained from scratch and evaluated in this study.

3D U-Net is an extension of the original U-Net architecture. It uses 3D convolutions to capture spatial relationships in all three dimensions, leading to more accurate segmentation results for 3D medical data.

Attention U-Net incorporates attention blocks [] during the upsampling phase of the 3D U-Net, allowing the upsampling stage to focus on important spatial features. Attention gate consistently improves the prediction performance of U-Net while preserving computational efficiency.

UNETR follows a U-shaped design for its encoder-decoder architecture. It uses a Transformer [12] as the encoder to capture global features, making it well-suited for segmenting complex 3D MRI data.

2.3 Experimental setting

In this study, we utilized multi-modal MRI-based 3D deep learning models for brain tumor segmentation. The input consists of different combinations of MRI modalities (T1, T1c, T2, and FLAIR), which are stacked into a multi-channel 3D tensor and fed into the model. As shown in Fig. 2 the model will output the prediction of the three tumor regions ET, ED, and NCT.

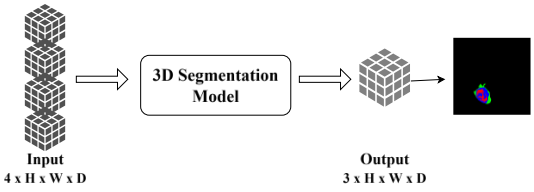


Fig. 2 Example of input and output of the models

As pre-processing, the T1c, T1, T2, and FLAIR data were normalized and resized to 160×160×128, corresponding to height (H),

width , and depth. Data augmentation included random rotations, flips, and Gaussian noise to enhance variability and model generalization.

The optimization strategies of models were experimentally determined. For the 3D U-Net and Attention U-Net, Stochastic Gradient Descent (SGD) optimizer was utilized with a momentum of 0.9, an initial learning rate of 1×10^{-3}, and a weight decay of 5×10^{-4}. The UNETR model employed the Adam optimizer with an initial learning rate of 1×10^{-3} and a weight decay of 1×10^{-5}. To further enhance the training process, all models were implemented with a cosine annealing learning rate scheduler, which adjusted the learning rate from a specified base value of 2×10^{-3} to a final minimum value of 1×10^{-3}. The duration of warmup epochs 10, and a starting warmup value of 5×10^{-4}. To ensure reproducibility, a fixed random seed was set for all experiments.

The loss function combined cross-entropy loss $\mathcal{L}_{\mathcal{CE}}$ and soft Dice loss $\mathcal{L}_{\mathcal{DL}}$ to improve medical image segmentation. $\mathcal{L}_{\mathcal{CE}}$ focuses on pixel-wise classification accuracy:

$$\mathcal{L}_{\mathcal{CE}}(y, \hat{y}) = -\sum_{i=1}^{N} w_i \cdot y_i \cdot log(\hat{y}_i) \quad (1)$$

$\mathcal{L}_{\mathcal{DL}}$ focuses on the overlap between predictions and annotations:

$$\mathcal{L}_{\mathcal{DL}} = 1 - \frac{2\sum_{i=1}^{N} \hat{y}_i \cdot y_i}{\sum_{i=1}^{N} \hat{y}_i + \sum_{i=1}^{N} y_i + \epsilon} \quad (2)$$

where y_i and $\hat{y}_i$ represent the predictions and annotations, respectively. Besides, w_i is the pixel weight, N denotes the number of pixels and ϵ is a very small constant, typically used to prevent division by zero.The total loss $\mathcal{L}_{total}$ is a weighted combination of both losses, where α represents the weighting factor.

$$\mathcal{L}_{total} = (1 - \alpha) \cdot \mathcal{L}_{\mathcal{CE}} + \alpha \cdot \mathcal{L}_{\mathcal{DL}} \quad (3)$$

3. Result

As shown in Fig. 3, the blue, green, and red represent enhancing tumor, peritumoral edematous/invaded tissue, and necrotic tumor core, respectively.

Models were evaluated by their performance on segmenting each subregion. According to the official dataset, enhancing tumor (ET) refers to the actively enhancing part of the tumor; tumor core (TC = ET + NRC) includes both the enhancing tumor and the necrotic core of the tumor; and whole tumor (WT= NCR + ED + ET) encompasses the entire tumor region.

Dice similarity metric was used to measure the overlap between the ground truth and the segmentation result:

$$Dice = \frac{2||X \cap Y||}{||X|| + ||Y||} \quad (4)$$

where X is the prediction result of the models, and Y is the ground truth.

Table 1 presents the segmentation results for both single-modality and multi-modality combinations. In the single-modality experiments, only the T1c modality effectively guided the model in segmenting the ET and TC, while the other three modalities primarily contributed to the segmentation of the WT. In the multi-modality experiments, removing T1c caused a significant decrease in the Dice scores of all three models in both ET and TC. In contrast, removing any single modality except T1c had a relatively minor impact on segmentation performance. It is difficult to distinguish which modality, FLAIR or T2, contributes more to WT segmentation in the single-modality experiments. However, in the experiments where one modality was

missing, a clearer trend can be observed. When FLAIR is missing, the WT segmentation performance of all three models drops more significantly, which can show that it has more contribution on segmenting WT region.

The best segmentation performance was achieved when using all four modalities, as showcased in Fig.3.

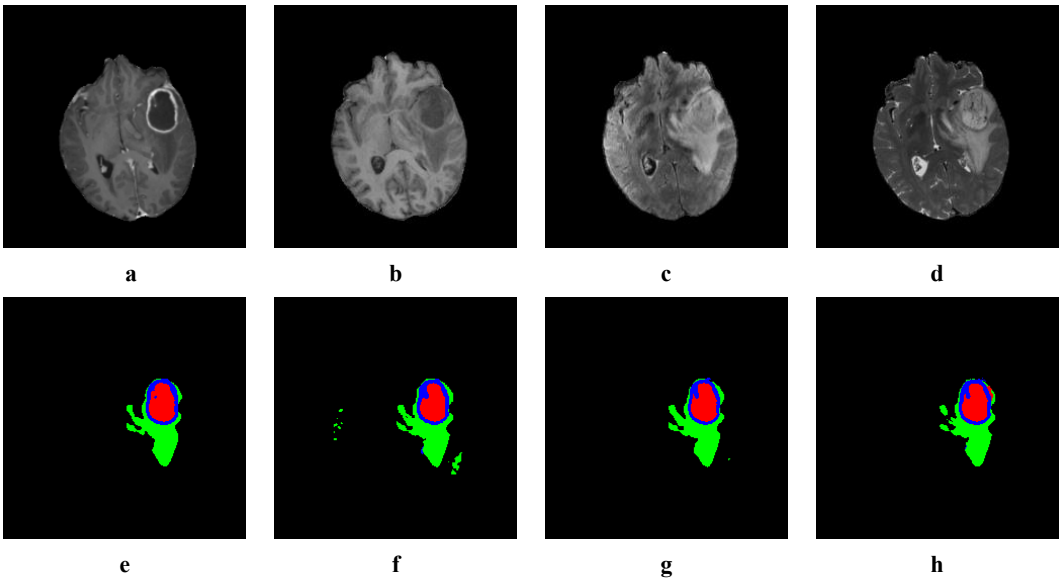


Fig. 3 An example of ground truth and segmentation results: **a**. T1c, **b.** T1, **c.** FLAIR, **d.** T2**, e**. ground truth, **f**. 3D U-Net, **g**. Attention U-Net, **h**. UNETR. (Segmentation annotation: the blue, green, and red represent enhancing tumor, peritumoral edematous/invaded tissue, and necrotic tumor core, respectively.)

Table 1 Dice score with different combinations of modalities

Modalities				**ET**			**TC**			**WT**		
T1c	T1	F	T2	3D U-Net	Att U-Net	UNERT	3D U-Net	Att U-Net	UNERT	3D U-Net	Att U-Net	UNERT
●	○	○	○	**0.729**	**0.694**	**0.553**	**0.746**	**0.714**	**0.572**	0.646	0.659	0.487
○	●	○	○	0.299	0.206	0.007	0.499	0.353	0.054	0.614	0.524	0.350
○	○	●	○	0.065	0.162	0.226	0.085	0.261	0.394	**0.676**	**0.731**	**0.783**
○	○	○	●	0.288	0.267	0.147	0.478	0.432	0.269	**0.730**	**0.630**	**0.637**
●	●	●	○	0.743	0.746	0.727	0.768	0.760	0.725	**0.824**	**0.856**	**0.841**
●	●	○	●	0.743	0.716	0.644	0.744	0.714	0.597	**0.809**	**0.748**	**0.700**
●	○	●	●	**0.757**	0.712	0.717	**0.773**	0.750	0.694	0.844	0.871	0.831
○	●	●	●	0.300	0.307	0.151	0.489	0.489	0.253	0.848	0.860	0.817
●	●	●	●	0.751	**0.773**	**0.765**	0.767	**0.784**	**0.774**	**0.856**	**0.875**	**0.892**

* ET is represented by blue, TC is represented by both blue and red, and WT is represented by all three colors.

4. Discussion and Conclusion

In this study, we analyzed the segmentation performance of three models—3D U-Net, Attention U-Net, and UNETR—on different combinations of MRI modalities (T1, T1c, T2, FLAIR) across brain tumor subregions. Based on the results, we had the followed findings: Firstly, as illustrated in Fig.3, the edges of ET and TC are distinctly visible in the T1c modality, which corresponds to that segmentation performance for ET and TC is optimal when the T1c modality is included in the input data. Meanwhile, FLAIR and T2 modalities perform best in the segmentation of the WT region. However, based on the results in Table 1, the specific contribution of each modality appears to vary depending on the model and input data. Models that leverage attention mechanisms, such as Attention U-Net and UNETR, seems more effective at extracting WT information from the T2 modality.

The limitations of this study include that all three models fuse the input modalities at an early stage, without comparing alternative fusion strategies. Additionally, due to time and resource constraints, experiments using only two modalities as input were not conducted. Future work will focus on completing experiments with two-modality inputs and testing models that incorporate different fusion strategies.

In conclusion, this study offers a comprehensive analysis of the effects of different MRI modality combinations on tumor subregion segmentation performance. We identified the most influential modalities for each subregion: T1c for ET and TC, and T2, FLAIR for WT. Complementary modality combinations were shown to effectively improve segmentation accuracy. This research lays the groundwork for future studies on multi-modal MRI feature fusion, emphasizing the importance of both modality selection and model choice.

Reference

1) K. Miller, Q. Ostrom, C. Kruchko, N. Patil, T. Tihan, G. Cioffi *et al.*: Brain and other central nervous system tumor statistics, *CA: A Cancer Journal for Clinicians*, 71-5, 381/406 (2021).
2) A. Tan, D. Ashley, G. López, M. Malinzak, H. Friedman, and M. Khasraw: Management of glioblastoma: State of the art and future directions, *CA: A Cancer Journal for Clinicians*, 70-4, 299/312 (2020).
3) U. Baid, S. Ghodasara, S. Mohan, M.Bilello, E. Calabrese *et al*. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. (2021).
4) Y. Cai, Y. Zhang, Z. Zhang, M. Gao, Y. Lin, S. Han, Y. Lu, X. Wu, and J. Han: Swin Unet3D: A three-dimensional medical image segmentation network combining vision transformer and convolution, *BMC Medical Informatics and Decision Making*, 23-1, 33 (2023).
5) A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, and D. Xu: Unetr: Transformers for 3D medical image segmentation, *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 574/584 (2022).
6) B. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, *et al.*: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10) (2015)
7) BraTS 2023 – Synapse:

https://www.synapse.org/Synapse:syn51156910/wiki/622351

8) S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, *et al.*, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017)
9) O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv., 234/241(2015).
10) Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger: 3D U-Net: Learning dense volumetric segmentation from sparse annotation, *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, Athens, Greece, Proceedings, Part II, 19, (2016).
11) O. Oktay, J. Schlemper, L. Folgoc, M Lee, M. Heinrich *et al.* "Attention u-net: Learning where to look for the pancreas. arXiv." *arXiv:1804.03999* 10 (2018).
12) A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, and I. Polosukhin. "Attention is all you need." In Advances in neural information processing systems, pages 5998/6008, (2017).