EAI Endorsed Transactions
on e-Learning

Research Article EAI.EU

ARM-Net: Improved MRI brain tumor segmentation method based on attentional mechanism and residual module

Ming Hu[1,*]

[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, PR China

Abstract

INTRODUCTION: Accurate tumor segmentation is a prerequisite for reasonable diagnosis and treatment of brain cancer. Gliomas, a highly prevalent and life-threatening type of brain tumor, pose a challenge in segmentation due to the intricate nature of brain structures and unpredictable appearances on brain MRIs.
OBJECTIVES: Current methods for brain tumor segmentation mainly rely on deep convolutional neural networks, which suffer from significant loss of feature information during encoding and decoding and the inability to capture tumor contours in detail.
METHODS: To address these challenges, this study rethinks the network architecture for MRI brain tumor segmentation. It proposes ARM-Net: an improved method for MRI brain tumor segmentation based on attention mechanisms and residual modules. Firstly, inverted external attention and dilated attention are employed in the last two layers of the encoder to enable the network to interact with both lesion areas and global information, facilitating better interaction among the four modalities. Secondly, different numbers of Res-Paths are added between encoder's first two layers and the decoder's last two layers to effectively mitigate the semantic gap issues caused by traditional skip connections.
RESULTS: Experiments on the BraTS 2021 dataset demonstrate that ARM-Net outperforms other similar models in terms of segmentation performance.
CONCLUSION: The experiment showed that the ARM-Net model could segment the contour structure of the tumor better than other methods.

Keywords: Brain tumor segmentation, Encoder-decoder, Attention mechanism, Residual module

Received on 30 April 2024, accepted on 07 May 2025, published on DD MM YYYY

doi: 10.4108/eetel.4953

*Corresponding author. Email: hu@home.hpu.edu.cn

1. Introduction

The brain, one of the most critical and sensitive parts of the human body, requires precise and timely diagnosis and treatment of brain diseases to ensure patient survival. Among these diseases, the highest incidence and mortality are attributed to gliomas, which arise from cellular abnormalities in the brain and are classified into high-grade gliomas (HGG) and low-grade gliomas (LGG)[1]. HGG exhibits greater invasiveness and growth rates, resulting in a high mortality rate among patients, whereas LGG can be benign or malignant, with slower growth rates and a higher likelihood of patient recovery, though they may progress to malignancy later on[2]. Accurate segmentation of brain tumors and surrounding tissues, such as edema, enhancing tumors, non-enhancing tumors, and necrotic areas, is crucial for diagnosing patient conditions and devising subsequent treatment plans[3].

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique that distinctly visualizes soft tissue abnormalities, thereby being extensively utilized in clinical diagnosis and monitoring tumor progression[4]. MRI employs specific sequences tailored to the tissue characteristics to differentiate various tumors. According to

the imaging technique used, it can be categorized into four modalities: T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR)[5]. Each modality emphasizes different aspects of the images, and the integration of multiple modalities can achieve complementary features. For instance, T1-weighted images are employed to distinguish healthy brain tissue, T1ce images facilitate the delineation of tumor boundaries, T2-weighted images are effective in detecting edema surrounding the tumor, and FLAIR images can differentiate edematous regions from cerebrospinal fluid.

Accurately segmenting lesion areas from multi-modal MRI brain tumors is crucial for clinical diagnosis and postoperative treatment of patients[6]. Due to the uncertainty of tumor location, size, and shape, as well as the issues of fuzzy boundaries and low imaging contrast between different tumor categories, distinguishing tumor areas from normal tissue is challenging, rendering MRI brain tumor segmentation a daunting task[7]. Previously, most brain tumor segmentation was manually conducted by physicians, leading to inter-rater variability and substantial time consumption, especially when dealing with large datasets. Hence, an automated segmentation approach is needed to assist physicians in MRI brain tumor segmentation.

The advancement of deep learning has brought about new opportunities for classification, segmentation, detection, and recognition tasks. With the evolution of convolutional neural networks, the Fully Convolutional Network (FCN)[8] and UNet[9] have emerged as the most commonly used and reliable methods in medical image segmentation. Building upon these models, numerous scholars have devised various segmentation models based on the encoder-decoder architecture to address the effectiveness of semantic segmentation and the diverse requirements of different segmentation targets. Zhou et al.[10] proposed the UNet++ network, an extension of UNet, which utilizes ensembles of UNet at different depths to determine the optimal network. Additionally, it employs deep supervision to synchronize encoders, facilitating collaborative learning. Despite achieving satisfactory results, UNet 2+ fails to adequately explore global feature information, leaving significant room for improvement. Addressing this limitation, Huang et al.[11] introduced UNet 3+, which incorporates full-scale skip connections and deep supervision strategies. This model effectively combines low-level details and high-level semantic information from different-sized receptive fields and aggregates feature maps across various levels through deep supervision, thereby enhancing segmentation accuracy across

Significant progress has been achieved in MRI brain tumor segmentation based on deep learning. Jiang et al.[12] proposed a cascaded segmentation approach, where the cascaded model segments brain tumor structures from coarse to fine, enabling end-to-end network training. Myronenko et al.[13] introduced an asymmetric network based on an encoder-decoder structure. Initially, large-scale encoders capture image features, followed by small-scale decoders reconstructing segmentation masks for precise lesion area delineation. Allah et al.[14] presented an Edge-UNet model merging boundary-related MRI data with brain MRI data, accurately extracting lesion features during encoding and combining raw MRI data of different sizes with contextual information during decoding, notably enhancing tumor segmentation performance. Cao et al.[15] proposed an MBANet convolutional neural network with a three-dimensional multi-branch attention mechanism. The model employs optimized "shuffle" units to form U modules, segmenting inputs along channels and performing group convolution operations, followed by a novel multi-branch 3D attention module SA extracting relevant information in the encoder. Zhu et al.[16] introduced a multimodal MRI brain tumor segmentation method integrating deep semantic and edge information fusion, comprising semantic segmentation, edge detection, and feature fusion modules. Swin former is utilized in the semantic segmentation module to extract feature information alongside a designed edge detection module and edge spatial attention module for fusing extracted semantic and edge features effectively across different stages.

While the aforementioned methods have shown promising results on brain tumor datasets, they also encounter certain limitations. These include subpar segmentation performance due to a pursuit of lightweight models or overly simplistic architectures that fail to achieve precise lesion delineation in the presence of varying levels of complexity across cases. Addressing these issues, this paper proposes a novel network architecture, termed ARM-Net, by refining attention mechanisms and residual modules atop an encoder-decoder framework. The specific research contributions of this study are outlined as follows:

(i) In this study, inverted external attention modules (IEA) and dilated gated attention modules (DGA) are applied in the last two layers of the encoder. The coordinated utilization of these two attention mechanisms effectively exploits global and local features of feature maps, learns to describe lesion characteristics from the dataset, integrates information from different scales, and efficiently delineates the contours of different tumor categories while reducing computational overhead.

(ii) Different numbers of Res-Path modules are sequentially introduced in the encoder's first two layers and the decoder's last two layers to address the semantic gap between shallow encoders and deep decoders in traditional skip connections, facilitating more efficient fusion of low-level and high-level information.

(iii) ARM-Net, a novel MRI brain tumor segmentation approach, is proposed in this paper, improving attention mechanisms and residual modules. Unlike other methods, this model precisely segments MRI brain tumors by enhancing attention mechanisms and residual modules.

(iv) Experimental validation conducted on the 2019 Brain Tumor Segmentation (BraTS) Challenge dataset demonstrates that ARM-Net outperforms previous improvements in U-shaped network segmentation, resulting in enhanced performance.

2. Method

This section describes the proposed approaches in detail, including an inverted External attention module (IEA) and an inflation-gated attention module (DGA), as well as a residual module (Res-Path) to mitigate the semantic gap.

2.1. ARM-Net network architecture

The proposed ARM-Net consists of an encoder, bottleneck layer, decoder, and skip connections. The encoder comprises four layers, with the first two layers employing two 3×3 convolutions for feature extraction, while the latter two layers integrate inverted external attention modules and dilated gated attention modules. The inverted external attention maximizes the utilization of multi-stage and multi-scale information, characterizing not only the overall dataset but also enhancing inter-modality relationships within samples. Meanwhile, the dilated gated attention simultaneously captures global and local information, where the global context aids in understanding the relationship between overall lesion structures and backgrounds for more precise lesion localization, while the local context facilitates capturing edge and corner details of lesion regions for more comprehensive predictions. The synergistic use of these attention modules extracts feature information from regions of interes[illegible] bottleneck layer comprises traditional dual convo[illegible]ion modules, as this stage's feature extraction contains abun[illegible]nt target information, and the attention-based output requi[illegible] traditional dual convolutions to narrow do[illegible]rget are[illegible] eliminating redundant information. [illegible]bsequ[illegible]y, skip connections are employed, with dif[illegible]ent numb[illegible] of Res-Path modules introduced in the first [illegible] layers [illegible] the semantic gap between deep en[illegible]ers a[illegible] sh[illegible]ow decoders, alleviating information di[illegible]arity and [illegible]serving more detailed information du[illegible] feature map [illegible]toration for effective fusion of feat[illegible]es. A[illegible]itionally, Res-[illegible] modules are not inserted between the [illegible]t two layers of skip connections, as th[illegible]atter two layers [illegible]he encoder utilize IEA and DGA mo[illegible]es, whose combined [illegible]ntion mechanisms effectively [illegible]itigate i[illegible]rmation disparity between the encoder a[illegible] decoder[illegible]e overall architecture of ARM-Net is depicted in F[illegible]re

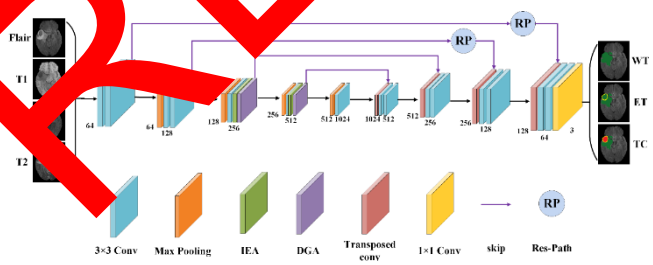

Figure 1. ARM-Net architecture diagram

2.2. Invert the external attention block

In the context of medical image segmentation, due to the inconsistent size, shape, and location of lesion areas, it is necessary to employ multi-scale attention to capture feature information from different locations. To address thi[illegible] an IEA module is proposed in this paper, which u[illegible]zes e[illegible]al attention to extract useful feature [illegible]ormation a[illegible] significantly enhances feature interacti[illegible]among differen[illegible] modalities. As depicted in Figure 2, the IE[illegible]odule utilizes two memory units to represent fea[illegible]e informa[illegible]among th[illegible] four modalities. Each meme[illegible] unit consists[illegible] a [illegible] convolution with shared para[illegible]eters, with[illegible]e channe[illegible]ber set to 128. The feature i[illegible]s unde[illegible] feature extraction through two conse[illegible]ve [illegible] units f[illegible]owed by convolution operato[illegible] This proce[illegible]maps the [illegible]ture maps to a high-dimensio[illegible]pace to compr[illegible]ensi[illegible]y describe the overall featur[illegible]nfor[illegible]ion among di[illegible]t modalities. The output features after con[illegible]olution operators become C×HW in dimensionality, which are [illegible]n reshaped to the same size as the [illegible]put [illegible]eatures. Subseq[illegible]tly, they are multiplied po[illegible]wise with the input features before reshaping, ensuring th[illegible] detailed inform[illegible]ion is not lost. Finally, feature mapping is [illegible]rformed with th[illegible]riginal input to enhance texture, edges, and[illegible]her detail info[illegible]nation.

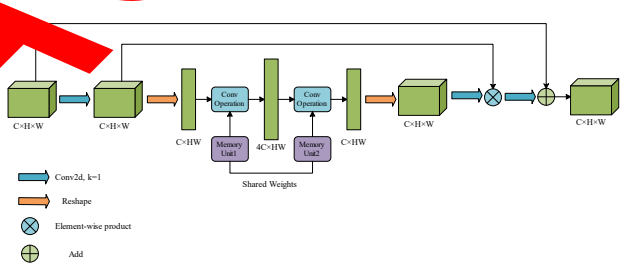

Figure 2. IEA module

2.3. Dilated gated attention block

After passing through the IEA module, numerous irrelevant pieces of information are filtered out, leaving behind feature information distributed across different locations. Subsequently, multi-scale attention is required to extract lesion features from various sizes of receptive fields. To this end, the present study employs the DGA module[17], which can simultaneously capture local and global information. From a local perspective, it extracts edge and corner details of lesion features, while from a global perspective, it comprehends the relationship between the overall lesion structure and background, enabling more accurate localization of lesion areas and more complete predictions.

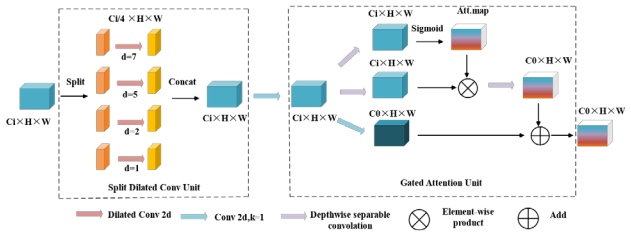

Figure 3. DGA module

As depicted in Figure 3, DGA is comprised of Separable Dilated Convolution (SDC) units and Gated Attention (GA) units. Specifically, the SDC divides the input feature map along the channel dimension into four parts, each of which undergoes depth-wise separable convolutions with varying dilation rates to acquire global and local feature information. Subsequently, the outputs are merged along the channel dimension to restore the size of the input feature map, followed by convolutional operations to facilitate interaction between local and global information. The output of the SDC units serves as input features to the GA unit, which focuses on the regions of interest to further extract lesion features. In the context of GA, the input feature map initially undergoes depth-wise separable convolutions to suppress irrelevant feature information generated by the SDC units and enhance crucial lesion features. Subsequently, residual mapping is employed along with other parallel modules to facilitate feature interaction and obtain the output features. The computational formula for the DGA module is as follows:

$$x_1, x_2, x_3, x_4 = Chunk_4(X) \quad (1)$$

$$x_1', x_2', x_3', x_4' = W_1(x_1), W_2(x_2), W_5(x_3), W_{\_4}) \quad (2)$$

$$T = W\ (Concat(x_1', x_2', x_3', x_4') \quad (3)$$

$$Att = \sigma(DW(T)) \quad (4)$$

$$Y = DW(DW(T) \otimes A\_) + W(T) \quad (5)$$

Here, $Chunk_4$ denotes the input feature map $X \in R^{C\times H\times W}$ divided into four parts along the channel dimension, where W_i represents the depth-wise separable convolution with dilation rate i. *Concat* indicates concatenation operation, W represents regular convolution operation, σ denotes the sigmoid activation function, DW represents depth-wise separable convolution, and $\otimes$ signifies element-wise multiplication. The output feature map $Y \in R^{C\times H\times W}$.

2.4. Semantic gap between encoder and decoder

The skip connections in the UNet network effectively propagate spatial information lost during pooling in the encoder to every layer of the decoder. While traditional skip connections preserve some spatial information, directly concatenating shallow encoders with deep decoders introduces a semantic gap issue to some extent. For instance, the skip connection of the first layer involves connecting the encoder before pooling with the decoder after transposed convolution. Since the encoder at this point is the network's initial layer, it extracts lower-level information. In contrast, the connected decoder is the network's final layer, extracting higher-level feature information. Consequently, there is a mismatch of information between the two, and fusing these significantly different features can lead to unnecessary impacts on semantic segmentation tasks.

To mitigate the semantic gap between the lower-level semantic information extracted by shallow encoders and the higher-level semantic information extracted by deep decoders, this study integrates Res-Path [28] with four residual modules between the first layer skip connections, as illustrated in Figure 4. The convolution layers in the module employ 3×3 convolutions, while the residual connections utilize 1×1 convolutions. Additionally, three residual modules are incorporated into the second layer skip connections, which exhibit significant information disparity. This integration aims to further process the feature information extracted during the encoder stage before channel concatenation with the corresponding decoder layer. Moreover, the residual structure employed in Res-Path simplifies feature learning. Res-Path is not used in the third and fourth layers of the skip connection in this paper, because the third and fourth layers of the encoder use the IEA and DGA modules, respectively, and the combination of the two attention mechanisms can effectively reduce the semantic gap with the same layer decoder.

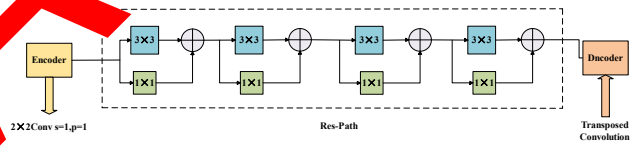

Figure 4. res-path module

3. Experiment and result

3.1. Data sets and preprocessing methods

The experimental setup of this study employs the BraTs 2019 dataset, comprising both training and testing sets. However, the testing set lacks labels, thus only the training set is utilized. The training set comprises a total of 335 3D MRI images, consisting of 259 HGG and 76 LGG images. In this paper, the training set is divided according to 8:2, where each sample contains four modes T1, T1CE, T2, and FLAIR, and the sample size is 240 × 240 × 155.

The BraTs 2019 dataset comprises four tissue labels: (1) necrotic core and non-enhancing tumor on T2-weighted images, (2) peritumoral edema, (3) background, and (4) enhancing tumor (ET) on T1CE images. To evaluate the results, a combination of labels 4 and 1 was used as the tumor core (TC), and a combination of labels 1, 2, and 4 was used as the entire tumor (WT).

Due to the large memory footprint, extended training cycles, and higher hardware demands associated with 3D

data, this study preprocesses the data into 2D images. Initially, the 3D data is sliced along the axial axis, resulting in 155 slices per sample, each sized at 240 × 240 × 1. Subsequently, blank slices devoid of lesion regions are discarded, followed by cropping the remaining slices to a size of 160 × 160 × 155.

3.2. Loss function

The overall performance of segmentation models depends not only on the network architecture but also on the choice of loss functions. Given the issue of class imbalance in brain tumor segmentation, conventional loss functions used in segmentation may not be suitable for training brain tumor models. Hence, this study employs the Generalized Dice Loss (GDL)[19] and the multi-class cross-entropy loss function[20], which adaptively weigh different classes to balance them and expedite convergence. The formulas for calculating these two loss functions are as follows:

$$L_{overall} = L_{GDL}(G,P) + \lambda \times L_{CE}(G,P) \quad (6)$$

The parameter λ is set to 1.25. The formula for computing the GDL is as follows:

$$L_{GDL}(G,P) = 1 - 2\frac{\sum_{j=1}^{C}(W_j \times \sum_{i=1}^{N}(g_{ij} \times p_{ij})) + \beta}{\sum_{j=1}^{C}(W_j \times \sum_{i=1}^{N}(g_{ij} + p_{ij})) + \beta} \quad (7)$$

In the formula, β represents the regularization constant, and W_j denotes the adaptive weight for the *j-th* class. The formula for calculating the multi-class cross-entropy loss function is as follows:

$$L_{CE}(G,P) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C}(g_{ij} \times Log(p_{ij})) \quad (8)$$

3.3. Evaluation index

This study employs the Dice similarity coefficient (DSC)[21] and the Hausdorff distance (HD)[22] to evaluate the segmentation results. The calculation formulas are as follows:

$$DSC = \frac{2TP}{2TP+FP+FN} \quad (9)$$

$$HD(X,Y) = max\{sup_{x \in X} D(x,Y), sup_{y \in Y} D(X,y)\} \quad (10)$$

Where DSC is used to compute the similarity between the segmentation result and the ground truth label, where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. HD is used to measure the distance between two subsets in space, and D is the Euclidean distance between point $x \in X$ and subset Y.

3.4. Ablation experiment

This study conducts ablation experiments on the BraTs 2019 dataset to validate the efficacy of the proposed method. To assess the performance of the proposed modules, IEA, DGA, and Res-Path were individually incorporated into the baseline model UNet for experimentation. The specific configurations were as follows:

(i) Baseline model UNet: adopts a U-shaped structure consisting of an encoder and a decoder.
(ii) UNet+IEA+DGA: IEA and DGA modules are sequentially added after the last two layers of the encoder in the baseline model UNet.
(iii) UNet+Res-Path: Different numbers of Res-Path modules are added to the first two layers of the skip connections in the baseline model UNet.
(iv) ARM-Net: IEA and DGA modules are sequentially added after the last two layers of the encoder in the baseline model UNet, and Res-Path modules with varying numbers of residuals are added to the first two layers of the skip connections.

The performance evaluation of this experiment on the BraTs 2019 dataset utilized the DSC and HD evaluation metrics, as shown in Table 1. Results from the ablation study indicate performance improvements when sequentially incorporating IEA and DGA modules into the baseline UNet model and further enhancements with the addition of Res-Path modules. Thus, it is evident that the IEA, DGA, and Res-Path modules contribute to enhancing the segmentation performance of MRI brain tumor images.

Table 1. Ablation experiments on DSC and HD indices

Method	DSC			HD		
	ET	WT	TC	ET	WT	TC
*	0.780	0.847	0.856	2.780	2.594	1.631
UNet+IEA+DGA	0.791	0.842	0.864	2.752	2.577	1.623
UNet+Res-Path	0.789	0.850	0.861	2.749	2.571	1.627
ARM-Net	**0.801**	**0.853**	**0.870**	**2.741**	**2.567**	**1.612**

Bold font is the optimal value

3.5. Comparison with other methods

The experiment was conducted on the BraTs 2019 dataset, comparing our proposed method, ARM-Net, with models such as UNet, DeepResNet, UNet2+, and UNet3+ in terms of DSC and HD evaluation metrics. Table 2 presents the experimental results.

Table 2. Compares the performance of different methods

Network	DSC			HD		
	ET	WT	TC	ET	WT	TC
UNet[9]	0.780	0.847	0.856	2.780	2.594	1.631
DeepResUNet [23]	0.797	0.821	0.867	2.881	2.604	**1.601**
UNet2+[10]	0.784	0.856	0.849	2.833	2.635	1.660
UNet3+[11]	0.757	**0.861**	0.815	2.754	**2.534**	1.651

Ours ARM-Net	**0.801**	0.853	**0.870**	**2.741**	2.567	1.612

Bold font is the optimal value

Upon observing Table 2, it is evident that the proposed method, compared to the baseline model UNet, demonstrates improvements in both DSC and HD metrics. When compared to other state-of-the-art models, ARM-Net achieves the highest scores in the ET and TC regions for the DSC metric, with scores of 80.1% and 87.0%, respectively. Additionally, there is an improvement in the ET region for the HD metric, with an optimal score of 2.741. These results indicate the successful segmentation of all tumor subregions by the proposed model.

Figure 5 depicts the segmentation results of ARM-Net compared to other state-of-the-art methods. Upon comparison, it is observed that the proposed approach exhibits strong performance in edge segmentation across different categories, enabling precise segmentation of MRI brain tumors.

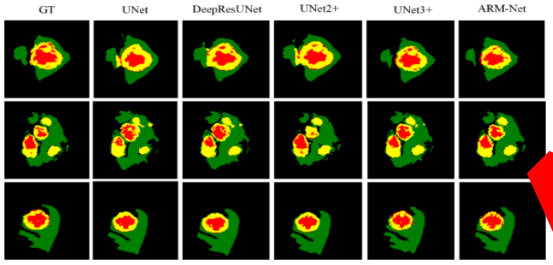

Figure 5. Visualization of results

4. Conclusion

This paper introduces a novel MRI brain tumor segmentation method, ARM-Net, which is enhanced by residual modules and attention mechanisms. Leveraging the learning feature extraction capabilities of deep neural networks, based on the UNet baseline model, traditional convolution modules are employed in the first two layers of the encoder to extract shallow-level feature information. With attention mechanisms adept at focusing on regions of interest and disregarding irrelevant areas, inverted external attention modules and channel gate attention modules are utilized in the last two layers of the encoder. This enables the network to interact with global information while focusing on local details and better-capturing interactions among the four modalities. Addressing the semantic gap in feature fusion between the shallow encoder and deep decoder, this paper introduces Res-Paths with varying residual block quantities in the first two layers of skip connections. The output features from the encoder layer are convolved through multiple residual blocks before being fused with the decoder layer's feature information, effectively reducing the semantic gap. Finally, comparative experiments are conducted on the BraTS 2019 dataset against other state-of-the-art methods, verifying that the proposed model outperforms in segmenting various subregions of brain tumors.

References

[1] KRISHNAPRIYA S, KARUNA Y. A survey of deep learning for MRI brain tumor segmentation methods: Trends, challenges, and future directions [J]. Health and Technology, 2023, 13(2): 181-201.

[2] MOHAMMED Y M, EL GAROUANI S, JELLOULI I. A survey of methods for brain tumor segmentation MRI images [J]. Journal of Computational Design and Engineering, 2023, 10(1): 266-93.

[3] FARAJZADEH N, SADEGHZADEH HASHEMZADEH M. Brain tumor segmentation and classification on MRI via deep hybrid representation learning [J]. Expert Systems with Applications, 2023, 224: 119963.

[4] LIU Z, TONG L, CHEN et al. Deep learning based brain tumor segmentation: a survey [J]. Complex & intelligent systems, 2023, 9(1): 1001-26.

[5] KUMAR A. Study and analysis of different segmentation methods for brain tumor MRI application [J]. Multimedia Tools and Applications, 2023, 82(5): 6911-39.

[6] FANG L, YANG Multi-input U model based on the integrated block and aggregation connection for MRI brain tumor segmentation [J]. Biomedical Signal Processing and Control, 2023, 79: 104027.

[7] AGGARWAL M, TIWARI A K, SARATHI M P, et al. An early detection and segmentation of Brain Tumor using Deep Neural Network [J]. BMC Medical Informatics and Decision Making, 2023, 23(1): 78.

[8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-40.

[9] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. 2015: 234-41.

[10] ZHOU Z, SIDDIQUEE M M R, TAJBAKHSH N, et al. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation [J]. IEEE transactions on medical imaging, 2019, 39(6): 1856-67.

[11] HUANG H, LIN L, TONG R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]//ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). 2020: 1055-9.

[12] JIANG Z, DING C, LIU M, et al. Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5. 2020: 231-41.

[13] MYRONENKO A. 3D MRI brain tumor segmentation using autoencoder regularization[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4. 2019: 311-20.

[14] ALLAH A M G, SARHAN A M, ELSHENNAWY N M. Edge U-Net: Brain tumor segmentation using MRI based on deep U-Net model with boundary information [J]. Expert Systems with Applications, 2023, 213: 118833.

[15] CAO Y, ZHOU W, ZANG M, et al. MBANet: A 3D convolutional neural network with multi-branch attention for

brain tumor segmentation from MRI images [J]. Biomedical Signal Processing and Control, 2023, 80: 104296.
[16] ZHU Z, HE X, QI G, et al. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI [J]. Information Fusion, 2023, 91: 376-87.
[17] RUAN J, XIANG S, XIE M, et al. MALUNet: A multi-attention and light-weight unet for skin lesion segmentation[C]//2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2022: 1150-6.
[18] IBTEHAZ N, RAHMAN M S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation [J]. Neural networks, 2020, 121: 74-87.
[19] HASHEMI N, MASOUDNIA S, NEJAD A, et al. A memory-efficient deep framework for multi-modal mri-based brain tumor segmentation[C]//2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2022: 3749-52.
[20] MAO A, MOHRI M, ZHONG Y. Cross-entropy loss functions: Theoretical analysis and applications[C]//International Conference on Machine Learning. 2023: 23803-28.
[21] NOORI M, BAHRI A, MOHAMMADI K. Attention-guided version of 2D UNet for automatic brain tumor segmentation[C]//2019 9th international conference on computer and knowledge engineering (ICCKE). 2019: 269-75.
[22] SOLTANI-GOL M, FATTAHI M, SOLTANIAN-ZADEH H, et al. DRAU-Net: Double Residual Attention Mechanism for automatic MRI brain tumor segmentation[C]//2022 30th International Conference on Electrical Engineering (ICEE). 2022: 587-91.
[23] LI H, JIANG G, ZHANG J, et al. Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images [J]. NeuroImage, 2018, 183: 650-65.