EoFormer: Edge-Oriented Transformer for Brain Tumor Segmentation

Dong She[1,2], Yueyi Zhang[1,2(✉)], Zheyu Zhang[1], Hebei Li[1], Zihan Yan[3], and Xiaoyan Sun[1,2(✉)]

[1] University of Science and Technology of China, Hefei 230026, China
{zhyuey,sunxiaoyan}@ustc.edu.cn
[2] Hefei Comprehensive National Science Center, Institute of Artificial Intelligence, Hefei 230088, China
[3] Beijing Tiantan Hospital, Capital Medical University, Beijing 100050, China

Abstract. Accurate segmentation of brain tumors in MRI images requires precise detection of the edges. However, this crucial information has been overlooked by existing methods. In this paper, we introduce the **E**dge-**o**riented Trans**former** (EoFormer) which specifically captures and enhances edge information for brain tumor segmentation. Our approach incorporates a CNN-Transformer encoder to comprehensively improve the feature representation capability. The CNN structure captures low-level local features in the image, while the Transformer structure establishes long-range dependencies between features to generate high-level global features. Additionally, the decoder of our approach utilizes two edge sharpening modules, the Edge-oriented Sobel and Laplacian modules, which enhance the edge information. We also introduce efficient attention and re-parameterization techniques that make EoFormer computationally efficient. Experimental results on the BraTS 2020 dataset and a private medulloblastoma dataset demonstrate the superiority of our approach compared with existing state-of-the-art methods. Moreover, our method achieves this with limtied model parameters and lower FLOPs, making it a promising approach for future research. The code is available at https://github.com/sd0809/EoFormer.

Keywords: Brain tumor segmentation · Edge-oriented module · Transformer

1 Introduction

Accurate segmentation of brain tumors from MRI images is of great significance as it enables more accurate assessment of tumor morphology, size, location, and distribution range, thereby providing clinicians with a reliable basis for diagnosis and treatment [16]. Physicians manually delineate the tumor regions based on the varying signal intensities between diseased and normal tissues. This signal

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43901-8_32.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14223, pp. 333–343, 2023.
https://doi.org/10.1007/978-3-031-43901-8_32

disparity constitutes the edge information in the images, making it essential for accurate tumor segmentation.

CNN-based networks, such as UNet [2], SegResNet [15], and nnUNet [8], have made significant progress in the field of medical image segmentation, including brain tumor segmentation. With the emergence of Transformer [19], which is capable of modeling long-range dependencies that CNNs struggle with, a number of CNN-Transformer hybrid networks have been proposed, such as TransBTS [21], UNETR [7], Swin-UNETR [6] and NestedFormer [23], leading to further improvements in brain tumor segmentation. However, the performance of existing brain tumor segmentation methods are still unsatisfactory, especially for the segmentation of edges between tumor lesion and normal tissues.

Considerable advancement has been achieved in the field of natural image segmentation by focusing on the edge information [3,11,18,25], and this idea is also being applied to medical image segmentation. Some methods utilize the distance-dependent objective functions to generate more accurate edge predictions. Karimi et al. [9] design a Hausdorff-based metric loss function to minimize Hausdorff distance (HD), which is used to measure the edge distance between two point sets. Other methods [1,12,20,22] involve post-processing uncertain regions to more accurately segment pixels near edges. For example, BAT [20] considers global context to coarsely locate lesion area and paying special attention to the ambiguous area to specify the exact edges of the skin cancers. Similarly, Xie et al. [22] use the confidence map to evaluate the uncertainty of each pixel to enhance the segmentation of the ambiguous edges of ultrasound images. However, the methods mentioned above are not suitable for brain tumor segmentation for two main reasons. (1) Efficiency. For instance, Karimi et al. [9] require the calculation of the HD at each iteration, which is both time-consuming and computationally demanding. Moreover, processing every slice of large volumes of MRI images at the pixel-level is impractical. (2) Task Complexity. Unlike many other medical image segmentation tasks that involve the segmentation of a single ROI, brain tumor segmentation requires the simultaneous segmentation of three regions: the whole tumor (WT), the tumor core (TC), and the enhancing tumor (ET) regions. Therefore, in addition to focusing on the edge between the tumor lesion and normal tissue to segment the WT, it is also necessary to consider the edges within the tumor in order to segment the TC and ET regions.

In this paper, we propose an **E**dge-**o**riented trans**Former** (EoFormer), for efficient and accurate brain tumor segmentation. We design a CNN-Transformer based encoder for more effective feature representation, called Efficient Hybrid Encoder (EHE). Specifically, the input image is first processed by the CNN blocks to extract low-level local features. Then, the extracted features are fed into the transformer blocks to create long-range dependencies, resulting in the formation of high-level semantic features. In addition, to provide more accurate edge predictions, we design two edge sharpening modules in the decoder, called Edge-oriented Sobel (EoS) and Laplacian (EoL) modules. By implicitly embedding Sobel and Laplacian filters into the convolution layers to extract 1st-order and 2nd-order differential features, the two modules could enhance the edge information contained in the feature maps. In order to reduce the computational and

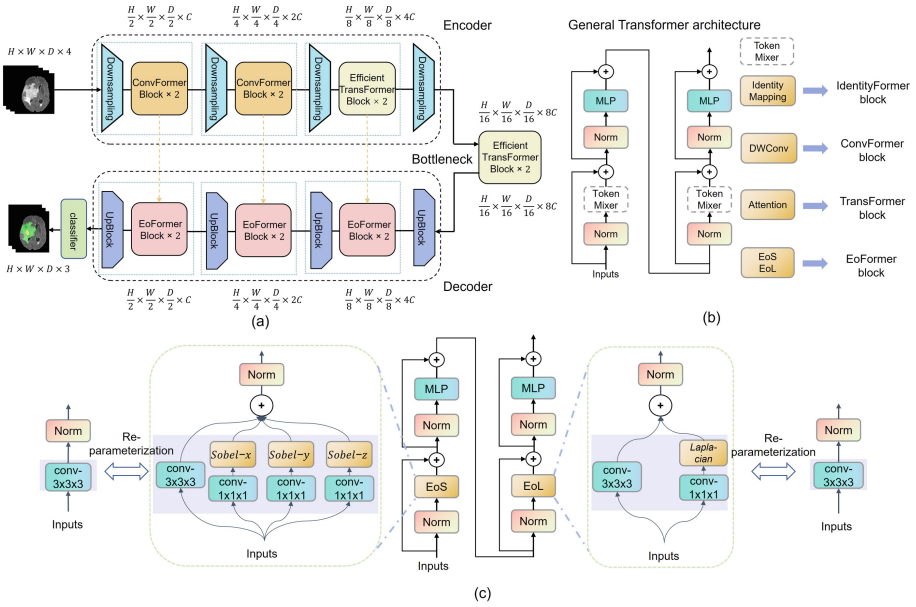

Fig. 1. (a) Overview of the proposed EoFormer. (b) The general transformer architecture and its variants. (c) The EoFormer block, the Edge-oriented Sobel Module (EoS) and the Edge-oriented Laplacian Module (EoL).

memory complexity of the model, we replace the vanilla attention module with our extended efficient attention module [17]. To simplify the model architecture and reduce inference time, we also introduce the re-parameterization technique [4,5]. Our model has been evaluated on both the publicly BraTS 2020 dataset and a private medulloblastoma segmentation dataset. The results demonstrate that EoFormer clearly outperforms the state-of-the-art methods with limited model parameters and lower FLOPs (see more in supplementary material).

2 Method

Figure 1(a) presents an overview of the proposed EoFormer architecture, which comprises two components: (1) an EHE encoder and bottleneck which are used to capture low-level local features and learn a comprehensive feature representation. (2) a decoder which incorporates edge-oriented modules to enhance the edge information in features.

2.1 Efficient Hybrid Encoder

The EHE, shown in Fig. 1(a), comprises four stages, each of which consists of a feature extraction module and a downsampling module. All four feature

extraction modules follow the same paradigm of the general transformer architecture (see Fig. 1(b)), which regards the attention module in the transformer as a token mixer [24]. In the first two stages of EHE, we use depth-wise convolution (DWConv) to instantiate the token mixer, called the ConvFormer block. In the third stage and bottleneck, we use the multi-head self-attention (MSA) to instantiate the token mixer, which is the typical transformer block. For each stage i, given an input feature map X, the output of the i^{th} block $X^{''}$ is computed as follows:

$$X^{'} = \text{TokenMixer}_i(\text{Norm}(X)) + X, \quad X^{''} = \text{MLP}(\text{Norm}(X^{'})) + X^{'}, \tag{1}$$

where the $\text{TokenMixer}_i(\cdot)$ corresponds to DWConv ($i \in \{0, 1\}$) and MSA ($i \in \{2, 3\}$), $\text{Norm}(\cdot)$ represents layer normalization, and $\text{MLP}(\cdot)$ denotes the Multilayer Perceptron. Our approach combines the strengths of CNN and transformer to create a more powerful encoder that can extract both local and global information from input data.

We address the computational and memory complexity issues that arise from 3D input by replacing the vanilla attention with our extended 3D efficient attention. Assuming the size of the input feature is n and the dimensionality is d, the input feature $X \in \mathbb{R}^{n \times d}$ pass through three linear layers to generate the queries $Q \in \mathbb{R}^{n \times d_k}$, keys $K \in \mathbb{R}^{n \times d_k}$ and values $V \in \mathbb{R}^{n \times d_v}$. The vanilla attention $D(\cdot)$ and the efficient attention $E(\cdot)$ are computed as follows:

$$D(Q,K,V) = \rho(QK^T)V, \quad E(Q,K,V) = \rho(Q)(\rho(K)^T V), \tag{2}$$

where $\rho(\cdot)$ is the softmax activation function, T represents the matrix transpose operator. The efficient attention reduces the memory complexity and computational complexity of vanilla attention from $\mathcal{O}(n^2)$ and $\mathcal{O}(dn^2)$ to $\mathcal{O}(dn + d^2)$ and $\mathcal{O}(nd^2)$, where $d = d_v = 2d_k$.

2.2 Edge-Oriented Transformer Decoder

We design the EoFormer block (see Fig. 1(c)) in the decoder, which instantiates the token mixer with our proposed Edge-oriented Sobel module (EoS) and Edge-oriented Laplacian module (EoL). Each edge-oriented module includes a normal $3 \times 3 \times 3$ convolution and an edge detection path to extract the 1st-order or the 2nd-order spatial derivatives from intermediate features. This design allows the edge-oriented module to efficiently extract the edges and textures of the features. Moreover, to boost the segmentation performance without sacrificing efficiency, we incorporate the re-parameterization technique in the decoder.

Edge-Oriented Sobel Module. We use a dual-branch structure, where the input feature X is simultaneously processed by two different branches. The first branch contains a $3 \times 3 \times 3$ convolution that extracts basic features from the input. The second branch, which is responsible for edge extraction, first uses a $C \times C \times 1 \times 1 \times 1$ convolution to enhance the interaction between channel

features of X, then utilizes a learnable scaled Sobel filter to extract the 1st-order differentiation edge information from X. This filter is capable of detecting edges in three directions (i.e. horizontal, vertical, and orthogonal directions), so it comprises three filters M_x, M_y, and M_z, each of which is represented by a 3 $\times$ 3 $\times$ 3 array. Take M_x as an example, which is described as:

$$M_x[0,:,:] = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, M_x[1,:,:] = \begin{bmatrix} -2 & 0 & +2 \\ -4 & 0 & +4 \\ -2 & 0 & +2 \end{bmatrix}, M_x[2,:,:] = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}.$$

We then apply a learnable scaling matrix $S \in \mathbb{R}^{C\times1\times1\times1}$ to M_x, which allows for dynamic adjustment of the scaling factor in each channel. The resulting feature extracted from the scaled Sobel-x filter is denoted as:

$$F_x = \mathrm{DWConv}_{S\cdot M_x}(\mathrm{Conv}_{1\times1\times1}(X)), \tag{3}$$

where the '$\cdot$' denotes channel-wise multiplication; the $\mathrm{DWConv}_{S\cdot M_x}$ indicates that $\mathrm{DWConv}(\cdot)$ applies a $S \cdot M_x$ learnable scaled filter as its kernel weight. Similarly, F_y and F_z are processed in the same way. The final output of the EoS module, denoted as F_{sob}, is given by:

$$F_{sob} = \mathrm{Norm}(\mathrm{Conv}_{3\times3\times3}(X) + F_x + F_y + F_z). \tag{4}$$

Edge-Oriented Laplacian Module. Different from the Sobel filter that only extracts edges in the horizontal, vertical, and orthogonal directions, the Laplacian filter can extract edges in all directions. After extracting the 1st-order differentiation edge information, the intermediate features are then fed into the EoL module for extracting the 2nd-order differentiation edge information. Similarly, the feature F, obtained from the learnable scaling Laplacian filter, and the final output of the EoL module, denoted as F_{lap}, are defined as:

$$\begin{aligned} F &= \mathrm{DWConv}_{S\cdot M_{lap}}(\mathrm{Conv}_{1\times1\times1}(X)), \\ F_{lap} &= \mathrm{Norm}(\mathrm{Conv}_{3\times3\times3}(X) + F). \end{aligned} \tag{5}$$

Re-parameterization of the Edge-Oriented Modules. We introduce the re-parameterization [4,5] into the edge-oriented modules to boost the segmentation performance while maintaining high efficiency. Specifically, we explain the re-parameterization of the EoL module as follows:

$$\mathrm{W}_{\mathrm{rep}} = \mathrm{W}_{\mathrm{conv}_{3\times3\times3}} + \mathrm{W}_{\mathrm{conv}_{1\times1\times1}} * \mathrm{W}_{\mathrm{conv}_{\mathrm{lap}}} \tag{6}$$

$$\mathrm{B}_{\mathrm{rep}} = \mathrm{B}_{\mathrm{conv}_{3\times3\times3}} + \mathrm{W}_{\mathrm{conv}_{\mathrm{lap}}} * \mathrm{up}(\mathrm{B}_{\mathrm{conv}_{1\times1\times1}}) + \mathrm{B}_{\mathrm{conv}_{\mathrm{lap}}} \tag{7}$$

where '$*$' represents the convolution operation, W_{conv} means the weights of the convolution and B_{conv} denotes the bias, and $\mathrm{up}(\cdot)$ is the spatial broadcasting operation ,which upgrades the bias $B \in \mathbb{R}^{1\times C\times1\times1\times1}$ into $\mathrm{up(B)} \in \mathbb{R}^{1\times C\times3\times3\times3}$. In the inference stage, the output feature F is produced by a normal $3 \times 3 \times 3$ convolution as follows:

$$F = W_{\mathrm{rep}} * X + B_{\mathrm{rep}}. \tag{8}$$

Table 1. Quantitative comparison on BraTS 2020 dataset. The '∗' means the FLOPs we recalculate.

Method	Param(M)	FLOPs(G)	Dice (%) ↑				HD95 (mm) ↓			
			WT	TC	ET	Ave	WT	TC	ET	Ave
3D-UNet [2]	5.75	1449.59	90.01	80.68	79.18	83.29	8.591	10.91	5.932	8.477
SegResNet [15]	18.79	185.23	87.59	85.50	81.20	84.77	4.941	4.653	3.822	4.472
nnUNet [8]	5.75	1449.59	91.15	86.12	81.67	86.32	**3.532**	4.901	3.561	3.998
TransBTS [21]	32.99	333.00	90.54	84.93	79.91	85.12	3.916	4.843	4.501	4.420
UNETR [7]	102.06	203.32	90.77	84.11	79.12	84.67	4.917	5.054	3.943	4.638
Swin-UNETR [6]	61.98	793.92	**91.50**	84.06	80.98	85.51	3.386	5.080	3.640	4.035
NestedFormer [23]	10.48	209.58*	91.07	85.30	80.10	85.49	3.583	4.735	4.391	4.236
EoFormer	6.28	91.81	90.84	**86.38**	**83.22**	**86.81**	3.974	**4.500**	**3.432**	**3.968**

Table 2. Quantitative comparison on MedSeg dataset.

Method	Dice (%) ↑			HD95 (mm) ↓		
	WT	ET	Ave	WT	ET	Ave
3D-UNet [2]	61.52	50.71	56.11	17.43	14.62	16.03
SegResNet [15]	76.76	55.60	66.18	7.810	9.411	8.611
TransBTS [21]	72.35	55.56	63.96	11.09	11.19	11.14
UNETR [7]	73.38	56.02	64.70	9.112	12.70	10.90
Swin-UNETR [6]	70.10	**60.79**	65.44	9.766	9.339	9.552
NestedFormer [23]	**79.89**	55.76	67.83	7.099	12.08	9.587
EoFormer	79.74	59.10	**69.42**	**6.978**	**7.104**	**7.041**

3 Experiment

3.1 Dataset and Evaluation Metric

In order to validate the performance of EoFormer, we conduct extensive experiments on both the publicly available BraTS 2020 dataset and a private medulloblastoma segmentation dataset (MedSeg).

The BraTS 2020 dataset [14] consists of MRI image data from 369 patients, with each patient having four modalities (T1, T1ce, T2 and T2-FLAIR) of skull-striped MRI, which are aligned to a standard brain template. The training/validation/test split follows 315/16/37 according to recent works [10,23].

The MedSeg dataset includes MRI images of T1, T1ce, T2, and T2 FLAIR modalities from 255 patients with medulloblastoma. The dataset includes manual annotations of the WT and ET regions. These annotated masks are reviewed by two experienced physicians to ensure the accuracy of the annotated results. The images are registered to the size of $24 \times 256 \times 256$. The training/validation/test split ratio is 3:1:1. Four-fold cross-validation is performed on this dataset.

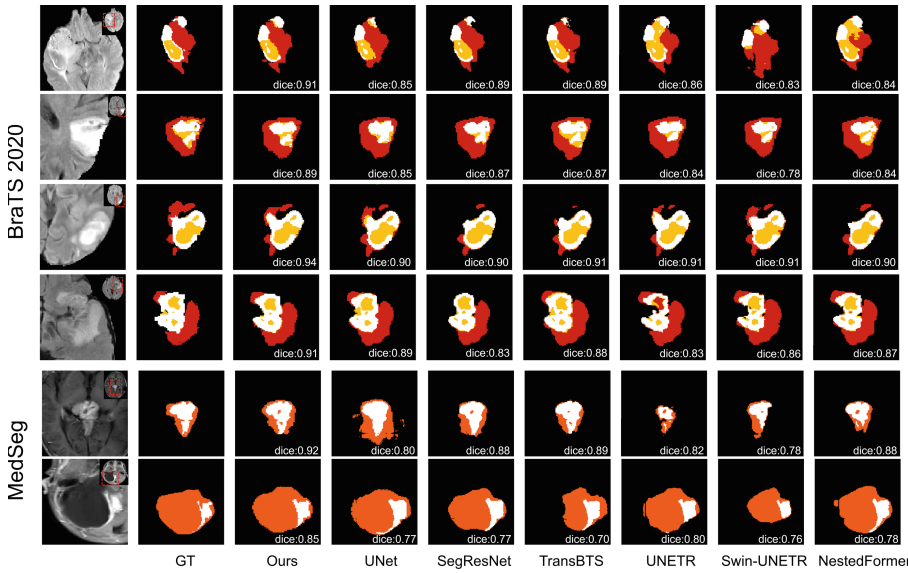

Fig. 2. Qualitative Comparison of segmentation results on BraTS and MedSeg. The red region represents WT, the yellow means TC and the white denotes ET (Color figure online).

3.2 Implementation Details

We implement EoFormer in Pytorch 1.11. Our model is trained from scratch for 300 epochs using two NVIDIA GTX 3090 GPUs. We select a combination of soft dice loss and cross-entropy as the loss function and utilize the AdamW optimizer [13] with a weight decay of 1×10^{-5}. The initial learning rate is 2×10^{-4}. For data augmentation, we apply image croping, flipping, identity scaling and shifting.

3.3 Results

We compare EoFormer with seven methods, including CNN-based methods (3D-UNet [2], SegResNet [15] and nnUNet [8]) and Transformer-based methods (TransBTS [21], UNETR [7], Swin-UNETR [6], NestedFormer [23]). The results are reproduced on our data split.

Table 1 displays the performance comparison of EoFormer against other methods on the BraTS 2020 dataset. EoFormer achieves the highest Dice scores on TC, ET, and the average. In addition, EoFormer attains the best HD95 scores on TC, ET, and the average. HD95 measures the edge distance between prediction and annotation, which is more sensitive to boundaries. Table 2 illustrates the performance of EoFormer and other methods on MedSeg. EoFormer outperforms the second-ranked NestedFormer by an average of 1.59% on Dice and achieves the top performance for both WT and ET on HD95. Furthermore, compared to the second-ranked SegResNet, EoFormer demonstrates an

Table 3. Ablation study on the encoder and decoder design.

Index	Encoder/Decoder	Dice (%) ↑				HD95 (mm) ↓			
		WT	TC	ET	Ave	WT	TC	ET	Ave
Enc1	UNet encoder	83.29	83.43	79.01	84.00	6.232	8.018	6.697	6.983
Enc2	CF×4	88.51	83.42	82.56	84.83	7.131	8.772	5.641	7.181
Enc3	CF×2+TF×2	**90.92**	**86.63**	81.05	86.20	**3.906**	5.227	5.637	4.923
Enc4	CF×2+ETF×2	90.84	86.38	83.22	**86.81**	3.974	**4.500**	3.432	**3.968**
Enc5	ETF×4	90.24	84.15	**85.40**	86.59	3.951	5.651	**3.405**	4.336
Dec1	UNet decoder	89.75	84.09	80.12	84.66	7.562	7.332	6.427	7.107
Dec2	IF×3	90.27	86.07	80.09	85.48	5.448	6.069	4.929	5.482
Dec3	CF×3	90.63	85.63	81.61	85.96	4.098	**4.467**	**3.023**	4.842
Dec4	EoF×3	**90.84**	**86.38**	**83.22**	**86.81**	**3.974**	4.500	3.432	**3.968**

average HD95 improvement of 1.57 mm, highlighting its superior performance in tumor boundary prediction. It is worth mentioning that EoFormer has the lowest FLOPs and limited model size. Fig. 2 represents the segmentation results on the BraTS 2020 and MedSeg datasets. The visualisation demonstrates that the EoFormer achieves the closest segmentation results to the ground truth. Specifically, EoFormer accurately segments both TC and ET region boundaries.

3.4 Ablation

We evaluate the effectiveness of our proposed EoFormer framework by conducting ablation experiments on the BraTS 2020. In Table 3, the abbreviations IF, CF, TF, ETF, and EoF represent IdentityFormer blocks (see Fig. 1(b)), ConvFormer blocks, TransFormer blocks, Efficient TransFormer blocks, and EoFormer blocks, respectively.

Encoder and Bottleneck Design. We compare our proposed EHE with different encoders and bottleneck in Table 3. Enc1 utilizes UNet encoder. Enc2 - 5 have the same encoder and bottleneck as EHE but with different configurations. Our results show that EHE outperforms other methods, with high average Dice and low average HD95. This is because a full CNN encoder (Enc2) is not good at capturing global dependencies, while a full Transformer encoder (Enc5) is inadequate at capturing low-level features. Our proposed EHE balances the strengths of both and achieves the best segmentation performance. Additionally, our extended efficient attention achieves better performance compared with Enc3 because it has a better ability to capture the periphery of objects [17].

Decoder Design. We compare the performance of various decoders in Table 3. Dec1 - 4 share the same EHE encoder, but employ different decoders: Dec1 uses the UNet decoder, Dec2 has three IdentityFormer blocks, Dec3 replaces

the IdentityFormer blocks with ConvFormer blocks, and Dec4 is our proposed EoFormer decoder. Our results show that the EoFormer decoder achieves the highest Dice scores, and achieves the lowest average HD95 score due to the incorporation of the EoS and EoL modules within the EoFormer block.

4 Conclusion

In this paper, we propose the EoFormer, a novel approach for brain tumor segmentation. Our method comprises the Efficient Hybrid Encoder and the Edge-oriented Transformer Decoder. The encoder effectively extracts features from images by striking a balance between CNN and Transformer architectures. The decoder integrates the Sobel and Laplacian edge detection filters into our edge-oriented modules that enhance the extraction capability of edge and texture information. Besides, we introduce the efficient attention mechanism and the re-parameterization technology to improve the model efficiency. Our EoFormer outperforms other state-of-the-art methods on both BraTS 2020 and MedSeg. Our model is computationally efficient and can be readily applied to other 3D medical image segmentation tasks.

References

1. Chen, S., Ding, C., Tao, D.: Boundary-assisted region proposal networks for nucleus segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 279–288. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_27
2. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
3. Ding, H., Jiang, X., Liu, A.Q., Thalmann, N.M., Wang, G.: Boundary-aware feature propagation for scene segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6819–6829 (2019)
4. Ding, X., Guo, Y., Ding, G., Han, J.: ACNet: strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1911–1920 (2019)
5. Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J.: RepVGG: making VGG-style convnets great again. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13733–13742 (2021)
6. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, 27 September 2021, Revised Selected Papers, Part I, pp. 272–284. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08999-2_22
7. Hatamizadeh, A., et al.: UNETR: transformers for 3D medical image segmentation. In: Proceedings of the IEEE/CVF winter Conference on Applications of Computer Vision, pp. 574–584 (2022)

8. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
9. Karimi, D., Salcudean, S.E.: Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. IEEE Trans. Med. Imaging **39**(2), 499–513 (2019)
10. Larrazabal, A.J., Martínez, C., Dolz, J., Ferrante, E.: Orthogonal ensemble networks for biomedical image segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12903, pp. 594–603. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_56
11. Li, X., et al.: Improving semantic segmentation via decoupled body and edge supervision. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12362, pp. 435–452. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58520-4_26
12. Lin, L., et al.: BSDA-Net: a boundary shape and distance aware joint learning framework for segmenting and classifying OCTA images. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12908, pp. 65–75. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_7
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
14. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2014)
15. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018, Part II. LNCS, vol. 11384, pp. 311–320. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11726-9_28
16. Sharma, N., Aggarwal, L.M., et al.: Automated medical image segmentation techniques. J. Med. Phys. **35**(1), 3 (2010)
17. Shen, Z., Zhang, M., Zhao, H., Yi, S., Li, H.: Efficient attention: attention with linear complexities. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3531–3539 (2021)
18. Tang, C., Chen, H., Li, X., Li, J., Zhang, Z., Hu, X.: Look closer to segment better: boundary patch refinement for instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13926–13935 (2021)
19. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
20. Wang, J., Wei, L., Wang, L., Zhou, Q., Zhu, L., Qin, J.: Boundary-aware transformers for skin lesion segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 206–216. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_20
21. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., Li, J.: TransBTS: multimodal brain tumor segmentation using transformer. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 109–119. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_11

22. Xie, Y., Liao, H., Zhang, D., Chen, F.: Uncertainty-aware cascade network for ultrasound image segmentation with ambiguous boundary. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention-MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part IV, pp. 268–278. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16440-8_26
23. Xing, Z., Yu, L., Wan, L., Han, T., Zhu, L.: NestedFormer: nested modality-aware transformer for brain tumor segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention-MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part V, pp. 140–150. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_14
24. Yu, W., et al.: MetaFormer is actually what you need for vision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10819–10829 (2022)
25. Zou, N., Xiang, Z., Chen, Y., Chen, S., Qiao, C.: Boundary-aware CNN for semantic segmentation. IEEE Access **7**, 114520–114528 (2019)