



OPEN Generative adversarial DacFormer network for MRI brain tumor segmentation

Muqing Zhang^{1,5}, Qiule Sun^{2,5}, Yutong Han¹, Mingli Zhang^{3,4}, Wei Wang¹ & Jianxin Zhang^{1✉}

Current brain tumor segmentation methods often utilize a U-Net architecture based on efficient convolutional neural networks. While effective, these architectures primarily model local dependencies, lacking the ability to capture global interactions like pure Transformer. However, using pure Transformer directly causes the network to lose local feature information. To address this limitation, we propose the Generative Adversarial Dilated Attention Convolutional Transformer (GDacFormer). GDacFormer enhances interactions between tumor regions while balancing global and local information through the integration of adversarial learning with an improved transformer module. Specifically, GDacFormer leverages a generative adversarial segmentation network to learn richer and more detailed features. It integrates a novel Transformer module, DacFormer, featuring multi-scale dilated attention and a next convolution block. This module, embedded within the generator, aggregates semantic multi-scale information, efficiently reduces the redundancy in the self-attention mechanism, and enhances local feature representations, thus refining the brain tumor segmentation results. GDacFormer achieves Dice values for whole tumor, core tumor, and enhancing tumor segmentation of 90.9%/90.8%/93.7%, 84.6%/85.7%/93.5%, and 77.9%/79.3%/86.3% on BraTS2019-2021 datasets. Extensive evaluations demonstrate the effectiveness and competitiveness of GDacFormer. The code for GDacFormer will be made publicly available at <https://github.com/MuqinZ/GDacFormer>.

Keywords Brain tumor segmentation, U-Net, Generative adversarial, Transformer

A brain tumor refers to an abnormal cells within the brain, which disrupts both its structural and functional integrity. MRI is a widely used non-invasive method for diagnosing and treating brain tumor¹. Accurate and reliable automatic segmentation of brain tumor is of critical clinical importance. Recently, with the rapid advancement of deep learning in medical analysis, methods based on deep learning have become the predominant approach for brain tumor segmentation².

The U-Net³, featuring a simple encoder-decoder structure with lateral connections, is the most renowned method for enhanced feature extraction and superior segmentation results. However, the limited availability of medical image datasets with high-quality annotations substantially constrains the capacity of networks to achieve optimal performance through training. To further enhance segmentation performance, researchers have integrated U-Net with other deep learning techniques, such as Generative Adversarial Networks (GAN)⁴, which have proven effective in various medical segmentation tasks^{5–7}. Using the GAN architecture, Xue et al.⁸ employ U-Net as a generator to achieve dynamic equilibrium in brain tumor segmentation, yielding commendable results. Building on this, Nema et al.⁹ proposed RescueNet, based on CycleGAN, using multiple generators and discriminators to ensure the generated image corresponds to the original. However, existing methods need further improvement to enhance the generator's feature extraction capability. To improve the generator, one can draw on a variety of advanced modules introduced in the U-Net architecture, including dilated convolution¹⁰, separable convolution¹¹, and the attention mechanism¹². Especially, to capture the overall voxel information of brain tumor, researchers incorporate Transformers¹³, which excel in establishing long-range dependencies in visual images for segmentation tasks^{14–16}. As a pioneering work, Wang et al.¹⁷ propose TransBTS by embedding Vision Transformer (ViT) into the bottleneck of U-Net, proving the effectiveness of Transformers for brain segmentation. Additionally, Roy et al.¹⁵ develop MedNeXt by embedding a Transformer into ConvNeXt,

¹School of Computer Science and Engineering, Dalian Minzu University, Dalian 116600, China. ²School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China. ³Montreal Neurological Institute, Quebec 116024, Canada. ⁴School of Computer Science and Technology, Shandong Technology and Business University, Yantai 264005, China. ⁵Muqing Zhang and Qiule Sun these have contributed equally to this work. ✉email: jxzhang0411@163.com

validating its effectiveness on organ, kidney, and brain tumor segmentation tasks. Chen et al.¹⁸ propose a dual-pathway Transformer module integrated into the bottleneck layer of U-Net to capture long-range interactions and global spatial dependencies. These methods described above direct integration of Transformer blocks into convolutional architectures, this way frequently induces notable disjointedness between global and local feature representations. Consequently, existing methods still need improvement in balancing global and local feature extraction for brain tumor segmentation.

In this work, we utilize the GAN architecture to build a robust brain tumor segmentation network. Such a design can solve the problem of limited accuracy improvement by training. Furthermore, to better balance global and local information of brain tumor, we introduce a new Transformer module, the Dilated Attention Convolution Transformer (DacFormer), into the generator. DacFormer comprises multi-scale dilated attention and a next convolutional block (NCB) with a Transformer construct. Leveraging these modules, we propose a novel generative adversarial transformer network (GDacFormer), as shown in Fig. 1. This network integrates adversarial learning and an improved Transformer to enhance segmentation performance. This combination captures both global and local information while using adversarial learning to enable the network to learn more effective information with limited training data. GDacFormer is extensively evaluated on the publicly available BraTS2019-2021 brain tumor segmentation datasets, and the results demonstrate its effectiveness. The main contributions are as follows:

- We propose GDacFormer, a novel approach that integrates adversarial learning and an improved transformer module for MRI brain tumor segmentation. This network enhances the interaction of features between tumor regions while balancing global and local information, resulting in more accurate segmentation results.
- We develop an improved Transformer module called DacFormer by introducing NCB and MSDA to enhance the brain tumor segmentation network. NCB enhances local feature extraction and complements global interactions. MSDA captures multi-scale feature representation and models long-range dependency. They collectively contribute to the enhanced performance of the GDacFormer.
- GDacFormer is extensively evaluated on the BraTS2019-2021 brain tumor segmentation datasets and achieves highly competitive results compared to state-of-the-art methods.

Materials and method

Overall architecture of GDacFormer

The GDacFormer architecture, shown in Fig. 1, adopts a GAN framework with two main components: the generator and the discriminator. The generator, a cascaded U-Net, produces brain tumor images progressively from coarse to fine, evaluated by the discriminator.

The generator's cascaded U-Net starts with a 3D U-Net for coarse segmentation, followed by a fine segmentation stage using an improved 3D U-Net. Input size for the coarse segmentation is $5 \times 128 \times 128 \times 128$, outputting $3 \times 128 \times 128 \times 128$. It is then concatenated with the original input and passed through the fine segmentation network for further processing. Additionally, to capture both global and local features of brain tumor, GDacFormer embeds three DacFormer layers in the generator's bottleneck. In DacFormer, the multi-scale dilated attention module focuses on capturing contextual semantic dependencies at different scales, while the NCB enhances local feature extraction, thereby improving segmentation performance.

DacFormer layer

The DacFormer layer, shown in Fig. 2, enhances GDacFormer by balancing global and local feature extraction in MRI brain tumor segmentation. Embedded in the generator's bottleneck, it integrates advanced attention mechanisms and convolutional operations. Comprising the Next Convolution Block (NCB) and Multi-Scale Dilated Attention (MSDA) module, it captures long-range dependencies and local features, providing a comprehensive approach to feature extraction.

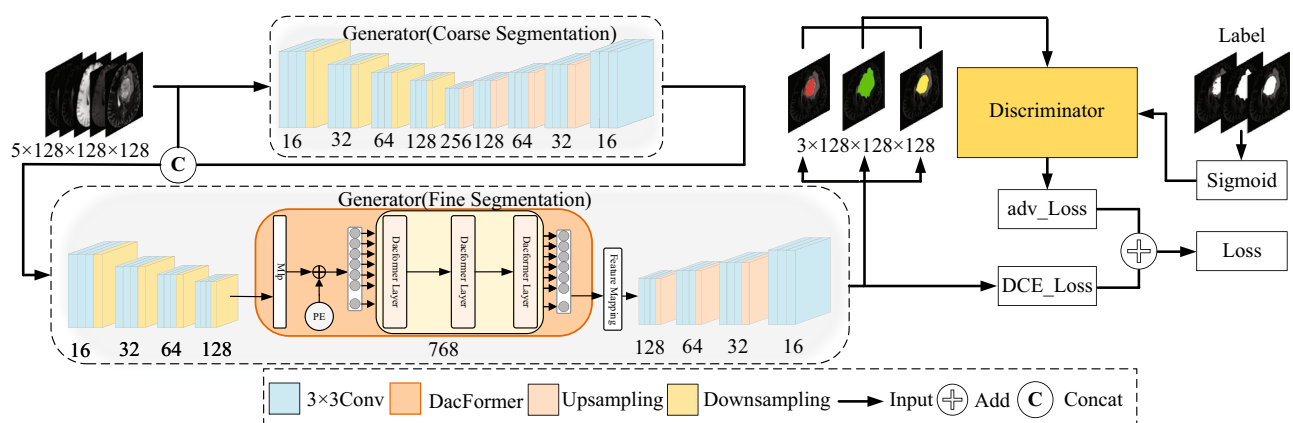


Fig. 1. The overall architecture of GDacFormer for brain tumor segmentation.

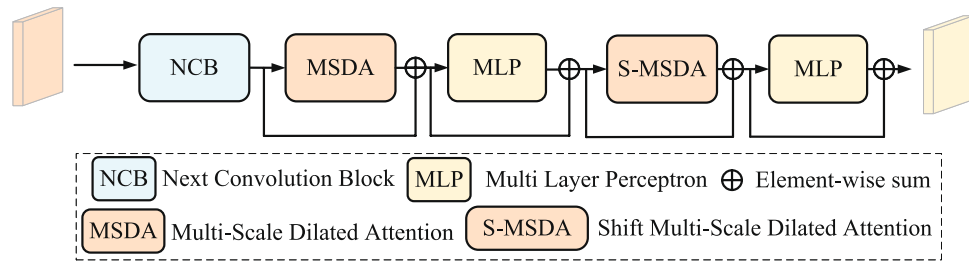


Fig. 2. The structure of DacFormer Layer.

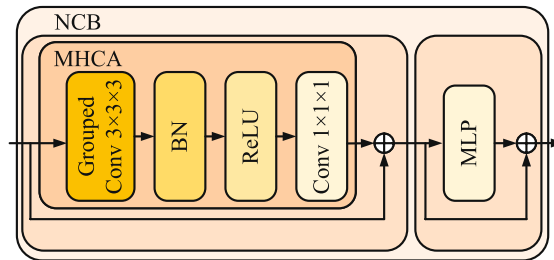


Fig. 3. The structure of Next Convolution Block (NCB).

As shown in Fig. 2, the input brain tumor image first passes through the NCB module, introduced in Sec. DacFormer layer and Fig. 4, where localized features are extracted using convolutional attention mechanisms. The processed image is then divided into 3D volumetric chunks, which are flattened and fed into the MSDA module, as shown in Fig. 4 of Sec. DacFormer layer. Here, multi-scale self-attention is applied to capture long-range dependencies and global details. The output from the MSDA module undergoes further transformation through a 2-layer MLP with GELU activation, introducing non-linear capabilities. The S-MSDA module enables the network to capture global features efficiently by enabling semantic information in different windows to interact to obtain cross-window links. Finally, the spatial information is reconstructed to form a 3D brain tumor image, integrating both detailed local features and broad contextual insights.

Next convolution block (NCB)

The NCB module, depicted in Fig. 3, retains the advantages of convolutional operations, such as capturing local features, while complementing the transformer's ability to model global interactions. The NCB includes a Multi-Head Convolutional Attention (MHCA) module, which serves as a token mixer, allowing information to be attended from different representational subspaces simultaneously (local and global). Eq. (1) detail the mathematical formulation of the NCB module, showing how the input is processed through MHCA and refined through a MLP with GELU activation. The formulations are as follow:

$$\begin{aligned}\tilde{X}^l &= \text{MHCA}(X^{l-1}) + X^{l-1}, \\ X^l &= \text{MLP}(\tilde{X}^l) + \tilde{X}^l\end{aligned}\quad (1)$$

where X^{l-1} denotes the input of $(l-1)$ th layer, and \tilde{X}^l and X^l are the outputs of MHCA and NCB.

Feature extraction is performed in the MHCA module through the Convolutional Attention (CA) module, which achieves efficient localized feature learning by means of a multi-head form in order to simultaneously attend to information from different representational subspaces at different locations, as shown in Eq. (2):

$$\text{MHCA}(X) = \text{Concat}(\text{CA}_1(X_1), \dots, \text{CA}_h(X_h))W. \quad (2)$$

Here, MHCA captures information from h parallel representation subspaces. $X = [X_1, X_2, \dots, X_h]$ denotes the partitioning of the input feature X into multiple heads in the channel dimension. W represents the projection layer equipped for MHCA to facilitate multi-head information interaction across. CA is the single-head convolutional attention, defined in Eq. (3):

$$\text{CA}(X) = \text{O}(W, (T_m, T_n)), \text{ where } T_{\{m,n\}} \in X, \quad (3)$$

where T_m and T_n are two tokens adjacent to the feature input X . O is an inner product operation with trainable parameters W and input tokens $T_{\{m,n\}}$. By independently computing attention maps on each subspace through CA, MHCA is able to capture diverse channel-specific dependencies. These multiple attention are then aggregated, enabling the model to integrate complementary information across subspaces, thus providing richer

and more discriminative feature. The MHCA is carried out with a group convolution (multi-head convolution), a point-wise convolution (conv 1x1x1 in Fig. 3), BatchNorm (BN) and ReLU are used for normalization and nonlinear activation. The ability of the DacFormer to extract local features is enhanced by adding the NCB module.

Multi-scale dilated attention (MSDA)

To leverage sparse tumor region information in brain tumor images, we construct the MSDA module to extract global, multi-scale fine semantic dependencies. The MSDA exploits the sparsity at different scales of the self-attention mechanism and models long-range dependencies using dilated attention (DA) in various feature subspaces. Besides self-attention computation, DA also increases the receptive field of the filters while maintaining the same number of parameters, making them efficient for capturing multi-scale features. Its structure is shown in Fig. 4. First, keys and values are sparsely selected in sliding windows centered on the query patch, and then self-attention is computed on the query patch. The detailed operation formula for a single DA is shown in Eq. (4):

$$X_{DA} = DA(Q, K, V, r), \quad (4)$$

where Q , K , and V represent the query, key, and value matrices, respectively, derived from the input feature, with each row of these matrices representing a query, key, or value vector. MSDA performs the self-attention computation within a sliding window centered on a region of size $w \times w \times w$ and w is set to 7. Additionally, the sparsity in the self-attention process is controlled by defining $r \in \mathbb{N}^+$. The computation of DA is given in Eq. (5):

$$\begin{aligned} x_{ijl} &= \text{Attention}(q_{ijl}, K_r, V_r), \\ &= \text{Softmax}\left(\frac{q_{ijl} K_r^T}{\sqrt{d_k}}\right) V_r, \\ 1 \leq i \leq W, 1 \leq j \leq H, 1 \leq l \leq D, \end{aligned} \quad (5)$$

where H , W , and D represent the height, width, and depth. k_r and V_r denote the keys and values selected from the feature maps K and V . d_k is the dimensionality of the keys. By sparsely selecting keys and values centered on queries, DA explicitly satisfies the properties of locality and sparsity, effectively and efficiently modeling long-range dependencies.

Similar to the NCB and vanilla vision transformer¹⁹, MSDA also employs multi-head DA, where each head performs DA (Eqs. 4–5) using vary dilation rates, it integrates multi-scale semantic information. We divide the channels of the feature map to n different heads and perform DA in each heads with different dilation rates. The MSDA is formulated in Eqs. (6):

$$\begin{aligned} h_i &= DA(Q_i, K_i, V_i, r_i), \quad 1 \leq i \leq n, \\ X_{MSDA} &= \text{Linear}(\text{Concat}[h_1, \dots, h_n]), \end{aligned} \quad (6)$$

where r_i is the expansion rate of the i -th head, and Q_i , K_i , and V_i are slices of the feature map fed into the i -th head. The outputs of all n heads $\{h_i\}_{i=1}^n$ are then concatenated and passed through a linear layer for feature aggregation. By this structural design, MSDA accurately combines semantic information at various scales within the receptive field and effectively reduces self-attention computation in background regions of brain tumor images.

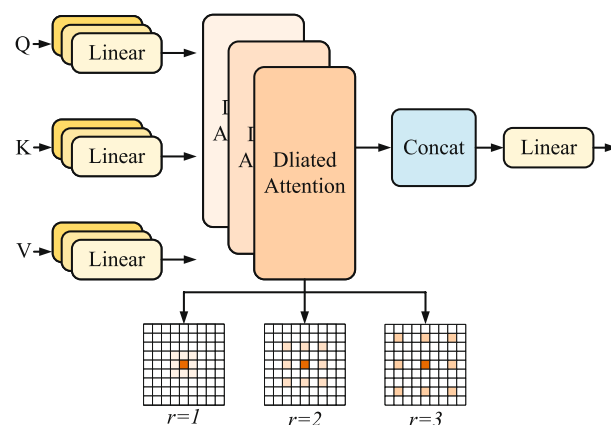


Fig. 4. The structure of Multi-Scale Dilated Attention (MSDA).

Discriminator

The discriminator, shown in Figure 5, consists of four four convolutional blocks, each consisting of a 3D convolutional layer with a kernel size of $1 \times 1 \times 1$, batch normalization (BN) layer and a LeakyReLU activation function. The $1 \times 1 \times 1$ convolution kernel acts as a fully connected layer to maintain the accuracy of the discriminator's judgment. The input to the discriminator is an image of size $3 \times 128 \times 128 \times 128$. The image generated by the generator module, along with the ground truth (GT), is fed into the discriminator.

Loss function

To supervise and optimize tumor segmentation results, we use a hybrid loss function comprising three parts: adversarial loss (L_{adv})⁴, Dice loss (L_{Dice})²⁰ and Cross-Entropy loss (L_{CE}). The adversarial loss ensures that the generated image is closer to the real image. Dice loss, frequently used alongside Cross-Entropy loss, is a common choice in medical image analysis to tackle class imbalance.

The hybrid loss is calculated as follows:

$$L_{adv}(\theta_G; \theta_C; y) = E_{(X, Y)} \sim y \left[- \sum_{a \in H} \sum_{b \in W} \{ (1 - \eta) \log(\psi(Y)[a, b]) + \eta \log(\psi(1 - \hat{Y})[a, b]) \} \right] \quad (7)$$

where y denotes the truth image, θ_G denotes the generated image, θ_c is used to discriminate the authenticity of the sample, \hat{y} denotes the segmentation result.

$$L_{Dice} = 1 - \frac{2 \sum_{l \in L} \sum_{i \in N} y_i^{(l)} \hat{y}_i^{(l)} + \varepsilon}{\sum_{l \in L} \sum_{i \in N} (y_i^{(l)} + \hat{y}_i^{(l)}) + \varepsilon}, \quad (8)$$

$$L_{CE} = - \sum_{i \in E} \sum_{l \in L} y_i^{(l)} \log \hat{y}_i^{(l)}, \quad (9)$$

where N is the set of all samples, L represents the set of all labels of the sample, $y_i^{(l)}$ is the one-hot coding (0 or 1) of the i th sample, labeled l , and $\hat{y}_i^{(l)}$ denotes the size of the predicted probability of the same sample i , labeled l . ε is the minima set to prevent the division by 0 condition from occurring in the computation and is set to 1×10^{-5} .

To maximize the convergence speed, the adversarial loss is multiplied by a coefficient α , set to 0.3 in this paper. The three components of the loss function are summed, and the complete loss function is expressed as:

$$Loss = \alpha \times L_{adv} + L_{Dice} + L_{CE} \quad (10)$$

With such a hybrid Loss Function design, L_{Dice} can effectively alleviate the common category imbalance problem in medical images by optimizing the spatial overlap between the predicted segmentation and the real labels, and enhance the segmentation effect on small targets such as ET. L_{CE} aims at pixel-level accuracy and provides stable and strong gradient information at the early stage of training, which helps the model to converge quickly and improve accuracy. L_{adv} is used to improve the structural realism and boundary coherence of the segmentation map output by the generator, which can further reduce the artifacts and unreasonable boundaries in the prediction map and make the segmentation results closer to GT.

Experiments

Experimental details

Experimental setup

The experiments are performed using the PyTorch deep learning framework, with the network trained for 1000 epochs, a batch size of 1, and the Adam optimizer. Exponential decay is applied to the learning rate, with an initial learning rate is 1×10^{-3} , and the weight decay coefficient is 1×10^{-4} . The hardware environment of the experiments is configured with 1 Intel(R) Xeon(R) Gold 5222 CPU @3.80GHZ and 1 Nvidia RTX3090 GPU graphics card (24GB).

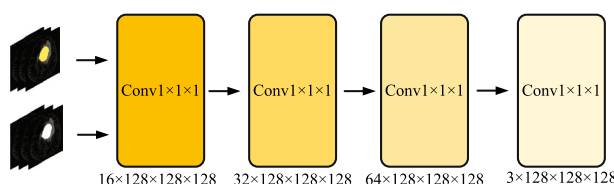


Fig. 5. The structure of the Discriminator.

Datasets and data processing

To thoroughly assess the performance of the proposed method on the MRI brain tumor segmentation task, we utilize the publicly available BraTS 2019–2021 brain tumor segmentation dataset^{21,22}. These datasets include both a training set and a validation set. The training set images are accompanied by ground truth (GT), manually annotated by a professional physician, and consist of four categories: healthy site (label 0), necrotic region (label 1), edematous region (label 2), and enhanced tumor (label 4). The GT labels of the validation set are not disclosed. The specific composition of the dataset is shown below: (1) BraTS 2019: 259 high-grade glioma(Hgg) cases and 76 low-grade glioma(Lgg) cases in the training set and 125 unknown tumor samples in the validation set. (2) BraTS 2020: 293 Hgg cases and 76 Lgg cases in the training set and 125 unknown tumor samples in the validation set. (3) BraTS 2021: 993 Hgg cases and 258 Lgg cases in the training set and 219 unknown tumor samples in the validation set.

As shown in Figure 6, each brain tumor data sample consists of images from four modalities: T1, T1ce, T2, and Flair. An image size of $240 \times 240 \times 155$ for each modality and all of them co-registered to a common anatomical template (SRI^{45}) and are resampled to $1mm^3$. In addition, the three regions used to evaluate MRI brain tumor segmentation are Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), where WT contains labels 1, 2, and 4, TC contains labels 1 and 4, and ET contains label 4. By uploading the segmented image results of the validation set to the CBICA online platform, we can receive the evaluated results.

Evaluation criterion

Model performance is assessed using Dice Similarity Coefficient (DSC) and Hausdorff distance, following^{15,17,23,24}. DSC measures the overlap between model segmentation and ground truth (GT) segmentation. The Dice Similarity Coefficient is calculated as in Eq. (11):

$$DSC = \frac{2TP}{FP + 2TP + FN} \quad (11)$$

where TP , FP , and FN represent the number of voxels correctly predicted to be tumor, the number of voxels incorrectly predicted to be tumor, and the number of voxels incorrectly predicted to be non-tumor, respectively.

Hausdorff distance is employed to measure the maximum distance between two image boundaries. The calculated results are generally multiplied by 95%, i.e., the formula is shown in Eq. (12).

$$HD(T, P) = \max \left\{ \sup_{i \in T} \inf_{p \in P} d(t, p), \sup_{p \in P} \inf_{i \in T} d(t, p) \right\} \quad (12)$$

where T and P represent the GT and prediction region, respectively, and t and p are points in the two regions. $d(t, p)$ denotes the distance function between point t and point p .

Results

Ablation study

To show the effectiveness of our network model, ablation experiments are conducted on the BraTS2020 validation set. The U-Net is used as the baseline approach. Based on the baseline network, GU-Net is constructed by incorporating GANs to verify the effectiveness of the adversarial learning approach. GU-Net+MHSA, GU-Net+MSDA and GDacFormer are used to evaluate the effectiveness of the MSDA module and NCB module in brain tumor segmentation. Where MHSA stands for multi-head self-attention in the ViT¹⁹.

Table 1 presents a detailed comparison of the performance of different models using the DSC and HD95 metrics on brain tumor segmentation tasks. Bold in all tables represents the highest value. Furthermore, the p -values presented in Table 1 are calculated using the Wilcoxon signed-rank test, comparing the GDacFormer with each ablated version of the model. Our main goal is to statistically demonstrate that GDacFormer performs significantly better than the versions marked with an asterisk. This provides evidence for the effectiveness of each component in our model. The baseline model, U-Net, achieves a mean DSC of 82.4% and a mean HD95 of 19.89 mm. Incorporating adversarial learning into the U-Net, resulting in the GU-Net, improves the mean

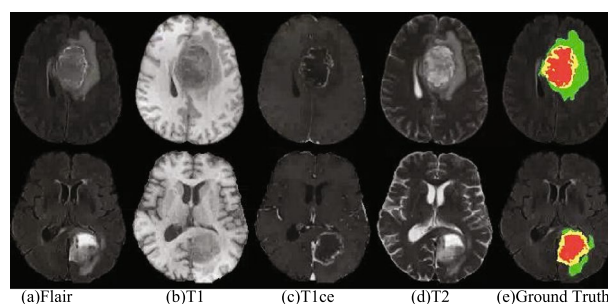


Fig. 6. MRI maps of brain tumor across four modalities, along with ground truth images. From left to right: Flair, T1, T1ce, T2 and Ground Truth. Each color corresponds to a tumor class (label): red (label 4) for necrosis and non-enhancing, green (label 2) for edema, and yellow (label 1) for enhancing tumor.

Methods	DSC ↑				HD95 ↓			
	ET	WT	TC	MEAN	ET	WT	TC	MEAN
U-Net (Baseline)	76.1*	89.7*	81.5*	82.4*	39.04*	8.02*	12.62*	19.89*
GU-Net	76.8*	90.6	82.6*	83.3*	27.41*	5.53	9.60*	14.18*
GU-Net+MHSA	78.4	90.3	83.9*	84.2	26.52*	6.02	8.93*	13.82*
GU-Net+MSDA	78.6	90.8	85.0	84.8	29.69*	4.61	5.98	13.43*
GDacFormer (Ours)	79.3	90.8	85.7	85.3	24.00	4.84	6.44	11.76

Table 1. Ablation study (%) on BraTS2020 Validation set. The ET, WT, and TC metrics are presented as the mean across all samples, while MEAN represents the average of ET, WT, and TC results. *Denotes significant difference compared with GTransU-Net by Wilcoxon signed-rank test (p value < 0.05).

Methods	DSC ↑				HD95 ↓			
	ET	WT	TC	MEAN	ET	WT	TC	MEAN
MBANet ²⁵	78.2	89.8	83.0	83.7	3.08	5.88	5.09	4.68
SMRAU-Net ²⁶	74.2	88.5	81.0	81.2	6.67	10.83	10.25	9.25
AMMGS ²³	76.8	89.3	81.1	82.4	5.18	8.22	7.23	6.88
SPA-Net ²⁷	77.1	89.6	83.3	83.3	—	—	—	—
TransBTs ¹⁷	78.9	90.0	81.9	83.6	3.73	5.64	6.05	5.14
IncompleteMriSeg ²⁸	78.3	90.3	83.1	83.9	5.55	4.82	6.93	5.76
GMetaNet ¹⁶	78.4	90.2	82.5	83.7	—	—	—	—
GAU-Net ²⁹	76.7	89.7	79.0	81.8	8.40	6.92	4.61	6.64
GVAT-Net ³⁰	75.2	89.9	81.5	82.2	12.56	7.39	8.06	9.34
GDacFormer (Ours)	77.9	90.9*	84.6*	84.5*	2.83*	4.42*	6.58	4.61*

Table 2. Comparison results of DSC (%) and HD95 (mm) metrics on the BraTS2019 Validation set. *Denotes significant difference compared with SOTA by Wilcoxon signed rank test (p value < 0.05).

DSC to 84.0% and reduces the mean HD95 to 14.26 mm, demonstrating the effectiveness of adversarial learning in enhancing segmentation accuracy and boundary alignment. Further to visualise the role of the MSDA module we introduce the traditional MHSA module in GU-Net and MSDA module in GU-Net respectively. These model achieves a mean DSC of 84.2%/84.8% and a mean HD95 of 13.82/13.43 mm. The results of GU-Net+MSDA model are better than GU-Net+MHSA model, benefiting from the long-term modeling capabilities and aggregates semantic multi-scale information of the MSDA, which enhances both the overlap and boundary metrics. Our proposed model, GDacFormer, which incorporates the NCB module in addition to the adversarial learning and MSDA, achieves the highest performance among all models. It records a mean DSC of 85.3% and a mean HD95 of 11.76 mm. This indicates that GDacFormer effectively balances global and local feature extraction, leading to superior segmentation accuracy and boundary precision. The improvements in both metrics highlight the robustness and competitiveness of GDacFormer in MRI brain tumor segmentation tasks, demonstrating its potential for clinical application.

Comparison with state-of-the-art

To better demonstrate the competitiveness of GDacFormer, we also compare it with the existing state-of-the-art brain tumor segmentation methods on the BraTS2019-2020 Validation sets and BraTS2021 Training set, and the comparative results on the three sets under DSC and HD95 metrics are reported in Tables 2, 3 and 4. In Tables 2, 3 and 4, the p -values were computed between GDacFormer and all compared methods based on the DSC and HD95. These values are used to highlight statistically significant improvements in each column of the results. This addition further supports the effectiveness of GDacFormer.

Table 2 showcases the performance of different models on the BraTS2019 validation set. GDacFormer demonstrates performance with a mean DSC of 84.5% and a mean HD95 of 4.61 mm. These results indicate that our model excels in both segmentation accuracy and boundary precision. Compared to CNN-based methods such as MBANet, SMRAU-Net, AMMGS, and SPA-Net, GDacFormer leads the DSC by 0.8%, 3.3%, 2.1% and 1.2%, showing clear improvements. This is due to the fact that GDacFormer uses both adversarial training and introduces global information compared to the above networks. Additionally, it outperforms Transformer-based approaches such as TransBTs, IncompleteMriSeg, and GMetaNet. GDacFormer leads the DSC by 0.9%, 0.6% and 0.8%. This is due to the fact that GDacFormer further enhances the ability of model global information to interact with local information. The performance is also better than GAN-based methods like GAU-Net(2.7%) and GVAT-Net(2.3%), highlighting the robustness and efficacy of GDacFormer in handling complex segmentation tasks. This is due to the fact that GDacFormer embeds the DacFormer layer on top of the adversarial training, which improves the model's ability to capture global information.

Methods	DSC \uparrow				HD95 \downarrow			
	ET	WT	TC	MEAN	ET	WT	TC	MEAN
dResU-Net ³¹	80.0	86.6	83.6	83.4	—	—	—	—
SMRAU-Net ²⁶	72.9	88.6	80.2	80.6	31.97	10.26	13.58	18.60
AMMGS ²³	78.0	88.3	81.7	82.7	23.62	7.17	7.99	12.93
SPA-Net ²⁷	77.8	90.0	83.2	83.7	—	—	—	—
Residual U-Net ³²	72.1	88.8	79.4	80.1	34.14	11.69	13.76	19.86
TransBTS ¹⁷	78.7	90.1	81.7	83.5	17.95	4.96	9.77	10.89
IncompleteMriSeg ²⁸	79.1	90.7	84.6	84.8	33.77	4.05	5.78	14.53
U-Net ¹⁴	71.2	88.3	75.9	78.5	34.46	8.18	10.63	17.76
MedNeXt ¹⁵	79.5	90.4	84.5	84.8	22.13	6.14	10.13	12.80
SRAU-Net ³³	81.9	88.6	74.3	81.6	—	—	—	—
GVAT-Net ³⁰	76.4	89.9	81.0	82.4	32.56	7.39	12.06	17.33
GATransU-Net ²⁴	70.8	90.3	81.5	80.9	37.58	4.91	7.49	16.66
GDacFormer (Ours)	79.3	90.8*	85.7*	85.3*	24.00	4.84*	6.44*	11.76*

Table 3. Comparison results of DSC (%) and HD95 (mm) metrics on the BraTS2020 Validation set. *Denotes significant difference compared with SOTA by Wilcoxon signed rank test (p value < 0.05).

Methods	DSC \uparrow				HD95 \downarrow			
	ET	WT	TC	MEAN	ET	WT	TC	MEAN
dResU-Net ³¹	84.0	86.0	82.2	84.1	—	—	—	—
nnU-Net ³⁴	84.8	92.8	88.7	88.8	13.07	6.78	8.79	9.55
MMEF-nnUNet ³⁵	87.3	92.8	90.1	90.1	10.09	5.10	9.68	8.29
SPA-Net ²⁷	83.5	90.7	88.9	87.7	22.07	6.10	9.21	12.46
Residual U-Net ³²	78.2	90.3	83.9	84.2	26.14	8.69	11.57	15.46
TransBTS ¹⁷	85.3	92.1	88.2	88.6	10.78	6.97	7.56	8.44
U-Net ¹⁴	86.2	92.5	91.8	90.2	11.28	7.74	7.85	8.96
nnFormer ³⁶	84.9	93.0	93.5	90.5	9.28	4.85	5.32	6.48
SSCFormer ³⁷	85.3	93.0	92.8	90.4	11.11	5.38	7.37	7.95
MedNeXt ¹⁵	89.4	93.3	91.5	91.4	8.32	7.83	4.79	6.98
SRAU-Net ³³	89.6	91.1	88.3	89.7	—	—	—	—
GVAT-Net ³⁰	81.4	90.8	85.4	85.9	21.83	5.37	8.56	11.92
GDacFormer (Ours)	86.3	93.7*	93.5*	91.2*	8.41	5.54	4.93	6.29*

Table 4. Comparison results of DSC (%) and HD95 (mm) metrics on the BraTS2021 Training set. *Denotes significant difference compared with SOTA by Wilcoxon signed rank test (p value < 0.05).

The evaluation results on the BraTS2020 validation set, as shown in Table 3, further validate the effectiveness of GDacFormer. The model achieves the highest mean DSC of 85.3% and a mean HD95 of 11.76 mm. Compared to CNN-based methods like dResU-Net, SMRAU-Net, AMMGS, SPA-Net, and Residual U-Net, our model leads the DSC by 1.9%, 4.7%, 2.6%, 1.4% and 5.2%. This is due to the fact that GDacFormer uses both adversarial training and introduces global information substantial enhancements in segmentation performance. GDacFormer also outperforms Transformer-based models such as TransBTS, IncompleteMriSeg, U-Net, and MedNeXt. GDacFormer leads the DSC by 1.8%, 0.5%, 6.8% and 0.5%. Demonstrates its superior ability to balance and integrate global context and local detail extraction. Furthermore, GDacFormer leads DSC the DSC by 3.7%, 2.9% and 4.4% than GAN-based methods such as SRAU-Net, GVAT-Net, and GATransU-Net, indicating that the integration of adversarial learning, MSDA, and NCB modules significantly increases segmentation accuracy and boundary precision.

Table 4 shows the results of the models on the BraTS2021 training set using 5-fold cross-validation. GDacFormer achieves an impressive mean DSC of 91.2% and a mean HD95 of 6.29 mm, reflecting excellent performance in segmentation tasks. This performance is notably higher than that of CNN-based methods like dResU-Net, nnU-Net, MMEF-nnUNet, SPA-Net and Residual U-Net. GDacFormer leads the DSC by 7.1%, 2.4%, 1.1%, 3.5% and 7.0%. Among transformer-based models, GDacFormer shows better DSC values than TransBTS, U-Net, nnFormer and SSCFormer by 2.6%, 1.0%, 0.7% and 0.8%. And outperforms MedNeXt in HD95 by 0.67mm. Furthermore, our model outperforms GAN-based methods such as SRAU-Net (by 1.5%) and GVAT-Net (by 5.3%), further demonstrating the effectiveness and robustness of GDacFormer in brain tumor segmentation. The experimental results demonstrate that GDacFormer exhibits strong competitiveness.

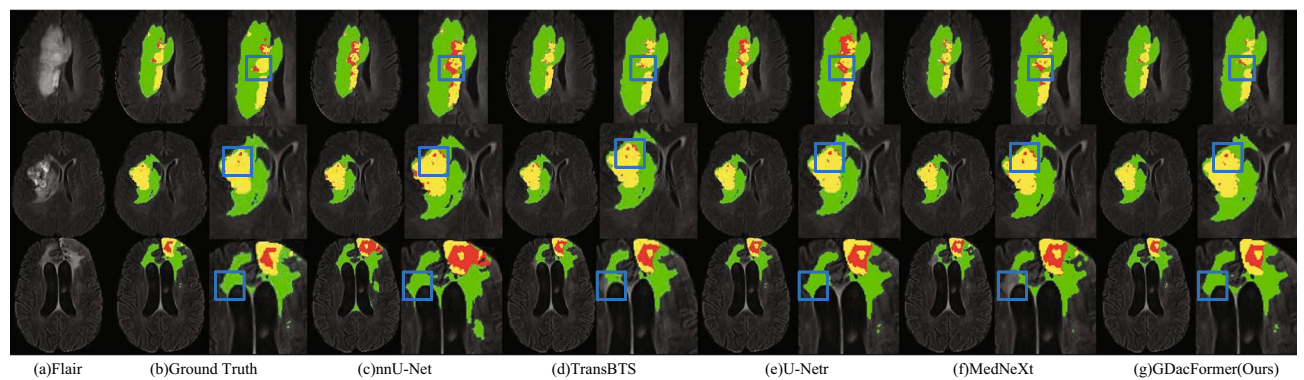


Fig. 7. Visualization of comparison with state-of-the-art methods on the BraTS2021 Training set. The visualized results contain green, yellow, and red regions, representing the edema region, enhanced tumor region, and necrotic region, respectively. (Transverse section).

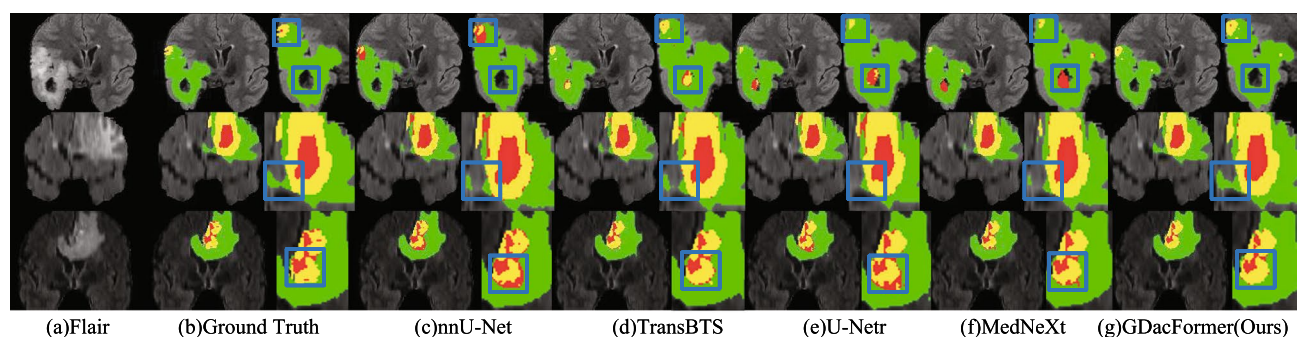


Fig. 8. Visualization of comparison with state-of-the-art methods on the BraTS2021 Training set. (Coronal section).

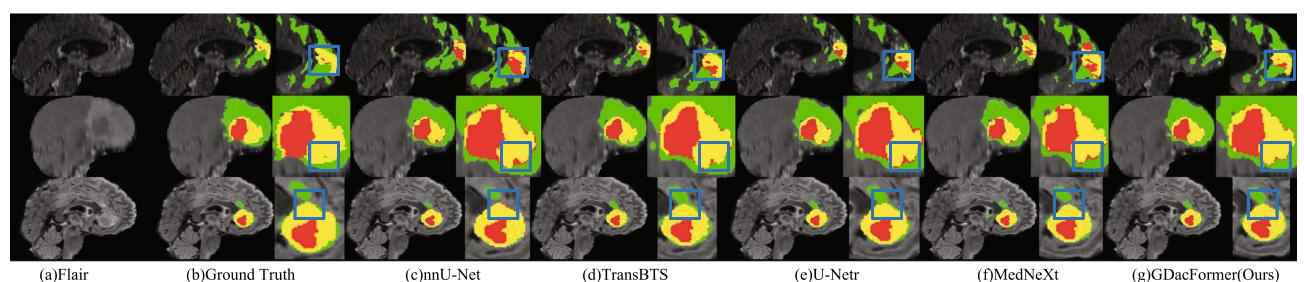


Fig. 9. Visualization of comparison with state-of-the-art methods on the BraTS2021 Training set. (Median sagittal section).

The combined analysis of Tables 2, 3 and 4 clearly demonstrates that GDacFormer consistently outperforms state-of-the-art models in different datasets and evaluation metrics. The model's integration of adversarial learning, MSDA, and NCB modules significantly enhances its capability to accurately segment brain tumor while maintaining precise boundary alignment. These comprehensive improvements highlight GDacFormer as a highly effective and competitive solution for MRI brain tumor segmentation. Consistent performance gains across different datasets emphasize the robustness and generalizability of our proposed model, making it a valuable tool for clinical applications in brain tumor segmentation.

Visualization results

We compare visualization results with state-of-the-art methods in the BraTS2021, shown in Figs. 7, 8 and 9. Each visualized result highlights three distinct regions within the brain tumor: edema (green), enhanced tumor (yellow), and necrotic (red).

First, as shown in Fig. 7, we show the transverse section visualization results. In the first case, the GDacFormer's segmentation results closely resemble the ground truth, particularly in the red region within the blue box. This indicates that GDacFormer accurately captures the necrotic regions of the tumor, providing a clear advantage over other models. Accurate detection and delineation of the necrotic region are critical to effective treatment planning and prognosis. In the second case, the segmentation results show the distribution of small pieces of the red region. GDacFormer demonstrates a clear advantage in accurately segmenting these small, scattered necrotic areas. This capability is vital to ensure that even the smallest tumor regions are identified and treated appropriately, which can significantly impact patient outcomes. The third case focuses on the green region within the blue box, which represents edema. GDacFormer outperforms other models by providing a more precise segmentation of the edema regions. Precise segmentation of the edema is vital to assess the extent of tumor-induced swelling and to plan appropriate surgical or therapeutic interventions.

Second, as shown in Fig. 8, we show the visualization results of the coronal section. The area in the blue box clearly shows the advantage of GDacFormer's segmentation, in the first case only GDacFormer has no incorrectly segmented red regions. In the second case, only the green region segmented by GDacFormer is coherent and closest to GT. In the third case, again focusing on the red region, all networks incorrectly segment a certain amount of the red region, but GDacFormer is the network with the smallest incorrectly segmented region.

Third, we show the visualization results of the Median sagittal section in Fig. 9. In the first and second cases, where we focus on the red region, the error region of the GDacFormer segmentation is clearly the smallest. In the third case, focusing on the green region of the segmentation, GDacFormer is the closest to GT.

In general, Figs. 7, 8 and 9 demonstrates the superior performance of GDacFormer, effectively capturing all tumor regions with high precision. This visual evidence supports the comparison results, confirming GDacFormer as a reliable method for MRI brain tumor segmentation.

Discussion

According to the evaluation experiments, it can be seen that the GAN-based segmentation method combines well the advantages of the U-Net semantic segmentation model and the adversarial learning architecture, thus achieving good segmentation results in the MRI brain tumor segmentation application. Meanwhile, after DacFormer is embedded, additional segmentation accuracy can be gained, which shows the importance of capturing both local and global feature information for segmentation.

Additionally, compared with state-of-the-art works, by well integrating adversarial learning as well as transformer mechanism, GDacFormer gains satisfactory results on three brain tumor datasets. Specifically, it achieves the optimal mean DSC values on both of BraTS2019 and BraTS2020 Validation sets. Although it is slightly worse than MedNeXt in the mean DSC value in the BraTS2021 training set, it outperforms MedNeXt on the WT and TC segmentation. These experimental results demonstrate the good competitiveness of the proposed GDacFormer on brain tumor segmentation task.

Finally, it should also be noted that although GDacFormer achieves the optimal mean DSC results on BraTS2019 and BraTS2020 Validation sets, the corresponding results are only 84.5% and 85.3%, respectively. Even on the BraTS2021 training set, MedNeXt only can achieve the highest mean DSC value of 91.4. These results also indicate that MRI brain tumor segmentation remains a very challenging medical task due to the small percentage of brain tissue, as well as various shapes and sizes of brain tumors.

Conclusion

This work proposes a novel MRI brain tumor segmentation method called GDacFormer, which integrates advanced Transformers with adversarial learning into a single network. The generator incorporates a new transformer module called DacFormer, which consists of Multi-Scale Dilated Attention (MSDA) and Next Convolution Block (NCB) modules. These components work synergistically to capture long-range dependencies at various scales and enhance local feature representations, offering a comprehensive approach to feature extraction. The discriminator ensures that the generated segmentation maps are as close to the ground truth as possible by distinguishing between real and generated images. Extensive experimental results on three brain tumor segmentation datasets demonstrate its effectiveness and competitiveness. In the future, we plan to evaluate the model in other medical image segmentation tasks to demonstrate its generalize ability. In addition, exploring other learning strategies within the adversarial learning or unsupervised³⁸ framework for brain tumor segmentation is a future research interest.

Received: 13 March 2025; Accepted: 15 May 2025

Published online: 22 May 2025

References

1. Ranjbarzadeh, R. et al. Brain tumor segmentation based on deep learning and an attention mechanism using mri multi-modalities brain images. *Sci. Rep.* **11**, 10930 (2021).
2. Jyothi, P. & Singh, A. R. Deep learning models and traditional automated techniques for brain tumor segmentation in mri: a review. *Artif. Intell. Rev.* **56**, 2923–2969 (2023).
3. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* **18**, 234–241 (Springer, 2015).
4. Goodfellow, I. et al. Generative adversarial nets. *Adv. Neural Inform. Process. Syst.* **27** (2014).
5. Asis-Cruz, D. et al. Fetalgan: automated segmentation of fetal functional brain mri using deep generative adversarial learning and multi-scale 3d u-net. *Front. Neurosci.* **16**, 887634 (2022).
6. Ding, Y. et al. Tostagan: An end-to-end two-stage generative adversarial network for brain tumor segmentation. *Neurocomputing* **462**, 141–153 (2021).

7. Mukherjee, D., Saha, P., Kaplun, D., Sinitca, A. & Sarkar, R. Brain tumor image generation using an aggregation of gan models with style transfer. *Sci. Rep.* **12**, 9141 (2022).
8. Xue, Y., Xu, T., Zhang, H., Long, L. R. & Huang, X. Segan: Adversarial network with multi-scale l1 loss for medical image segmentation. *Neuroinformatics* **16**, 383–392 (2018).
9. Nema, S., Dudhane, A., Murala, S. & Naidu, S. Rescuenet: An unpaired gan for brain tumor segmentation. *Biomed. Signal Process. Control* **55**, 101641 (2020).
10. Wang, S. et al. U-net using stacked dilated convolutions for medical image segmentation. arXiv preprint [arXiv:2004.03466](https://arxiv.org/abs/2004.03466) (2020).
11. Tang, P. et al. Efficient skin lesion segmentation using separable-unet with stochastic weight averaging. *Comput. Methods Programs Biomed.* **178**, 289–301 (2019).
12. Bougourzi, F., Distant, C., Dornaika, F. & Taleb-Ahmed, A. Pdatt-unet: Pyramid dual-decoder attention unet for covid-19 infection segmentation from ct-scans. *Med. Image Anal.* **86**, 102797 (2023).
13. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inform. Process. Syst.* **30** (2017).
14. Hatamizadeh, A. et al. Unetr: Transformers for 3d medical image segmentation. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 574–584 (2022).
15. Roy, S. et al. Mednext: transformer-driven scaling of convnets for medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 405–415 (Springer, 2023).
16. Lu, Y. et al. Gmetanet: Multi-scale ghost convolutional neural network with auxiliary metaformer decoding path for brain tumor segmentation. *Biomed. Signal Process. Control* **83**, 104694 (2023).
17. Wang, W. et al. Transbts: Multimodal brain tumor segmentation using transformer. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I* **24**, 109–119 (Springer, 2021).
18. Chen, B., Liu, Y., Zhang, Z., Lu, G. & Kong, A. W. K. Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation. *IEEE Transactions on Emerging Topics in Computational Intelligence* (2023).
19. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929) (2020).
20. Milletari, F., Navab, N. & Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, 565–571 (Ieee, 2016).
21. Menze, B. H. et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2014).
22. Bakas, S. et al. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Sci. Data* **4**, 1–13 (2017).
23. Liu, X., Hou, S., Liu, S., Ding, W. & Zhang, Y. Attention-based multimodal glioma segmentation with multi-attention layers for small-intensity dissimilarity. *J. King Saud Univ. Comput. Inform. Sci.* **35**, 183–195 (2023).
24. Huang, L. et al. A transformer-based generative adversarial network for brain tumor segmentation. *Front. Neurosci.* **16**, 1054948 (2022).
25. Cao, Y. et al. Mbanet: A 3d convolutional neural network with multi-branch attention for brain tumor segmentation from mri images. *Biomed. Signal Process. Control* **80**, 104296 (2023).
26. Akbar, A. S., Fatchah, C. & Suciati, N. Single level unet3d with multipath residual attention block for brain tumor segmentation. *J. King Saud Univ. Comput. Inform. Sci.* **34**, 3247–3258 (2022).
27. Liu, H., Huang, J., Li, Q., Guan, X. & Tseng, M. A deep convolutional neural network for the automatic segmentation of glioblastoma brain tumor: Joint spatial pyramid module and attention mechanism network. *Artif. Intell. Med.* 102776 (2024).
28. Ting, H. & Liu, M. Multimodal transformer of incomplete mri data for brain tumor segmentation. *IEEE J. Biomed. Health Inform.* (2023).
29. Hamghalam, M., Lei, B. & Wang, T. Brain tumor synthetic segmentation in 3d multimodal mri scans. In *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**, 153–162 (Springer, 2020).
30. Peiris, H., Chen, Z., Egan, G. & Harandi, M. Reciprocal adversarial learning for brain tumor segmentation: a solution to brats challenge 2021 segmentation task. In *International MICCAI Brainlesion Workshop*, 171–181 (Springer, 2021).
31. Raza, R., Bajwa, U. I., Mehmood, Y., Anwar, M. W. & Jamal, M. H. dresu-net: 3d deep residual u-net based brain tumor segmentation from multimodal mri. *Biomed. Signal Process. Control* **79**, 103861 (2023).
32. An, D. et al. Dynamic weighted knowledge distillation for brain tumor segmentation. *Pattern Recogn.* **155**, 110731 (2024).
33. Jia, Z., Zhu, H., Zhu, J. & Ma, P. Two-branch network for brain tumor segmentation using attention mechanism and super-resolution reconstruction. *Comput. Biol. Med.* **157**, 106751 (2023).
34. 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**, 203–211 (2021).
35. Huang, L., Denoeux, T., Vera, P. & Ruan, S. Evidence fusion with contextual discounting for multi-modality medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 401–411 (Springer, 2022).
36. Zhou, H.-Y. et al. nnformer: Volumetric medical image segmentation via a 3d transformer. *IEEE Trans. Image Process.* **32**, 4036–4045 (2023).
37. Xie, Q., Chen, Y., Liu, S. & Lu, X. Sscformer: Revisiting convnet-transformer hybrid framework from scale-wise and spatial-channel-aware perspectives for volumetric medical image segmentation. *IEEE J. Biomed. Health Inform.* (2024).
38. Yang, Y. et al. Integrating fuzzy clustering and graph convolution network to accurately identify clusters from attributed graph. *IEEE Trans. Network Sci. Eng.* **12**, 1112–1125 (2025).

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61972062, the Applied Basic Research Project of Liaoning Province under Grants 2023JH2/101300191 and Liaoning Province Science and Technology Plan Joint Program (Key Research and Development Program Project 2024JH2/102600089).

Additional information

Correspondence and requests for materials should be addressed to J.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025