뇌종양 분할을 위한 3D 이중 융합 주의 네트워크

Hoang-Son Vo-Thanh[2], Tram-Tran Nguyen Quynh[1,2], Nhu-Tai Do[1], 김수형[1]
[1]전남대학교 인공지능융합학과
[2]베트남 호치민 외국어 정보 기술 대학 정보 기술과
hoangson.vothanh@gmail.com, tramtnq@huflit.edu.vn, ntdo@jnu.ac.kr, shkim@jnu.ac.kr

3D Dual-Fusion Attention Network for Brain Tumor Segmentation

Hoang-Son Vo-Thanh[2], Tram-Tran Nguyen Quynh[1,2], Nhu-Tai Do[1], Soo-Hyung Kim[1]
[1]Dept. of Artificial Intelligence Convergence, Chonnam National University
[2]Dept. of Information Technology, HCMC University of Foreign Language Information Technology, Vietnam

Abstract

Brain tumor segmentation problem has challenges in the tumor diversity of location, imbalance, and morphology. Attention mechanisms have recently been used widely to tackle medical segmentation problems efficiently by focusing on essential regions. In contrast, the fusion approaches enhance performance by merging mutual benefits from many models. In this study, we proposed a 3D dual fusion attention network to combine the advantages of fusion approaches and attention mechanisms by residual self-attention and local blocks. Compared to fusion approaches and related works, our proposed method has shown promising results on the BraTS 2018 dataset.

1. Introduction

Gliomas is a type of brain tumor with a poor prognosis, rapid development, and disease with a high mortality rate. Many works have used deep learning [1, 2, 3] to identify tumor from MRI modalities. However, this issue is still a challenge by the appearance and location diversity of tumor.

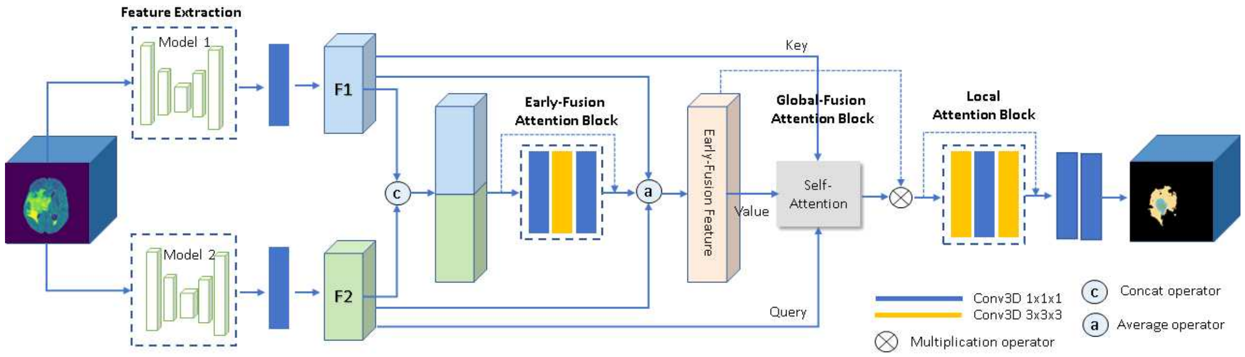
Recently, many researchers have focused on attention mechanisms [4]. The models can learn the visual attention masks, focusing on necessary regions to grasp the essential contents for identifying tumors. In addition, fusion approaches [5] utilize mutual benefits from multi-models to boost the performance results by late fusion and early fusion from the last feature maps of pre-train models.

In this study, we proposed a 3D dual fusion attention network to learn mutual benefits from the last feature maps of pre-trained models where global and local attention blocks exploit the relations of the input feature maps. We conducted experiments on BratTS 2018 [6] and archived good results compared to related works and fusion techniques.

2. Proposed Method

Overview: Our problem is to classify every voxel that belongs to one of four classes: whole tumor (WT), tumor core (TC), enhancing tumor (ET), and background. The inputs are the last feature maps of two pre-trained models denoted by $F_1 \in R^{h \times w \times d \times f_1}$ and $F_2 \in R^{h \times w \times d \times f_2}$. The goal is to learn mutual benefits of two feature maps to maximize the performance. Our proposed model is shown in Fig.1.

Early-Fusion Attention block: We first employ two Conv3D 1x1x1 layers right after last output feature maps of pre-train models to normalize the size of the feature maps and concatenate two feature maps to the joint feature map. The block inspired by [7] plays the role of an early fusion approach to learn the joint feature



(그림 1) 3D Dual-Fusion Attention Network

map using two Conv3D layers 3x3x3, a Conv3D layer 1x1x1 in the middle of them and a shortcut path for residual learning. Finally, the average operator is used to merge the output feature map, two normalized input features and then pass through to global and local attention blocks.

Global-Fusion Attention Block: This block, inspired by scale dot-product attention [8], exploits the contextual relationships of a query and key-value pairs at every voxel. The query Q and key K are taken from two normalized input feature maps, while the value V is the early-fusion feature map. We calculate the output by scale dot-product equation and apply residual learning by element-wise multiplication of the result with V.

$$G(Q,K,V)=\left(softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V\right)\otimes V$$

where K, Q, V are key, query, and value, respectively. ⊗ is the element-wise multiplication.

Local Attention Block: The goal of the block is to learn the spatial context of the global-fusion feature map. It involves two Conv3D layers 1x1x1, 1 Conv3D layer 3x3x3 in the middle using GELU activation. A skip connection is used to avoid the dead blocks from the learning process.

3. Experiments and Results

Experimental Setup: We conducted the experiments on BratTS 2018 [6] with four modalities FLAIR, T1w, T2w, and T1Gd. There are 285 MRI scans in the training set and 66 MRI scans in the validation set without ground truth. Every output label consists of three tumor regions: whole tumor (WT), tumor core (TC), and enhancing tumor (ET).

The training data was split following a ratio of 8:2 for training and validating with augmentation by random rotation, flip, and sensitivity. The input patch size was 128 x 128 x 128. The training configuration was set up as follows: the number of epochs was 100, and early stopping was applied after six patient epochs. The optimal function was Adam, with an initial learning rate of 3e-1. The loss and metric functions were dice loss and dice score.

Results and Discussion: Firstly, we trained DynUnet [9], and SegResnet [10] models with the performance results shown in Table 1.

<표 1> Performance of pretrained models

Validation	Dice Score		
	WT	TC	ET
DynUnet	0,8623	0,7857	0,7749
SegResnet	0,81753	0,74489	0,6681

After that, we applied the fusion techniques [x] with late and early-fusion methods to evaluate the performance of two pre-train models. The late fusion used the average on probability scoring to produce the result. The early fusion employed last feature maps of two pre-trained models to make the join feature map by the concatenating operator, followed by two Conv3D 1x1x1 layers.

<표 2> Comparision to fusion approaches

Validation	Dice Score		
	WT	TC	ET
Late fusion	0,88829	0,81713	0,73897
Early fusion	0,89303	0,79925	0,76008
Our method	**0,90267**	**0,84107**	**0,77842**

The performance results of our method comparing to fusion approaches were shown in

Table 2. The early fusion gave results better than the late fusion, with dice score values of 89% and 88% in WT, respectively. In contrast, our method helped the dice score achieve 90%.

Moreover, our method had the best results comparing to related works shown in Table 3.

<표 3> Comparision to related works

Validation set	Year	Dice		
		WT	TC	ET
Zhang et al.[1]	2021	0,9	0,836	**0,791**
Zhou et al. [2]	2022	0,876	0,784	0,688
Akbar et al.[3]	2022	0,896	0,798	0,777
Ours method		**0,9027**	**0,8411**	0,77842

4. Conclusion

This study proposed a novel 3D Dual-Fusion attention network to exploit mutual benefits from the last feature maps of two pre-train models. Our method uses the early-fusion block to learn the joint feature map of dual models. It exploits the global context using global-fusion attention models by a query, key-value pairs from the joint feature map, and two input feature maps. Finally, we use a local attention block to learn the local context of the global-fusion feature map and avoid unintended features by residual connection. Our experimental results on BratTS 2018 dataset have proven to be effective.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1I1A3A04036408) and also supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-02068, Artificial Intelligence Innovation Hub).

References

[1] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, and Y. Yu, “Cross-modality deep feature learning for brain tumor segmentation,” Pattern Recognit., vol. 110, p. 107562, Feb. 2021, doi: 10.1016/j.patcog.2020.107562.

[2] T. Zhou, S. Ruan, P. Vera, and S. Canu, “A Tri-Attention fusion guided multi-modal segmentation network,” Pattern Recognit., vol. 124, p. 108417, Apr. 2022, doi: 10.1016/j.patcog.2021.108417.

[3] A. S. Akbar, C. Fatichah, and N. Suciati, “Single level UNet3D with multipath residual attention block for brain tumor segmentation,” J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 6, Part B, pp. 3247–3258, Jun. 2022, doi: 10.1016/j.jksuci.2022.03.022.

[4] M. Guo, T. Xu, J. Liu, Z. Liu, P. Jiang, T. Mu, S. Zhang, R. Martin, M. Cheng, and S. Hu, “Attention mechanisms in computer vision: A survey,” Computational Visual Media, vol. 8, no. 3, pp. 331–368, Sept. 2022.

[5] K. Gadzicki, R. Khamsehashari, and C. Zetzsche, “Early vs Late Fusion in Multimodal Convolutional Neural Networks,” in 2020 IEEE 23rd International Conference on Information Fusion (FUSION), Jul. 2020, pp. 1–6. doi: 10.23919/FUSION45008.2020.9190246.

[6] A. Myronenko, “3D MRI Brain Tumor Segmentation Using Autoencoder Regularization,” in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum, Eds., Cham: Springer International Publishing, 2019, pp. 311–320.

[7] F. Chollet, “Xception: Deep Learning With Depthwise Separable Convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017.

[8] A. Vaswani et al., “Attention is all you need,” Adv. Neural Inf. Process. Syst., vol. 30, 2017.

[9] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” Nat. Methods, vol. 18, no. 2, Art. no. 2, Feb. 2021, doi: 10.1038/s41592-020-01008-z.

[10] A. Myronenko, “3D MRI brain tumor segmentation using autoencoder regularization,” in 4th International Workshop Brainlesion, MICCAI, Sept. 2019, pp. 311–320.