Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 4, No. 2, 2025

Brain Tumor Image Segmentation with Convolutional Neural Networks: A Review

Beibei Hou, Tiansong Sheng

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

Abstract: Brain tumor segmentation is essential in medical image analysis for clinical diagnosis, treatment planning, and prognosis. Despite significant progress, challenges remain, including limited data annotation, high computational costs, and poor model generalization. To address these, researchers have proposed CNN-based models (e.g., FCN, U-Net, U-Net++) and advanced architectures like large kernel convolution (LKC), deformable convolution (DCN), and CNN-transformer hybrids. This paper examines the widely used BraTS dataset and evaluation metrics such as Dice coefficients and Hausdorff distances, while addressing current challenges. Researchers are also exploring strategies like joint learning, self-supervised learning, multimodal fusion, and lightweight model design. These advances aim to improve segmentation performance and expand clinical applications.

Keywords: Brain tumor segmentation, Deep Learning, Medical image analysis, BraTS Dataset, LKC.

1. Introduction

The brain serves as the central regulatory system of the human body, coordinating physiological functions and ensuring the proper operation of the nervous system essential to overall health. Brain tumors are common neurological disorders caused by the abnormal proliferation of cells within brain tissue. Brain tumors pose significant challenges to clinical diagnosis and treatment due to their pathological complexity and potentially severe complications [1, 9, 31]. Accurate and early diagnosis is critical for selecting effective treatment and improving patient prognosis. Medical imaging techniques, particularly magnetic resonance imaging (MRI), play a crucial role in detecting, diagnosing, and classifying brain malignancies [2].

Image segmentation is a key technique in medical imaging that divides an image into meaningful regions, helping to emphasize important structures, reduce complexity, and support further analysis [3]. In imaging modalities such as MRI and CT, particularly for brain tumor diagnosis, segmentation is essential for accurately delineating tumor boundaries [4]. CNNs have demonstrated strong performance in brain tumor segmentation by learning complex patterns in medical images, eliminating the need for handcrafted rules or traditional classification approaches. CNN-based algorithms typically achieve higher accuracy in brain tumor diagnosis and classification compared to conventional methods [5].

However, brain tumor segmentation remains challenging due to the wide variation in tumor morphology, size, and location, as well as irregular or blurred boundaries. In non-contrast MRI scans, the tumor region often occupies a much smaller area than the background, which can lead to models to focus excessively on non-lesion regions and compromise segmentation accuracy during training [6]. Addressing these issues requires developing more robust and adaptive segmentation techniques.

Convolutional neural networks (CNNs) have demonstrated strong capabilities in learning meaningful feature representations in medical image analysis, enabling improved accuracy and computational efficiency in tumor segmentation tasks. This review systematically explores CNN-based segmentation algorithms, analyzes their strengths and limitations, and discusses potential strategies for improvement.

2. Classical Models in Brain Tumor Segmentation

2.1. FCN

The Fully Convolutional Network (FCN) [7] was the first end-to-end pixel-wise prediction model to achieve a breakthrough in semantic segmentation. Its core innovation lies in replacing fully connected layers with convolutional layers, enabling the model to process input images of varying sizes and extract multilevel features through hierarchical downsampling (Long et al., 2015). In addition, FCN achieves dense predictions through transposed convolution, which restores low-resolution feature maps to the original input size.

Long et al. (2015) reported that in the PASCAL VOC semantic segmentation task, the inference time of FCN-8s was approximately 175 ms, significantly accelerating the process compared to the SDS method, with CNN computation achieving a speedup of up to 114× [7]. This indicates that FCN, relying on the end-to-end fully convolutional architecture, not only improves the computational efficiency of semantic segmentation, but also reduces the dependence on candidate regions and post-processing of traditional methods, thus reducing the computational overhead.

However, due to its reliance on pooling operations, FCN may lose spatial information during downsampling, potentially reducing segmentation accuracy—particularly in regions with small tumors or lesions and blurred boundaries. A schematic of the FCN architecture is presented in Fig. 1.

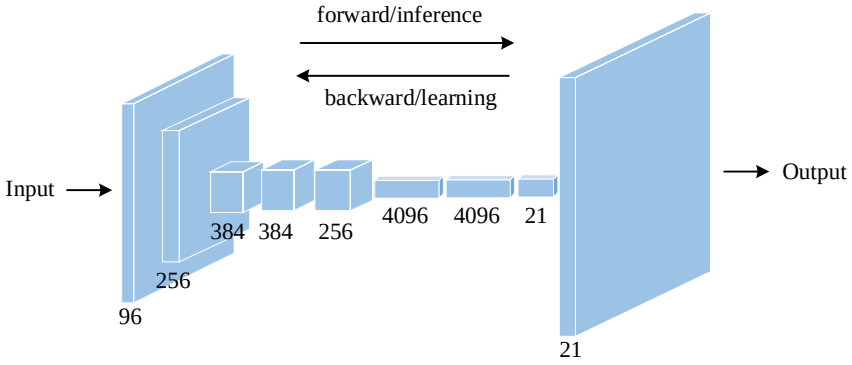

Figure 1. The architecture of the FCN model (drawn based on [7])

2.2. U-Net

To address the loss of spatial information in FCNs for medical image segmentation, U-Net [8], a variant of the fully convolutional network, was proposed. U-Net extracts high-level semantic features through an encoder and incorporates shallow spatial information during decoding via skip connections, enabling the model to maintain pixel-level accuracy while learning abstract representations [8]. These skip connections transmit low-level features from the encoder directly to the corresponding decoder layers, helping to recover fine details and boundary information [9].

his design significantly enhances spatial information preservation in U-Net, making it more accurate in medical image segmentation. This is especially true for brain tumors, where preserving boundary details in complex regions is critical. Compared to FCN, U-Net recovers spatial information more efficiently and mitigates the resolution loss caused by pooling, improving segmentation accuracy.

However, U-Net is not without limitations. Because skip connections directly link corresponding encoder and decoder layers, they may introduce redundant or irrelevant features. Moreover, feature fusion occurs only at a fixed scale, which limits U-Net's adaptability to morphologically complex structures [10]. A schematic of the U-Net architecture is presented in Fig. 2.

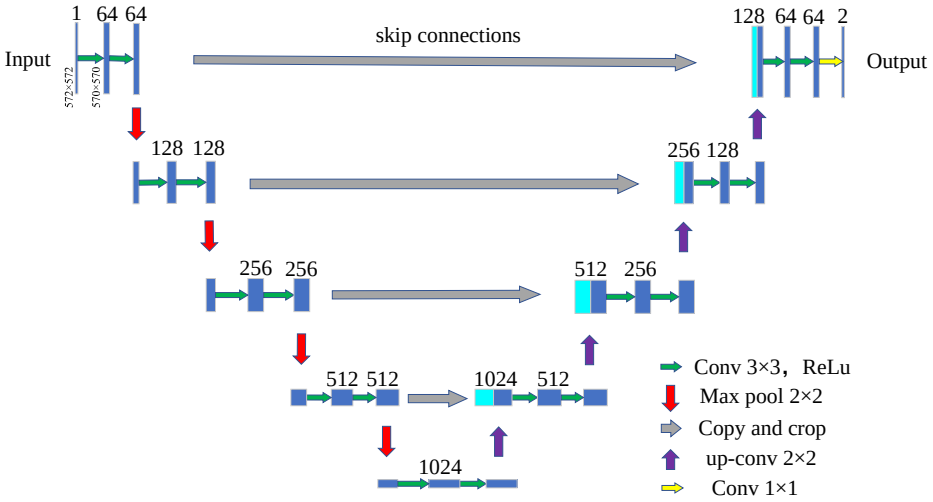

Figure 2. The architecture of the U-Net model (drawn based on [8])

2.3. U-Net++

U-Net++ [11] i is an enhanced version of U-Net, improving segmentation performance through multilevel skip connections and nested pathways. It not only builds skip connections between the encoder and decoder, but also introduces multiple nested paths across different encoding and decoding stages. Its core advantage lies in capturing both low-level detailed features and high-level semantic information, which helps preserve tumor-related features when segmenting brain tumors with complex morphology, low contrast, and ambiguous boundaries. The architectural diagram of U-Net++ is presented in Fig. 3.

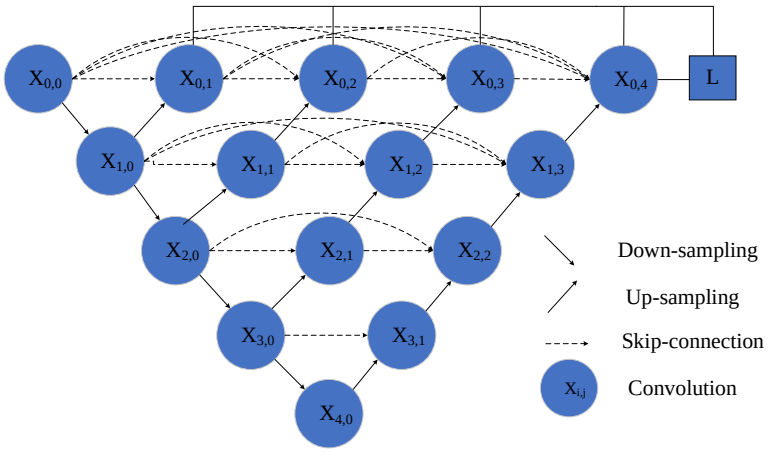

Figure 3. The architecture of the U-Net++ model (drawn based on [11])

On the BraTS 2019 dataset, U-Net++ achieved Dice coefficients of 87.12%, 78.17%, and 71.92% for the whole tumor (WT), tumor core (TC), and enhanced tumor (ET) regions, respectively. These results represent improvements of 1.37%, 3.56%, and 2.71% compared to those of U-Net [11]. The improvements in the tumor core (TC) and enhancing tumor (ET) regions are more significant.

Compared to the traditional U-Net, U-Net++ achieves higher segmentation accuracy on medical images with complex structures. However, its complex architecture incurs significant computational cost. Therefore, optimizing computational efficiency and reducing inference time while maintaining segmentation accuracy remains a key area for further investigation.

3. LKC in Medical Image Segmentation

3.1. Theoretical Advantages of LKC

3.1.1. Limitations of Fixed Kernels

U-Net++ has improved segmentation accuracy, but its fixed convolutional kernel still presents limitations when dealing with tumors with complex morphology or large-scale variations. T Traditional fixed convolution limits model flexibility and its ability to adapt to tumors of varying shapes and sizes. To address this problem, researchers have introduced more adaptive techniques, such as large kernel convolution (LKC) [13], which can overcome the limitations of fixed convolution. Large kernel convolution enables the model to recognize complex or irregular targets more efficiently. It improves segmentation accuracy while reducing computational cost, making it especially suitable for medical image segmentation tasks.

3.1.2. Receptive Field Benefits

A key advantage of large kernel convolution is expanding the receptive field, allowing the model to capture more global contextual information. Compared to traditional small-kernel convolution (e.g., 3×3), large-kernel convolution (e.g., 31×31) covers a larger area in a single pass, enabling the model to extract richer features and better understand image structures [12, 18]. In medical image segmentation, tumors often exhibit complex morphology and fuzzy boundaries, making it challenging for small-kernel convolutions to distinguish subtle differences between tumors and surrounding tissues. By expanding the receptive field, large-kernel convolution enables the model to capture more critical information, thereby improving tumor segmentation accuracy and maintaining stability across varying scales and morphologies [14].

3.1.3. Efficiency and Optimization

While large-kernel convolution enhances performance by expanding the receptive field, it also incurs substantial computational and memory overhead. When the kernel size scales to dozens or even hundreds of pixels, training and inference efficiency can decline sharply. To address this issue, researchers have developed several optimization strategies aimed at maintaining the benefits of large kernels while reducing computational cost.

One such strategy is sparse convolution, which limits computation to key regions of interest rather than processing the entire feature map uniformly [15]. Another commonly used technique is kernel decomposition, where a large kernel (e.g., 31×31) is decomposed into multiple smaller kernels (e.g., 3×3 or 5×5), significantly reducing computational complexity while retaining representational power [16]. These approaches improve the practicality of large kernel convolution, especially in resource-constrained tasks like medical image segmentation.

3.2. SLaK in Medical Image Segmentation

3.2.1. Core Techniques and Design

Building on these advancements, SLaK [17] (Sparse Large Kernel Network) reduces the computational overhead of large-kernel convolutions in medical image segmentation while preserving their global receptive field. Its core techniques, such as dynamic sparse kernel design and kernel decomposition, reduce computational burden while maintaining performance.

Its design revolves around two key innovations: dynamic sparse kernel selection and kernel decomposition. The former reduces computational cost by 40–50% by computing only the most informative weights, while the latter further lowers the burden—by up to 30%—by breaking down large kernels (e.g., 51×51) into smaller ones (e.g., 5×5 or 7×7), all without sacrificing the model's ability to learn complex spatial features.

3.2.2. Performance and Application Potential

SLaK offers strong global context modeling and demonstrates competitive performance on benchmarks like ADE20K. Its sparse large-kernel strategy not only reduces computational cost but also retains global information modeling capability, holding promise for applications in medical image segmentation, such as brain tumor segmentation on the BraTS dataset.

Compared to traditional small kernel convolution, large kernel convolution captures subtle differences between the

tumor and surrounding tissues, especially in cases of complex morphology and blurred boundaries, improving segmentation accuracy [18]. Therefore, SLaK represents a promising solution for resource-constrained scenarios that demand high segmentation accuracy (e.g., portable MRI devices and bedside ultrasound systems). Future research may further optimize its computational efficiency and segmentation accuracy to enhance its practical value in medical imaging applications.

3.2.3. Schematic Representation

Figure 4 illustrates the architectures of ConvNeXt, RepLKNet, and SLaK with large convolutional kernels (e.g., 51×51). Dark green squares represent dense weights, and light green squares denote sparse weights. These models demonstrate how combining dense and sparse kernels can enhance computational efficiency and segmentation performance.

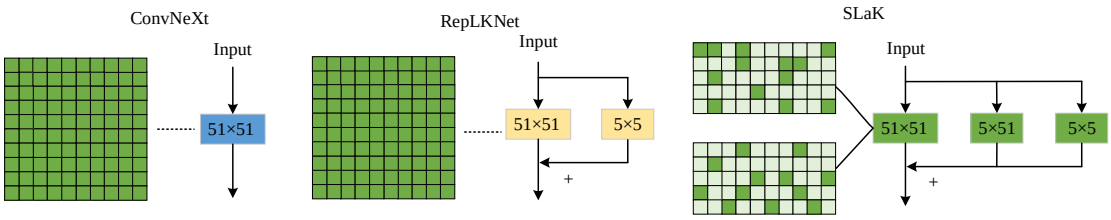

Figure 4. Large kernel convolution in ConvNeXt, RepLKNet, and SlaK (drawn based on [17])

3.3. UX-Net in 3D Segmentation

3.3.1. Architecture Overview

3D UX-Net [19] is an efficient hybrid architecture designed for 3D medical image segmentation, combining Large Kernel Convolution (LKC), Convolutional Neural Networks (CNNs), and Transformers. Compared to traditional 2D methods, 3D UX-Net employs 3D convolutions to capture volumetric information, making it ideal for medical data such as brain MRI and CT scans. In tumor segmentation tasks with complex morphology and blurred boundaries, it effectively extracts features and enhances accuracy.

3.3.2. Technical Highlights

3D UX-Net utilizes large kernel convolution (LKC) to extend the receptive field, enabling the capture of global context while integrating local features to enhance detailed segmentation accuracy. Additionally, the Swin Transformer improves computational efficiency via a window-based self-attention mechanism and strengthens long-range dependency modeling, making it well-suited for segmenting structurally complex medical images [20]. Meanwhile, 3D UX-Net integrates Depthwise Separable Convolution (DWC) with the Swin Transformer to reduce computational cost and improve inference efficiency without sacrificing accuracy.

3.3.3. Experimental Results and Analysis

Experiments on the FeTA 2021, FLARE 2021, and AMOS 2022 datasets show that 3D UX-Net outperforms traditional CNN-based methods in segmentation accuracy, efficiency, and transfer learning. It achieves a Dice score of 93.4% on the FLARE 2021 dataset, outperforming SwinUNETR by 0.5%. Additionally, the model shows a 27.8% reduction in computational cost and a 22% increase in speed. On the AMOS 2022 dataset, 3D UX-Net improves the Dice score by 2.27% compared to nnU-Net, demonstrating excellent generalization ability (Lee et al., 2022).

Figure 5 illustrates the overall architecture of 3D UX-Net, which includes the 3D UX-Net block and a downsampling module, and demonstrates how multi-layer feature extraction enhances 3D data perception. Experimental results indicate that 3D UX-Net achieves superior performance in segmenting medical images with low contrast and complex backgrounds, offering an accurate and efficient approach for 3D medical image analysis tasks such as fetal MRI and abdominal CT.

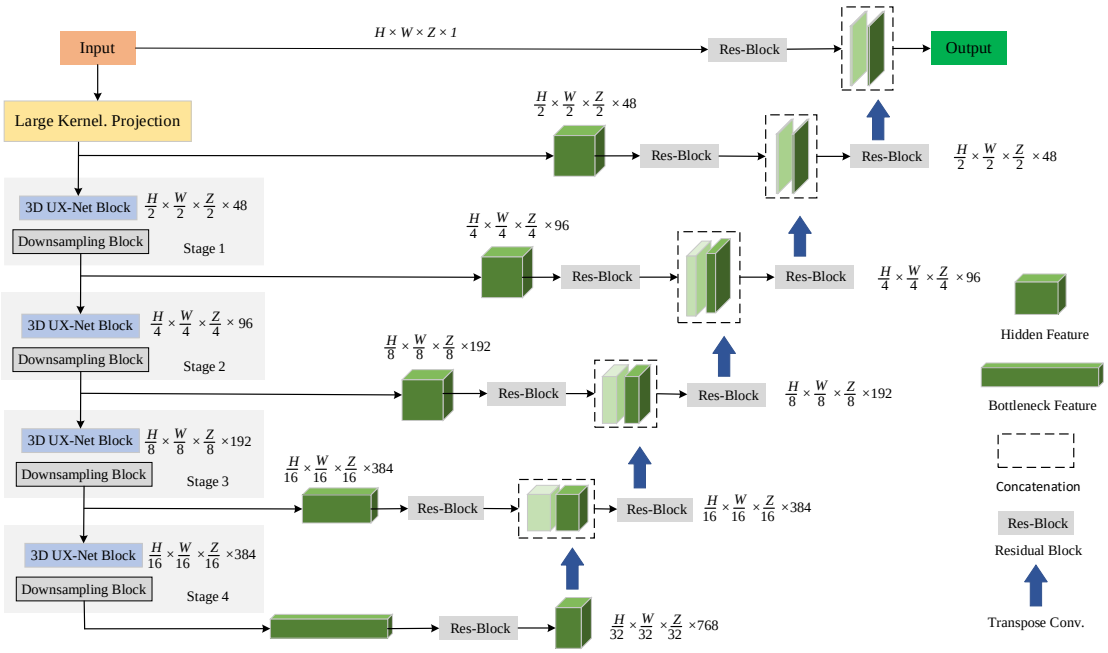

Figure 5. The architecture of 3D UX-Net model (drawn based on [19])

4. Advances in DCN and Transformer Segmentation

4.1. Evolution of DCN

4.1.1. Offset Learning

Traditional convolution operations use a fixed convolution kernel that slides over the feature map to extract local features. A key limitation of traditional convolutional operations is their poor adaptability to complex geometries, especially for targets with irregular shapes or blurred boundaries—posing challenges in medical image segmentation.

A widely used solution to this problem is Deformable Convolutional Network (DCNv1) [21], which introduces two modules to enhance CNN adaptability to deformed targets. Deformable Convolution dynamically adjusts sampling positions by learning offsets, while Deformable RoI Pooling refines the bin positions in RoI Pooling [22]. DCNv1 improves the adaptability of CNNs for deformed targets and enhances feature extraction for boundary detection and target localization.

4.1.2. Modulation Mechanism

DCNv2 [23] builds on DCNv1 by introducing a modulation mechanism that enhances the flexibility of convolutional operations and improves feature representation, while also expanding the use of deformable layers for modeling complex, deformable targets. It improves sampling precision across feature hierarchies by enabling more convolutional layers to learn offsets and further incorporates modulation to adjust both spatial position and feature amplitude for each sampling point.

This mechanism allows each sampling point to learn both spatial offsets and adaptive amplitude weights, thereby enhancing the model's flexibility and representational capacity. By refining feature extraction through modulation, DCNv2 improves the model's adaptability to multi-scale and deformable structures, thereby enhancing segmentation accuracy and generalization in medical imaging.

4.1.3. Lightweight Design

Although DCNv1 and DCNv2 significantly improve segmentation accuracy, traditional deformable convolutions still incur substantial computational overhead. DCNv3 [24] introduces minor enhancements to the offset learning strategy of DCNv2, enabling a more lightweight and efficient architecture. It reduces the computational cost to less than 1%, though memory access operations still account for 99% of the overall overhead. This observation prompted researchers to revisit the operator implementation and identify redundant memory accesses in the forward pass, which were subsequently optimized in the faster and more efficient DCNv4 [25].

DCNv4 eliminates the softmax operation, enhancing dynamic adaptability and computational performance. It serves as a dynamic sparse operator that streamlines computation by redefining deformable behavior and minimizing redundant memory access. Consequently, DCNv4 surpasses DCNv3 in speed and efficiency, demonstrating superior performance and adaptability in large-scale medical image segmentation tasks.

4.2. SDAH-UNet: A DCN-Transformer Model

4.2.1. Limitations of DCN

Although DCN has made notable progress in medical image segmentation, it still faces challenges, including high computational cost, low sensitivity to small targets, and limited generalization ability [26]. Researchers are currently exploring combinations of DCN with new techniques such as Transformers and the self-attention mechanism to handle complex structures, reduce computational overhead, and improve segmentation performance [27]. Moreover, refining DCN for better identification of diverse tumor regions and leveraging multimodal fusion in brain tumor segmentation can further enhance model stability and adaptability [28].

4.2.2. Deformable U-Net

Deformable U-Net [29] demonstrates strong performance in medical image segmentation, thanks to its combination of U-Net's encoder-decoder architecture with deformable convolution (DCN), which enhances its ability to model morphologically irregular and boundary-ambiguous structures. DCN enables the convolution kernels to adaptively adjust sampling positions based on feature information, overcoming the limitations of fixed receptive fields and improving contour accuracy. Although Deformable U-Net performs well in local feature modeling, its capacity to capture global dependencies remains limited, particularly in handling long-range structure-dependent medical images. In addition, its high computational and memory demands limit its scalability for processing high-resolution medical images and large-scale datasets.

4.2.3. SDAH-UNet

To enhance Deformable U-Net's modeling of global features, SDAH-UNet [30] ccombines Swin Transformer to improve long-range dependency modeling and reduce computational cost. The Swin Transformer employs a sliding window mechanism, which not only reduces computational overhead but also improves the model's ability to capture long-range features.

Experimental results show that SDAH-UNet achieves a Dice coefficient of 86.90% on the BraTS2020 dataset, marking a 4.63% improvement over traditional U-Net and a 3.26% improvement over nnU-Net [30]. In addition, SDAH-UNet demonstrates improved segmentation accuracy and robustness in tumor regions with low contrast and blurred boundaries, outperforming existing methods across multiple datasets.

4.2.4. Network Architecture

Figure 6 illustrates the overall architecture of SDAH-UNet. The model is based on the classical U-Net structure and combines the Self-Distilling Attention Mechanism (SDMSA) with convolutional and anti-convolutional extensions. Through skip connections and hierarchical feature extraction, the network is able to enhance the perception of global and local information while maintaining high resolution. The SDAPC module integrates deep convolution and a self-attention mechanism, further improving the model's accuracy and efficiency in medical image segmentation tasks.

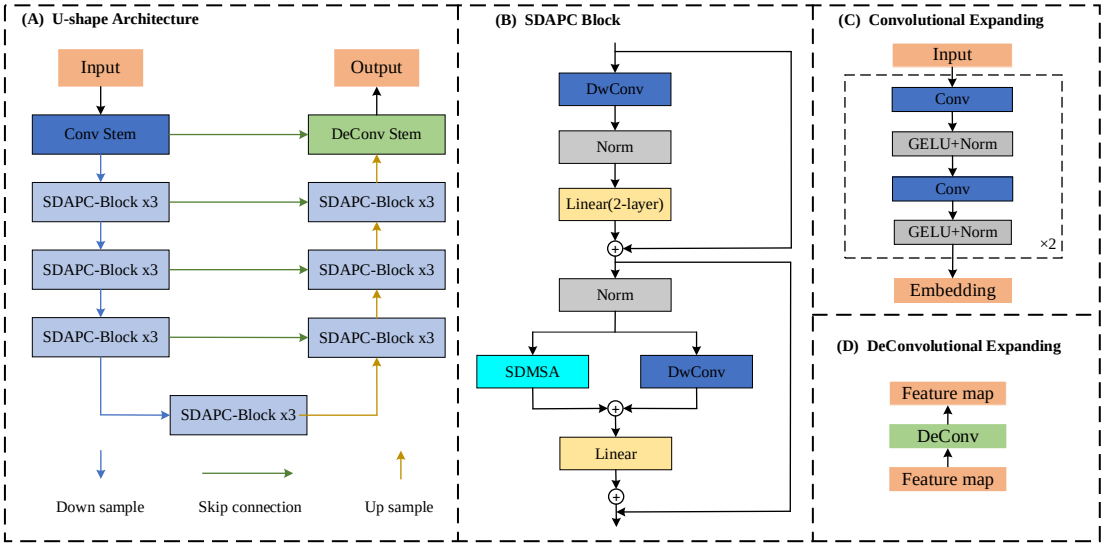

Figure 6. Overall architecture of the SDAH-UNet model (drawn based on [30])

4.3. Optimizing DCN–Transformer Integration

DCN is primarily used for local feature extraction but struggles with modeling long-range dependencies, which limits its ability to capture global structural information. To address this limitation, researchers have explored combining Transformers with DCN to enhance global feature modeling and improve computational efficiency [31]. Specifically, optimizing the sparse computation in DCN and incorporating a lightweight Transformer structure can reduce computational overhead and improve inference speed. Moreover, in medical image analysis, different modalities (e.g., MRI, CT, PET) possess unique characteristics. Designing modality-specific DCN–Transformer hybrid models improve segmentation accuracy and adaptability, enhancing their value in clinical applications [31].

5. Dataset and Performance Evaluation

5.1. Datasets

To advance brain tumor medical image segmentation, researchers have widely used the BraTS dataset series for algorithm training, validation, and evaluation. The series includes various challenges, such as the BraTS Main Task (Glioma Segmentation) and BraTS-MEN (Meningioma Segmentation Subtask).

Since 2012, BraTS has been jointly organized by multiple international research institutions to promote automated brain tumor segmentation, particularly focusing on the accurate delineation of gliomas [32]. The challenge provides multimodal MRI scans, including T1, T2, FLAIR, and contrast-enhanced T1 (T1ce) sequences, with annotations for three tumor subregions: whole tumor (WT), tumor core (TC), and enhancing tumor (ET) [33]. BraTS is evaluated using metrics such as the Dice coefficient (DSC) and Hausdorff distance (HD), and has become the preferred dataset for brain tumor segmentation research [34].

BraTS-MEN is a subtask of the BraTS challenge, focusing on the automatic segmentation of meningiomas. Unlike gliomas, meningiomas are typically benign tumors that appear as homogeneously enhanced lesions but are more challenging to segment due to their irregular morphology, blurred boundaries, and low contrast. BraTS-MEN employs multimodal MRI data, including T1, T2, and FLAIR sequences, and applies the same evaluation metrics as the main BraTS challenge [35]. Due to the characteristics of meningiomas, this subtask imposes higher demands on segmentation algorithms, advancing the development of intelligent diagnostic systems for brain tumors.

With the advancement of automated segmentation techniques, the BraTS dataset has gradually become a benchmark in the field of brain tumor segmentation. Each dataset not only advances segmentation technology but also tackles different challenges and tasks. Table 1 compares the BraTS datasets from key years, highlighting their key features and the progress made in their applications.

Table 1. Comparison of BraTS Datasets from Key Years

Dataset	Task Type	Samples	Key Features
BraTS 2015	Glioma Segmentation	~250 samples	First use of T1ce; focus on glioma
BraTS 2019	Glioma and Meningioma Segmentation	~500 samples	Added BraTS-MEN subtask; complex tumor morphology
BraTS 2021	Glioma Meningioma Radiogenomics	~500 samples	Introduced Radiogenomics; integrated genomic data
BraTS 2023	Glioma Meningioma Radiogenomics	~800 samples	Largest dataset; optimized for clinical use

5.2. Evaluation Metrics

In brain tumor segmentation tasks, commonly used evaluation metrics include the Dice coefficient, Intersection over Union (IoU), Average Symmetric Surface Distance (ASSD), Relative Volume Difference (RVD), and Hausdorff Distance (HD) [32] which are used to assess model performance. Table 2 summarizes these key metrics, presenting their descriptions, value ranges, and units for evaluating segmentation performance.

Table 2. Common Evaluation Metrics for Tumor Segmentation

Metric	Description	Range	Unit
Dice	Overlap ratio between prediction and ground truth	[0, 1]	None
IoU	Intersection over Union	[0, 1]	None
ASSD	Average symmetric surface distance	[0, ∞)	mm
RVD	Relative volume difference	(-∞, +∞)	%
HD	Max distance between boundaries	[0, ∞)	mm

6. Challenges and Advances

6.1. Data and Solutions

Medical image segmentation is advancing rapidly but still faces many challenges. Chief among them are limited sample sizes and a lack of labeled data, which hinder the performance and segmentation accuracy of deep learning models. The acquisition and annotation of medical images rely on experienced medical experts, making the process both costly and time-consuming, and limiting the overall data volume. Researchers are exploring solutions to mitigate data scarcity.

In this context, federated learning, self-supervised learning, and diffusion modeling have emerged as promising approaches. Federated learning enables data sources to train models locally without sharing raw data, preserving privacy and addressing scarcity [36]. Self-supervised learning reduces reliance on manual annotation by designing auxiliary tasks that enable models to learn features from unlabeled data [37]. Additionally, diffusion models can generate synthetic medical images, expand training datasets, and enhance model generalization [38]. Beyond data-driven strategies, multimodal data fusion is another promising approach that can improve segmentation performance, particularly for tumors with blurred boundaries or complex morphology [39].

6.2. Lightweight Models and Interpretability

Deep learning models have achieved impressive results but still face several practical challenges in real-world clinical applications. Among these, lightweight deployment is a key factor in making medical image segmentation models clinically viable [39]. Since medical devices have limited computational resources, model efficiency directly affects the speed at which clinicians can obtain diagnostic results. Techniques such as quantization, pruning, and knowledge distillation help reduce computational complexity while maintaining segmentation accuracy. These methods enable the deployment of high-performance segmentation algorithms in resource-constrained environments such as portable MRI scanners and bedside ultrasound systems.

IIn clinical settings, physicians using AI-assisted systems focus on outcomes and seek to understand the model's decision-making process [40]. As a result, uncertainty quantification and model interpretability have become critical to the clinical adoption of segmentation technologies.

6.3. Emerging Techniques

With the continued advancement of artificial intelligence, medical image segmentation is gradually integrating emerging technologies. For instance, combining 3D convolution with spatio-temporal modeling allows models to more comprehensively analyze the three-dimensional structure of tumors and their dynamic changes, thereby improving segmentation accuracy [38].

These methods also enable deeper extraction of critical information from medical images, offering enhanced support for clinical diagnosis. As these technologies mature, improvements in segmentation accuracy, computational efficiency, and clinical applicability are anticipated, driving further progress in intelligent medical imaging and precision healthcare.

7. Conclusion

Brain tumor segmentation is a crucial task in computer vision and medical image analysis. Recently, deep learning techniques have advanced rapidly, with CNN-based models (e.g., U-Net, U-Net++), large kernel convolution (LKC), deformable convolution (DCN), and CNN–Transformer hybrids demonstrating strong performance on publicly available datasets like BraTS. This review outlines the development of brain tumor segmentation, analyzes the strengths and limitations of various methods, and highlights the value of the BraTS benchmark and its sub-challenges.

However, the field still faces challenges such as limited annotated data, high computational costs, and poor model generalization. To overcome these challenges, researchers have proposed various strategies including data augmentation, self-supervised learning, multimodal fusion, model compression (e.g., pruning, quantization, knowledge distillation), and federated learning.

Looking forward, segmentation technologies are expected to become more efficient and intelligent. Techniques like multimodal fusion, lightweight deployment, and emerging methods (e.g., diffusion modeling, 3D convolution) will enhance their application in clinical diagnostics and intelligent healthcare.

Acknowledgements

Key scientific and technological projects in Henan province (242102211042)

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

References

[1] Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A Summary. Acta Neuropathol. 2016, Vol. 131 (No. 6), p. 803–820.

[2] Aboussaleh I, Riffi J, El Fazazy K, et al. 3DUV-NetR+: A 3D hybrid semantic architecture using transformers for brain tumor segmentation with multimodal MR images. Results Eng. 2024, Vol. 21, p. 101892.

[3] Bahadure NB, Ray AK, Thethi HP, et al. Comparative approach of MRI-based brain tumor segmentation and classification using genetic algorithm. J Digit Imaging. 2018, Vol. 31 (No. 4), p. 477–489.

[4] Benabid A, Yuan J, Elhassan MAM, et al. CFNet: Cross-scale fusion network for medical image segmentation. J King Saud Univ Comput Inf Sci. 2024, Vol. 36 (No. 7), p. 102123.

[5] Al-Zoghby AM, Al-Awadly EMK, Moawad A, et al. Dual Deep CNN for Tumor Brain Classification. Diagnostics. 2023, Vol. 13 (No. 12), p. 2050.

[6] Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017, Vol. 36, p. 61–78.

[7] Long J, Shelhamer E, Darrell T, et al. Fully convolutional networks for semantic segmentation. IEEE Conf Comput Vis Pattern Recognit. Boston, USA, 2015, p. 3431–3440.

[8] Ronneberger O, Fischer P, Brox T, et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. Med Image Comput Comput Assist Interv. Munich, Germany, 2015, p. 234–241.

[9] Liu Z, Tong L, Chen L, et al. Deep learning based brain tumor segmentation: a survey. Complex Intell Syst. 2022, Vol. 8 (No. 4), p. 1–27.

[10] Wang H, Cao P, Wang J, et al. UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-wise Perspective with Transformer [J]. arXiv preprint arXiv:2109.04335, 2021.

[11] Zhou Z, Siddiquee MMR, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation. Lecture Notes in Computer Science. Springer, 2018, Vol. 11045, p. 3–11.

[12] Liu Z, Mao H, Wu C-Y, et al. ConvNeXt: Revisiting Convolutions for Image Recognition. IEEE Conf Comput Vis Pattern Recognit. New Orleans, USA, 2022, p. 4929–4939.

[13] Peng C, Zhang X, Yu G, et al. Large kernel matters—improve semantic segmentation by global convolutional network [J]. arXiv preprint arXiv:1703.02719, 2017.

[14] Azad R, Niggemeier L, Hüttemann M, et al. Beyond self-attention: deformable large kernel attention for medical image segmentation [J]. arXiv preprint arXiv:2309.00121, 2023.

[15] Liu B, Wang M, Foroosh H, et al. Sparse convolutional neural networks. IEEE Conf Comput Vis Pattern Recognit. Boston, USA, 2015, p. 806–814.

[16] Jaderberg M, Vedaldi A, Zisserman A, et al. Speeding up convolutional neural networks with low rank expansions [J]. arXiv preprint arXiv:1405.3866, 2014.

[17] Liu S, Chen T, Chen X, et al. More ConvNets in the 2020s: Scaling up Kernels Beyond 51×51 Using Sparsity [J]. arXiv preprint arXiv:2207.03620, 2022.

[18] Ding X, Zhang X, Zhou Y, et al. RepLKNet: scaling up kernels beyond 51×51 with reparameterized large kernel design [J]. arXiv preprint arXiv:2203.06717, 2022.

[19] Lee HH, Bao S, Huo Y, et al. 3D UX-Net: a large kernel volumetric ConvNet modernizing hierarchical transformer for medical image segmentation [J]. arXiv preprint arXiv:2209.15076, 2022.

[20] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. IEEE Int Conf Comput Vis. Montreal, Canada, 2021, p. 10012–10022.

[21] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks. IEEE Int Conf Comput Vis. Venice, Italy, 2017, p. 764–773.

[22] Girshick R. Fast R-CNN. IEEE Int Conf Comput Vis. Santiago, Chile, 2015, p. 1440–1448.

[23] Zhu X, Hu H, Lin S, et al. Deformable ConvNets v2: more deformable, better results. IEEE Conf Comput Vis Pattern Recognit. 2019, p. 9308–9316.

[24] Wang W, Dai J, Chen Z, et al. InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions [J]. arXiv preprint arXiv:2211.05778, 2022.

[25] Zhu X, Su W, Wang L, et al. Efficient deformable ConvNets: rethinking dynamic and sparse operator for vision applications [J]. arXiv preprint arXiv:2303.11301, 2023.

[26] Zhang Y, Wang Y, Xiao X, et al. A survey on deformable convolutional networks for medical image analysis [J]. arXiv preprint arXiv:2304.12345, 2023.

[27] Chen J, Lu Y, Yu Q, et al. TransUNet: transformers make strong encoders for medical image segmentation [J]. arXiv preprint arXiv:2102.04306, 2021.

[28] Hatamizadeh A, Tang Y, Nath V, et al. UNETR: Transformers for 3D Medical Image Segmentation. IEEE/CVF Winter Conf Appl Comput Vis. Waikoloa, HI, USA, 2022, p. 574–584.

[29] Dong S, Zhao J, Zhang M, et al. DeU-Net: Deformable U-Net for 3D Cardiac MRI Video Segmentation. Med Image Comput Comput Assist Interv. Lima, Peru, 2020, p. 98–107.

[30] Wang L, Huang J, Xing X, et al. Swin deformable attention hybrid U-Net for medical image segmentation [J]. arXiv preprint arXiv:2302.14450, 2023.

[31] Guo M-H, Xu T-X, Liu J-J, et al. Attention mechanisms in computer vision: a survey [J]. arXiv preprint arXiv:2111.07624, 2021.

[32] Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BraTS) [J]. IEEE Trans Med Imaging. 2015, Vol. 34 (No. 10), p. 1993–2024.

[33] 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.

[34] Baid U, Ghodasara S, Mohan S, et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification [J]. arXiv preprint arXiv:2107.02314, 2021.

[35] LaBella D, Baid U, Khanna O, et al. Analysis of the BraTS 2023 intracranial meningioma segmentation challenge [J]. arXiv preprint arXiv:2405.09787, 2024.

[36] Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning [J]. npj Digit Med. 2020, Vol. 3, p. 119.

[37] Chaitanya K, Erdil E, Karani N, et al. Contrastive learning of global and local features for medical image segmentation with limited annotations [J]. arXiv preprint arXiv:2006.10511, 2020.

[38] Pinaya WHL, Tudosiu PD, Dafflon J, et al. Brain imaging generation with latent diffusion models [J]. arXiv preprint arXiv:2209.07162, 2022.

[39] Zhou Y, Wang Y, Wang Y, et al. Medical image segmentation: state of the art, applications, and future directions [J]. Sci Rep. 2024, Vol. 14 (No. 1), p. 58665.

[40] Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine [J]. WIREs Data Mining Knowl Discov. 2019, Vol. 9 (No. 4), p. e1312.