Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 16, No. 1, 2025

Brain tumor segmentation using deep learning: A Review

Beibei Hou, Saizong Guan

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China

Abstract: Brain tumor segmentation is a crucial task in medical image analysis, as accurate delineation of tumor regions is vital for clinical diagnosis, treatment planning, and prognosis assessment. Traditional Convolutional Neural Network (CNN)-based models have demonstrated significant success in capturing local features, but they face challenges in modeling global context, which is essential for complex segmentation tasks. This review examines recent advancements in brain tumor segmentation, with a focus on CNNs, Transformers, Mamba, and Graph Neural Networks (GNNs), as well as their hybrid models. This review critically evaluates the strengths and limitations of each approach with respect to architecture, segmentation accuracy, and real-world applicability. Additionally, it addresses key challenges such as computational complexity and data scarcity, and proposes future research directions to enhance the practical use of these methods in clinical settings.

Keywords: Brain tumor segmentation; Deep learning; Transformer; CNN; GNN; Mamba.

1. Introduction

Medical image segmentation is a fundamental task in medical image analysis, playing a central role in identifying and evaluating anatomical structures, diseases, or regions of interest [1], [2]. By accurately delineating these areas, segmentation forms the basis for estimating disease prognosis and devising targeted treatment strategies, making it indispensable for advancing precision medicine and improving clinical outcomes [3]. Brain tumors refer to abnormal cell growth within or near the brain, including primary brain tumors and brain metastases [4]. Gliomas [5], one of the most common primary brain tumors, are classified into High-Grade Gliomas (HGG, III–IV) and Low-Grade Gliomas (LGG, I–II) [6]. Brain tumor segmentation is one of the most challenging tasks in medical image analysis due to the complex and diverse nature of brain tumors, variations in tumor shapes, sizes, and locations, as well as the difficulty of distinguishing tumors from surrounding healthy tissue [7].

MRI is preferred over CT in brain tumor imaging due to its high contrast, low radiation [8], and diverse sequences (e.g., T1, T2, FLAIR, T1c), which are crucial for diagnosing and evaluating conditions like gliomas, neurodegenerative disorders, and treatment outcomes [9].

The process of brain tumor segmentation can be formally described as follows: given one or multiple images from various imaging modalities (e.g., different MRI sequences), the goal is to automatically assign each voxel or pixel in the input data to a specific predefined sub-region, effectively separating tumor areas from normal tissues.

Accurate segmentation of brain tumors allows for localization and delineation of tumor boundaries, which is essential for surgical planning, radiation therapy, and monitoring treatment efficacy [10]. The diverse shapes, sizes, and locations of brain tumors present significant challenges for conventional segmentation approaches. As a result, achieving automatic and precise brain tumor segmentation is essential to minimize human errors, enhance clinical outcomes, and facilitate timely interventions.

As shown in Figure 1, an example of a high-grade glioma (HGG) case is presented, demonstrating input data from several MRI modalities and their corresponding ground truth 3D volume segmentation map. Each MRI sequence provides distinct tissue characteristics, while the ground truth includes manual annotations with different colors representing specific tumor sub-regions: Peritumoral Edema (green), Necrotic and Non-Enhancing Tumor Core (red), and Enhancing Tumor (yellow).

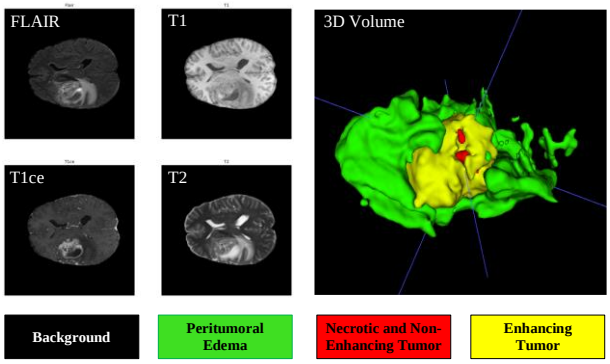

Fig. 1 Example of input MRI modalities and ground truth 3D volume segmentation map

Over the past decade, the rise of deep learning techniques, particularly CNNs, has revolutionized segmentation methods [11]-[15]. CNNs have revolutionized medical image analysis by effectively capturing local features, with architectures like U-Net [16] further enhancing segmentation tasks through the use of encoder-decoder structures and skip connections.

In recent years, the field of brain tumor segmentation has seen significant advancements through the integration of various deep learning architectures, each addressing unique challenges. Figure 2 illustrates the Keyword Frequency of Brain Tumor Segmentation in MICCAI (2021–2024). Traditional models like U-Net and its variants primarily rely on convolutional operations, which excel at capturing local features but fall short when it comes to modeling long-range dependencies and global relationships—critical for complex tasks like brain tumor delineation. To address these limitations, Vision Transformers (ViTs) [17] have emerged as a transformative approach, leveraging self-attention mechanisms to capture global context effectively and enhance

segmentation performance.

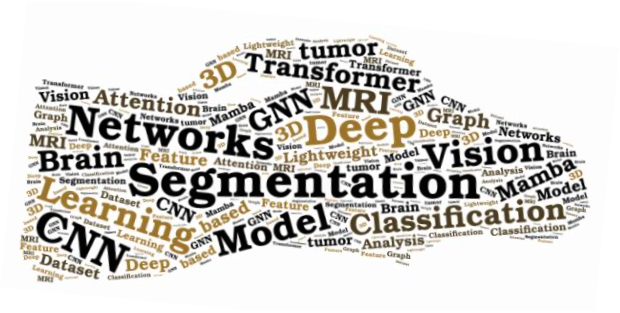

Fig. 2 Keyword Frequency Map of Brain Tumor Segmentation in MICCAI (2021–2024)

While ViTs address global dependencies, GNNs [18] provide a unique advantage by modeling irregular data structures and spatial relationships. Unlike CNNs that focus on local pixel-level features and Transformers that model global dependencies, GNNs excel at capturing complex anatomical and spatial relationships. These networks can effectively represent the connectivity between different regions of an image, allowing for a more nuanced understanding of the tumor's structure.

Furthermore, Mamba [19], as a state-space model (SSM) [20], has shown immense promise in bridging the gap between local and global feature extraction. Mamba's ability to model long-range dependencies efficiently and maintain linear computational complexity positions it as a key player in medical image segmentation. As Mamba continues to evolve, it is expected to drive significant advancements not only in medical imaging but also in a wide range of other domains.

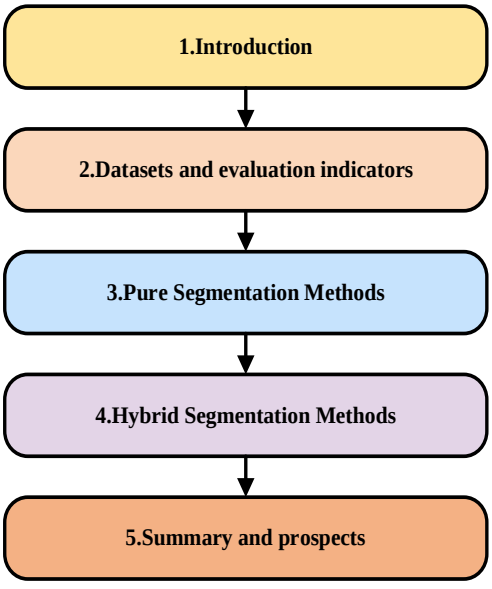

Fig. 3 The structure of the survey paper

As shown in Figure 3, this paper aims to provide a comprehensive review of the current methodologies used in brain tumor segmentation, highlighting advancements, challenges, and future directions. The paper examines the strengths and weaknesses of traditional models like CNNs, as well as emerging approaches like ViTs, Mamba, and GNNs. By exploring hybrid models that combine these techniques, this review seeks to offer insights into how these architectures contribute to more accurate and efficient segmentation. The findings aim to guide future research efforts in overcoming existing challenges, such as computational complexity and data limitations, and to inspire innovations that will enhance the clinical applicability and real-world performance of brain tumor segmentation methods.

2. Datasets and evaluation indicators

2.1. Datasets

In order to evaluate and compare brain tumor segmentation methods objectively, we used the official BraTS 2018, BraTS 2019, and BraTS 2020 datasets[21][23], which have been widely used in brain tumor segmentation research. These datasets offer multi-modal MRI images with manually annotated ground truth labels, making them ideal for benchmarking segmentation algorithms.

The BraTS 2018 dataset consists of 285 patients in the training set and 66 unlabeled patients in the validation set. The BraTS 2019 dataset includes 335 patients (259 HGG and 76 LGG) in the training set, with 125 unlabeled patients in the validation set. The BraTS 2020 dataset contains 369 patients in the training set and 125 unlabeled patients in the validation set. Each image in the training sets includes four MRI modalities: T1, T1c, T2, and FLAIR, which capture distinct features of brain tumors. All images have dimensions of 240×240×155 mm³ and a voxel spacing of 1×1×1 mm³.

2.2. Evaluation indicators

Brain tumor segmentation is a critical task in medical imaging, requiring precise and reliable methods to delineate tumor sub-regions. The datasets used for benchmarking, such as BraTS 2018, BraTS 2019, and BraTS 2020, provide manually annotated ground truth labels for four categories: Background (label 0), Necrosis and Non-Enhancing Tumor (NCR/NET, label 1), Peritumoral Edema (ED, label 2), and Enhancing Tumor (ET, label 4). During evaluation, the ground truth is reformatted into three primary regions for comparison: WT (labels 1, 2, and 4), TC (labels 1 and 4), and ET (label 1).

In brain tumor segmentation tasks, two common evaluation metrics are the Dice Similarity Coefficient (Dice score) and the Hausdorff Distance (95%), which provide complementary insights into segmentation accuracy and boundary alignment. Then they can be expressed as follows:

$$\text{Dice} = \frac{2|A \cap B|}{|A|+|B|} = \frac{2TP}{2TP+FP+FN} \tag{1}$$

$$HD = \max\{\sup_{p\in B}\inf_{q\in A} d(p,q), \sup_{q\in A}\inf_{p\in B} d(p,q)\} \tag{2}$$

A and B are the sets representing the ground truth and predicted segmentation, respectively. The terms TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative voxel counts.

3. Pure Segmentation Methods

Table 1 presents the segmentation results of various CNN-based or Transformer-based models on the BraTS dataset. It includes the Dice coefficients for three tumor regions: WT, TC, and ET. The Dice scores provide a measure of segmentation accuracy for each tumor region.

Table 1. Segmentation results of various models on the BraTS dataset.

Method	Dataset	Dice_WT	Dice_TC	Dice_ET
3D-UNet	BraTS2018	0.760	0.885	0.718
3D-UNet	BraTS2020	0.882	0.830	0.782
SegResNet	BraTS2020	0.903	0.845	0.796
nnUNet	BraTS2020	0.907	0.848	0.814
SwinUNet	BraTS2020	0.872	0.809	0.744
V-Net	BraTS2019	0.739	0.887	0.766

3.1. Pure CNN

CNN-based brain tumor segmentation models have been extensively studied and are widely used in the field. 3D U-Net [24] is a cornerstone in pure CNN-based segmentation. As a 3D extension of the classic U-Net, it processes volumetric data using 3D convolutions and pooling operations. The encoder-decoder structure with skip connections ensures that spatial context is captured effectively while preserving important details, improving segmentation accuracy. This model laid the foundation for future advancements by offering computational efficiency and robustness.

Building on the success of U-Net, V-Net [25] introduces improvements such as residual convolutions in its encoder-decoder structure. This modification enhances gradient flow and model stability. Furthermore, V-Net integrates Dice Loss, addressing the common issue of class imbalance in medical images, making it particularly useful for tasks like brain tumor segmentation. The use of "same" convolutions improves model interpretability, and its 3D convolutions continue the legacy of robust volumetric data processing started by 3D U-Net.

SegResNet [26] introduces an additional layer of complexity by incorporating a Variational Autoencoder (VAE). This VAE module encourages better feature learning by reconstructing the original input during training, enhancing the encoder's ability to capture a more detailed latent representation. While its CNN-based framework remains the core, the addition of VAE boosts the model's ability to generalize and capture fine-grained details necessary for accurate segmentation. This innovation complements the skip connection approach of U-Net and V-Net, allowing for more comprehensive feature extraction[27].

Finally, nnU-Net [28] takes automation a step further by streamlining the entire segmentation pipeline. While earlier models like 3D U-Net, V-Net, and SegResNet require manual adjustments, nnU-Net automates preprocessing, architecture selection, and training. This makes it highly adaptable to a wide range of datasets, eliminating the need for task-specific optimization. nnU-Net also maintains the strengths of its predecessors, ensuring that it performs consistently across different challenges, setting a new benchmark for efficiency and generalization.

3.2. Pure Transformer

The architecture based solely on Transformer, which only includes ViT layers, has limited applications in medical image segmentation because both global and local information are crucial for dense prediction tasks such as segmentation. Karimi et al. proposed a pure Transformer model for 3D segmentation, utilizing self-attention across linear embeddings of 3D patches. However, a significant drawback of pure Transformer models is the quadratic complexity of self-attention with respect to input image dimensions, which limits their applicability to high-resolution medical images.

To overcome these limitations, Swin-Unet [29] was introduced as the first U-shaped pure Transformer-based architecture tailored for 2D medical image segmentation. It incorporates Swin Transformer [30] blocks with shifted window attention, enabling efficient multi-scale feature extraction and the modeling of both local and global context. The architecture comprises an encoder, bottleneck, decoder, and skip connections, which together facilitate precise spatial detail recovery and high-resolution segmentation predictions. By demonstrating superior performance over CNN and hybrid approaches, Swin-Unet highlights the potential of pure Transformer models in advancing medical image segmentation.

3.3. Pure Mamba

Mamba is an innovative state-space model that integrates selective information processing, hardware-aware algorithms, and a simplified SSM architecture. By parameterizing the input of the SSM, it enables selective attention, enhancing efficiency while maintaining performance. VM-UNet [31] is the first pure state-space model (SSM)-based architecture designed for medical image segmentation. It integrates the Vision Mamba model into a U-Net framework, leveraging the Visual State Space (VSS) block to efficiently capture long-range dependencies with linear computational complexity. The architecture features an asymmetric encoder-decoder design with components like patch embedding, VSS blocks, and skip connections, offering a scalable and efficient solution for medical image segmentation tasks.

Mamba-UNet [32]is a novel architecture for medical image segmentation that synergizes with U-Net. Mamba-UNet utilizes a pure Visual Mamba (VMamba)-based encoder-decoder structure, incorporating skip connections to preserve spatial information at different scales. This design facilitates comprehensive feature learning, capturing intricate details and broader semantic contexts in medical images. Additionally, we introduce a novel integration mechanism within the VMamba block to ensure seamless connectivity and information flow between the encoder and decoder paths, further enhancing segmentation performance.

3.4. Pure GNN

The use of pure GNN models in medical image segmentation has been relatively underexplored. Unlike classification tasks, which benefit from GNNs' strength in modeling relationships and global dependencies, medical image segmentation requires precise delineation of complex anatomical structures at a pixel level. This precision is

difficult to achieve with GNNs alone due to their tendency to produce coarse boundaries rather than sharp, accurate edges. Consequently, most research focuses on hybrid approaches, combining GNNs with convolutional architectures (e.g., CNNs or Transformers) to leverage the strengths of both models. As of now, there is little to no established work on pure GNN-based medical segmentation models in mainstream literature, as researchers prioritize methods that achieve higher accuracy and boundary refinement through integrated frameworks.

4. Hybrid Segmentation Methods

Hybrid architectures that combine the strengths of multiple deep learning models have gained significant attention in medical image segmentation, particularly in tasks like brain tumor segmentation. By integrating different types of neural networks, these hybrid models leverage the complementary strengths of each architecture, enhancing the model's ability to capture both fine-grained local features and broader global dependencies. In the following sections, we categorize these hybrid models into four main types: CNN-Transformer hybrids, CNN-Mamba hybrids, GNN-Transformer hybrids, and CNN-GNN hybrids. Each of these approaches combines different architectural components to address unique challenges in segmentation tasks, offering a more effective solution than using a single architecture alone. Table 2 presents segmentation results for hybrid architectures combining deep learning models, focusing on brain tumor segmentation.

Table 2. Segmentation results of various models on the BraTS dataset.

Method	Dataset	Dice_WT	Dice_TC	Dice_ET
TransBTS	BraTS2020	0.910	0.855	0.791
UNETR	BraTS2020	0.899	0.842	0.788
NestedFormer	BraTS2020	0.920	0.864	0.800
SwinUNet	BraTS2020	0.872	0.809	0.744
S2CA-Net	BraTS2020	0.804	0.914	0.852
M2GCNet	BraTS2019	0.856	0.843	0.783

4.1. CNN-Transformer hybrids

CNN-Transformer hybrid models combine the strengths of CNN's local feature extraction and Transformer's ability to model global dependencies, making them highly effective for 3D brain tumor segmentation tasks. One significant advancement in this area is the SegFormer3D [33] model, which leverages hierarchical ViTs to enhance global contextual understanding while addressing the challenges of limited datasets and large computational requirements. Unlike traditional CNN-based models, which primarily focus on local feature extraction, SegFormer3D computes attention across multiscale volumetric features, improving segmentation accuracy. By utilizing an all-MLP decoder to aggregate local and global attention features, SegFormer3D avoids the need for complex decoders, making the design more memory-efficient. With 33× fewer parameters and a 13× reduction in GFLOPS compared to state-of-the-art models, SegFormer3D achieves competitive performance on benchmark datasets such as Synapse, BraTS, and ACDC, providing a lightweight yet highly accurate solution for 3D brain tumor segmentation.

Building upon the success of hybrid architectures, S2CA-Net [34] introduces a novel approach to brain tumor segmentation by addressing challenges such as variations in tumor shape, size, and location. While CNN-based models struggle with these challenges due to their limited receptive fields, S2CA-Net introduces a shape-scale co-awareness mechanism that learns both shape-aware and scale-aware features simultaneously. Key components like the Local-Global Scale Mixer (LGSM), Multi-level Context Aggregator (MCA), and Multi-Scale Attentive Deformable Convolution (MS-ADC) work together to enhance segmentation accuracy and robustness. S2CA-Net outperforms existing methods on benchmark datasets, demonstrating its effectiveness in handling the complexities inherent in brain tumor segmentation.

Further advancing the encoder-decoder design, TransBTS [35] combines the strengths of both 3D CNNs and Transformers, specifically targeting MRI-based brain tumor segmentation (BTS). The model employs 3D CNNs in the encoder to extract compact spatial and depth-aware feature maps, which are then passed into Transformer layers for global feature modeling. This combination ensures that TransBTS captures both local and global dependencies effectively. The decoder utilizes 3D CNN-based progressive feature upsampling to restore high-resolution segmentation outputs. Skip connections further enhance segmentation accuracy by preserving finer spatial details. Through this hybrid approach, TransBTS achieves a balance between computational efficiency and segmentation accuracy, making it well-suited for 3D medical imaging tasks.

In the realm of multi-modal MRI segmentation, NestedFormer [36] introduces a novel architecture that explicitly models both intra-modal and inter-modal dependencies for brain tumor segmentation. The key innovation is the Nested Modal-Aware Feature Aggregation (NMaFA) module, which uses nested transformers to capture long-range spatial dependencies within each modality and across modalities. This is complemented by the Global Poolformer Encoder to enhance global feature extraction and a Multi-scale Fusion strategy with Modal-Sensitive Gating (MSG) to combine multi-modal features efficiently. NestedFormer stands out by effectively handling multi-modal fusion while maintaining computational efficiency, further advancing segmentation capabilities in multi-modal MRI datasets.

Lastly, UNETR [37] and its improved version, UNETR++ [38], represent a shift toward Transformer-based architectures for 3D brain tumor segmentation. UNETR replaces the traditional CNN encoder in the U-Net structure with pure Transformers, capturing long-range dependencies and global context directly from the input volumetric data. This design is particularly well-suited for brain tumor segmentation, allowing for the effective modeling of complex structures through tokenized patches. Building upon this, UNETR++

addresses the computational bottleneck of self-attention in Transformers by introducing the Efficient Paired Attention (EPA) block. This block learns spatial and channel dependencies through two mutually dependent branches using spatial and channel attention, ensuring linear complexity in spatial attention calculations. By sharing query and key mapping weights between the branches, UNETR++ reduces overall network parameters while maintaining high segmentation performance, offering an efficient and effective solution for 3D segmentation tasks.

4.2. CNN-GNN hybrids

DGRUnit [39] is a novel approach for brain tumor segmentation, comprising two parallel graph reasoning modules: a spatial reasoning module using a Graph Convolutional Network (GCN) to capture long-range spatial dependencies, and a channel reasoning module using a Graph Attention Network (GAT) to model contextual interdependencies between image channels. This approach significantly improves segmentation performance, particularly in tumor regions, and is highly flexible and generalizable.

Building on this concept, M2GCNet [40] introduces the multi-modal graph convolution module (M2GCM), which incorporates two key components: the spatial-wise graph convolution module (SGCM) to capture spatial dependencies and the channel-wise graph convolution module (CGCM) to model contextual relationships between different image channels. M2GCNet further improves feature learning through the introduction of a multi-modal correlation loss, which captures nonlinear relationships between modality pairs, allowing for a more comprehensive understanding of the tumor's characteristics across multiple MRI modalities.

4.3. CNN- Mamba hybrids

HC-Mamba [41] is a novel medical image segmentation model that integrates hybrid convolutional techniques with the modern state-space model Mamba. To address the challenges of image resolution reduction and information loss due to downsampling in medical images, HC-Mamba introduces dilated convolutions, enabling the model to capture broader contextual information without increasing computational costs. Additionally, the model utilizes depthwise separable convolutions, significantly reducing the number of parameters and computational requirements. By combining these techniques, HC-Mamba enhances the model's receptive field while maintaining high performance at a lower computational cost, making it effective for large-scale medical image data processing.

LKM-UNet [42] introduces a novel approach to medical image segmentation by incorporating Large-Kernel Mamba (LM) blocks, which excel at modeling local spaces compared to smaller kernels used in traditional CNNs and Transformers. The architecture leverages Hierarchical and Bidirectional Mamba blocks, designed to enhance both global and local spatial modeling capabilities of the Mamba mechanism in visual inputs. Additionally, Pixel-level Spatial Semantic Modeling (PiM) and Patch-level Spatial Semantic Modeling (PaM) are implemented to capture pixel-level neighborhood information and long-range dependencies respectively, further enhancing segmentation performance.

Extending the concepts from LKM-UNet, LMa-UNet [43] focuses on overcoming the limitations of small-kernel CNNs and Transformer-based models, particularly their restricted receptive fields. By leveraging large-window-based Mamba networks, LMa-UNet enhances local spatial modeling while maintaining efficiency in global context modeling. This is achieved without the quadratic complexity of self-attention, making the model computationally efficient for large-scale medical image segmentation. Like LKM-UNet, LMa-UNet incorporates hierarchical and bidirectional Mamba blocks to improve both global and neighborhood space modeling. Additionally, the introduction of Pixel-level SSM (PiM) and Patch-level SSM (PaM) further boosts the model's ability to extract local pixel features and model long-range dependencies.

LightM-UNet [44] integrates Mamba and UNet within a lightweight framework, using residual visual Mamba layers to extract deep semantic features and model long-range spatial dependencies with linear complexity. With a parameter count of only 1M, LightM-UNet surpasses existing state-of-the-art models in validation on 2D and 3D real-world datasets. This approach represents a novel attempt to incorporate Mamba into UNet as a lightweight optimization strategy, aiming to address computational resource constraints in practical medical applications.

4.4. GNN-Transformer hybrids

SGFormer [45] is a simplified graph Transformer designed for large-scale graph processing, utilizing a single-layer global attention mechanism for linear computational complexity. By combining this with a traditional GNN, SGFormer efficiently updates node representations and models interactions between all nodes. It eliminates the need for position encoding, edge embedding, and augmented loss functions, offering a simpler and more computationally efficient solution. Its ability to scale to graphs with billions of nodes makes SGFormer an ideal tool for handling large and complex medical datasets, enhancing segmentation accuracy while maintaining efficiency.

4.5. Mamba-Transformer hybrids

MambaVision [46] is a novel hybrid Mamba-Transformer backbone designed for vision applications, with a focus on efficiently modeling visual features. The key innovation involves redesigning the Mamba formulation to enhance global context learning, alongside integrating multiple self-attention blocks in the final layers to better capture long-range spatial dependencies. For medical image segmentation, MambaVision's ability to model complex spatial relationships and capture long-range dependencies can significantly improve the accuracy and efficiency of segmenting intricate anatomical structures and pathology, benefiting tasks like tumor detection and organ segmentation.

5. Summary

Brain tumor segmentation plays a vital role in medical image analysis, with accurate tumor delineation being essential for diagnosis, treatment planning, and prognosis assessment. This review has explored recent advancements in brain tumor segmentation, focusing on various deep learning techniques, including CNNs, Transformers, Mamba, and GNNs, as well as their hybrid models. The review highlights the strengths and limitations of these approaches, particularly in capturing both local features (via CNNs) and global context (via Transformers and other mechanisms). Traditional CNN-based models have demonstrated excellent performance in capturing local information but face challenges in modeling

long-range dependencies, which are crucial for accurately segmenting complex tumor regions. Hybrid models, such as those combining CNNs and Transformers, have been effective in overcoming these limitations by enabling the simultaneous capture of detailed local features and global semantic information.

Among the reviewed models, transformer-based architectures have gained significant attention for their ability to model long-range dependencies and improve segmentation accuracy. However, their high computational complexity remains a challenge. To address this, many models have incorporated transformer-based components with lower-resolution feature maps or used them in specific parts of the architecture. Additionally, modifications to the traditional transformer structure, including the use of multi-head self-attention mechanisms and MLP blocks, have helped reduce the parameter count while maintaining high segmentation performance.

While the application of transformers has shown promise, challenges such as data scarcity, computational efficiency, and real-world applicability remain. In particular, the need for large annotated datasets is a significant hurdle, and self-supervised, semi-supervised, and weakly-supervised learning approaches are emerging as promising solutions to mitigate this issue. Moreover, the integration of transformers with other techniques, such as Mamba and GNNs, holds potential for enhancing segmentation accuracy and improving model robustness.

Acknowledgements

Key scientific and technological projects in Henan province (242102211042)

Doctoral Fund Project of Henan Polytechnic University (B2022-14)

References

[1] Wang R, Lei T, Cui R, et al. Medical image segmentation using deep learning: A survey[J]. IET image processing, 2022, 16(5): 1243-1267.

[2] Jha D, Riegler M A, Johansen D, et al. Doubleu-net: A deep convolutional neural network for medical image segmentation[C]//2020 IEEE 33rd International symposium on computer-based medical systems (CBMS). IEEE, 2020: 558-564.

[3] Graham S, Vu Q D, Raza S E A, et al. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images[J]. Medical image analysis, 2019, 58: 101563.

[4] Pichaivel M, Anbumani G, Theivendren P, et al. An overview of brain tumor[J]. Brain Tumors, 2022, 1: 1-10.

[5] Nabors L B, Portnow J, Ahluwalia M, et al. Central nervous system cancers, version 3.2020, NCCN clinical practice guidelines in oncology[J]. Journal of the National Comprehensive Cancer Network, 2020, 18(11): 1537-1570.

[6] Fangusaro J. Pediatric high grade glioma: a review and update on tumor clinical characteristics and biology[J]. Frontiers in oncology, 2012, 2: 105.

[7] Nadeem M W, Ghamdi M A A, Hussain M, et al. Brain tumor analysis empowered with deep learning: A review, taxonomy, and future challenges[J]. Brain sciences, 2020, 10(2): 118.

[8] Maqsood S, Damaševičius R, Maskeliūnas R. Multi-modal brain tumor detection using deep neural network and multiclass SVM[J]. Medicina, 2022, 58(8): 1090.

[9] Ranjbarzadeh R, Zarbakhsh P, Caputo A, et al. Brain tumor segmentation based on optimized convolutional neural network and improved chimp optimization algorithm[J]. Computers in Biology and Medicine, 2024, 168: 107723.

[10] Rasool N, Bhat J I. A Critical Review on Segmentation of Glioma Brain Tumor and Prediction of Overall Survival[J]. Archives of Computational Methods in Engineering, 2024: 1-45.

[11] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[13] Simonyan K. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[14] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[15] 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: 3431-3440.

[16] 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. Springer International Publishing, 2015: 234-241.

[17] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[18] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. IEEE transactions on neural networks, 2008, 20(1): 61-80.

[19] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces[J]. arXiv preprint arXiv:2312.00752, 2023.

[20] Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces[J]. arXiv preprint arXiv:2111.00396, 2021.

[21] Menze B H, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS)[J]. IEEE transactions on medical imaging, 2014, 34(10): 1993-2024.

[22] Bakas S, Akbari H, Sotiras A, et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features[J]. Scientific data, 2017, 4(1): 1-13.

[23] Bakas S, Reyes M, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge[J]. arXiv preprint arXiv:1811.02629, 2018.

[24] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer International Publishing, 2016: 424-432.

[25] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). Ieee, 2016: 565-571.

[26] Myronenko A, Siddiquee M M R, Yang D, et al. Automated head and neck tumor segmentation from 3D PET/CT HECKTOR 2022 challenge report[M]//3D Head and Neck Tumor Segmentation in PET/CT Challenge. Cham: Springer Nature Switzerland, 2022: 31-37.

[27] Heidari M, Kazerouni A, Soltany M, et al. Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2023: 6202-6212.

[28] Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature methods, 2021, 18(2): 203-211.

[29] Cao H, Wang Y, Chen J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 205-218.

[30] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.

[31] Ruan J, Xiang S. Vm-unet: Vision mamba unet for medical image segmentation[J]. arXiv preprint arXiv:2402.02491, 2024.

[32] Wang Z, Zheng J Q, Zhang Y, et al. Mamba-unet: Unet-like pure visual mamba for medical image segmentation[J]. arXiv preprint arXiv:2402.05079, 2024.

[33] Perera S, Navard P, Yilmaz A. SegFormer3D: an Efficient Transformer for 3D Medical Image Segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 4981-4988.

[34] Zhou L, Jiang Y, Li W, et al. Shape-Scale Co-Awareness Network for 3D Brain Tumor Segmentation[J]. IEEE Transactions on Medical Imaging, 2024.

[35] Wenxuan W, Chen C, Meng D, et al. Transbts: Multimodal brain tumor segmentation using transformer[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. 2021: 109-119.

[36] Xing Z, Yu L, Wan L, et al. NestedFormer: Nested modality-aware transformer for brain tumor segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 140-150.

[37] Hatamizadeh A, Tang Y, Nath V, et al. Unetr: Transformers for 3d medical image segmentation[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022: 574-584.

[38] Shaker A M, Maaz M, Rasheed H, et al. UNETR++: delving into efficient and accurate 3D medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2024.

[39] Ma Q, Zhou S, Li C, et al. DGRUnit: Dual graph reasoning unit for brain tumor segmentation[J]. Computers in Biology and Medicine, 2022, 149: 106079.

[40] Zhou T. M2GCNet: Multi-Modal Graph Convolution Network for Precise Brain Tumor Segmentation Across Multiple MRI Sequences[J]. IEEE Transactions on Image Processing, 2024.

[41] Xu J. HC-Mamba: Vision MAMBA with Hybrid Convolutional Techniques for Medical Image Segmentation[J]. arXiv preprint arXiv:2405.05007, 2024.

[42] Wang J, Chen J, Chen D, et al. LKM-UNet: Large kernel vision mamba unet for medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2024: 360-370.

[43] Wang J, Chen J, Chen D, et al. Large window-based mamba unet for medical image segmentation: Beyond convolution and self-attention[J]. arXiv preprint arXiv:2403.07332, 2024.

[44] Liao W, Zhu Y, Wang X, et al. Lightm-unet: Mamba assists in lightweight unet for medical image segmentation[J]. arXiv preprint arXiv:2403.05246, 2024.

[45] Wu Q, Zhao W, Yang C, et al. Simplifying and empowering transformers for large-graph representations[J]. Advances in Neural Information Processing Systems, 2024, 36.

[46] Hatamizadeh A, Kautz J. Mambavision: A hybrid mamba-transformer vision backbone[J]. arXiv preprint arXiv:2407.08083, 2024.