ISSN: 2688-8645

TransDeeplab: A Novel Architecture for Medical Image Segmentation

Zhiyu Lin [a], Lei Guo [b]

School of Electronics and Information, Southwest Minzu University, Chengdu Sichuan, 610000, China

[a] 15238507626@163.com, [b] 17695089252@163.com

Abstract

Early and accurate diagnosis of brain tumors plays a critical role in improving patient prognosis. Magnetic Resonance Imaging (MRI) serves as the core diagnostic modality; however, the inherent heterogeneity of brain tumors and the complexity of their morphological boundaries pose significant challenges for traditional segmentation methods, limiting their clinical applicability. To address this issue, we propose the TransDeepLab model, aiming to develop an efficient and precise brain tumor segmentation algorithm. The model integrates DeepLabV1 and Vision Transformer (ViT) within an Encoder–Decoder U-shaped architecture, leveraging the strengths of local feature extraction and global context modeling. Moreover, we introduce two novel components: a Feature Information Exchange Module to enhance feature fusion and a Trainable Gated Selection Module to optimize feature utilization. Using the BraTS2021 dataset and implemented within the PyTorch framework, the proposed approach is rigorously evaluated against state-of-the-art segmentation methods in terms of accuracy, robustness, and computational efficiency. The study aims to provide an automated, high-precision brain tumor segmentation tool to advance intelligent healthcare in the field of neuro-oncology.

Keywords

Brain Tumor Image Segmentation; Deep Learning; TransDeepLab Model; Vision Transformer (ViT); BraTS2021 Dataset.

1. Introduction

In the diagnosis and treatment of brain tumors, Magnetic Resonance Imaging (MRI) serves as a pivotal diagnostic tool. However, the heterogeneity of brain tumors, the complexity of their morphological boundaries, as well as noise and artifacts in MRI images, render traditional segmentation methods insufficient for meeting the high-precision requirements of clinical applications. The rapid advancement of deep learning technologies offers novel solutions to this challenge. In particular, the local feature extraction capabilities of Convolutional Neural Networks (CNNs[1]) and the global context modeling advantages of Vision Transformers (ViTs) have emerged as promising directions for enhancing brain tumor segmentation performance.

Building upon this foundation, the present study focuses on achieving **efficient and precise brain tumor image segmentation** by designing the **TransDeepLab** model, which integrates the architectures of **DeepLabV1[2]** and **Vision Transformer [3][4](ViT)**. The model adopts an **Encoder–Decoder U-shaped structure**: in the encoder stage, MRI image patches are first processed to extract **global contextual features** via ViT, followed by **multi-scale local feature extraction** using the **atrous convolutions** of DeepLabV1. The decoder then fuses multi-source features and progressively restores the image resolution. Furthermore, two novel components

ISSN: 2688-8645

are introduced—the **Feature Information Exchange Module** to enhance feature fusion and the **Trainable Gated Selection Module** to optimize the selection of salient features.

The experiments are conducted using the **BraTS2021** dataset, which provides multi-modal MRI images and thus ensures sufficient high-quality data for model training and validation. The proposed model is implemented and trained within the **PyTorch** framework, leveraging high-performance computing servers to ensure computational efficiency. A comprehensive evaluation is performed by comparing the proposed approach with mainstream segmentation methods across multiple dimensions, including **accuracy**, **robustness**, and **computational efficiency**. Ultimately, this study aims to deliver a reliable and automated brain tumor segmentation tool to facilitate the advancement of **intelligent healthcare** in the field of **neuro-oncology**.

2. Methodology

2.1. Data Processing

This study utilizes the BraTS2021 dataset[5], which comprises 1,251 training samples and 219 validation samples. Each sample corresponds to multi-modal Magnetic Resonance Imaging (MRI) scans of a single patient, encompassing four imaging modalities: T1-weighted imaging (T1), T1 contrast-enhanced imaging (T1CE), T2-weighted imaging (T2), and T2 Fluid-Attenuated Inversion Recovery (T2-FLAIR).

The segmentation task aims to identify and delineate three clinically significant tumor regions, namely the Whole Tumor (WT), the Tumor Core (TC), and the Enhancing Tumor (ET).

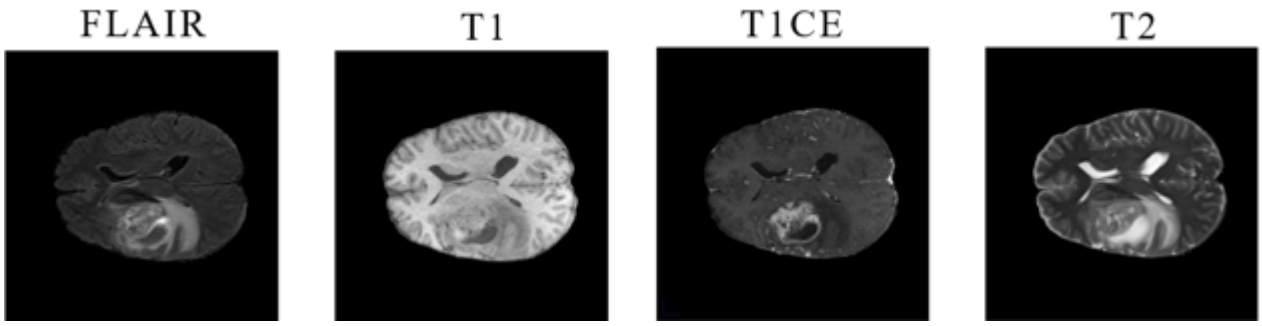

Fig 1. Sample Dataset Visualization

First, the images are normalized and resized using Z-score standardization to eliminate intensity variations across different modalities. Brain tissue is then cropped to reduce computational load. Next, data augmentation techniques—including random translation, rotation, scaling, flipping, and noise addition—are applied to increase data diversity and enhance the robustness of the model. Finally, the label data are carefully inspected and standardized to ensure consistency and accuracy.

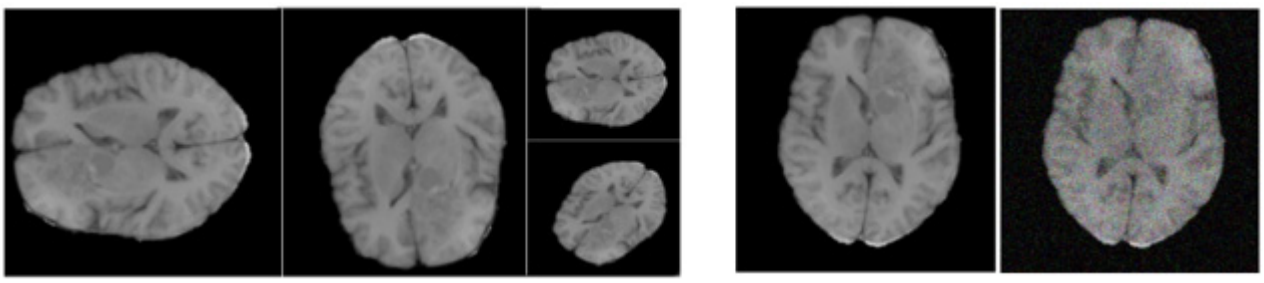

Fig 2. Data Preprocessing

ISSN: 2688-8645

2.2. Network Architecture

2.2.1. TransDeeplab Network

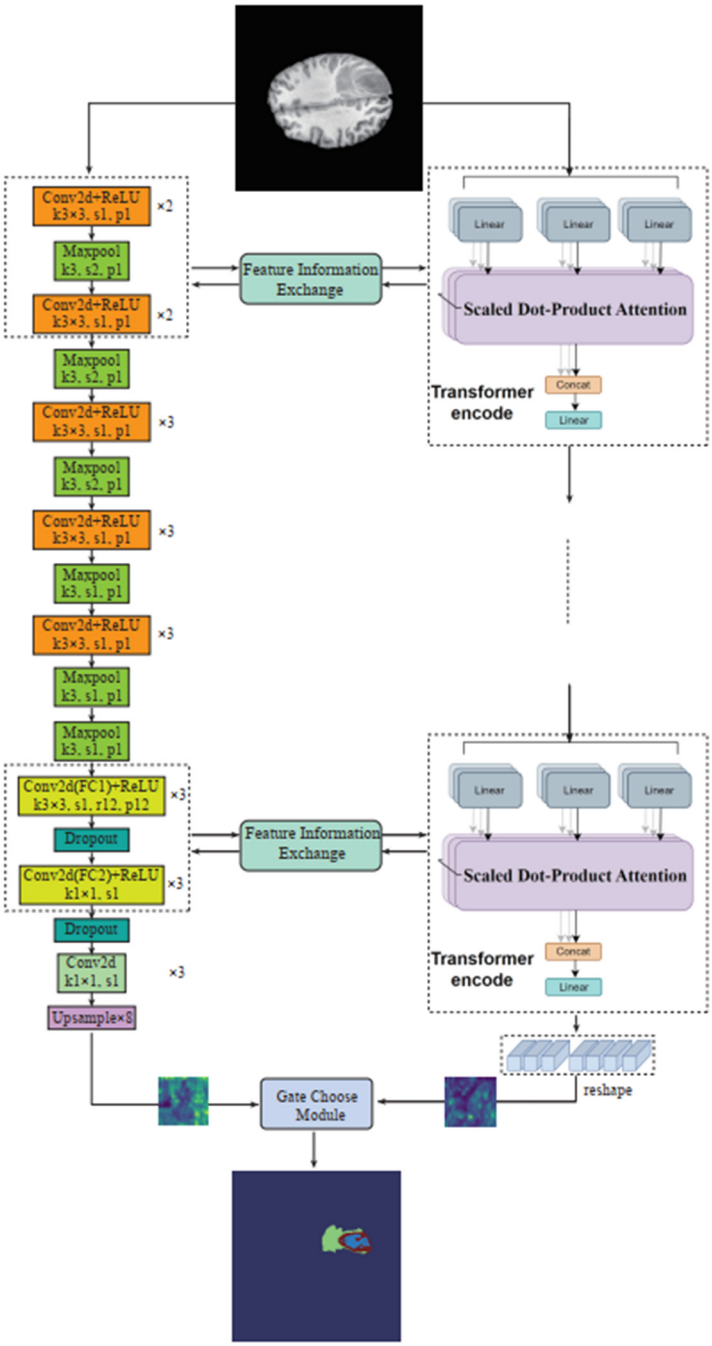

Fig 3. TransDeepLab

This study constructs an overall architecture of TransDeepLab, comprising an encoder and a decoder. The proposed model integrates DeepLabV1 with Vision Transformer (ViT), aiming to enhance the performance of brain tumor image segmentation[6].

In the encoder, the input MRI images are first divided into non-overlapping patches of size 4 × 4. Each patch is transformed into a fixed-dimensional feature representation via a linear embedding layer. These feature representations are then fed into the Vision Transformer (ViT), which leverages the self-attention mechanism to capture global information across the image, thereby effectively extracting high-level contextual features.

ISSN: 2688-8645

Following global feature extraction by ViT, the features are processed by the DeepLabV1 module. DeepLabV1 employs atrous convolution to enlarge the receptive field, enabling the extraction of richer spatial information, which is particularly beneficial for the complex and heterogeneous boundaries of brain tumors. By handling multi-scale image features, DeepLabV1 further improves segmentation accuracy.

In the decoder, a symmetric structure is designed to integrate the output features from both ViT and DeepLabV1. Skip connections are employed to fuse multi-scale features[7] from the encoder with the upsampled features in the decoder, thereby restoring detailed spatial information. The upsampling operation in the decoder is implemented using a specialized patch expansion layer, gradually recovering the resolution until it matches the original input size. The final output of the model is generated via a linear projection layer, producing pixel-level segmentation predictions.

By combining ViT and DeepLabV1, the model maintains the capability of global feature learning while effectively capturing local details, thereby achieving higher accuracy and robustness in brain tumor image segmentation tasks.

2.2.2. FIEM(Feature Information Exchange Module)

In the TransDeepLab model, a Feature Information Exchange Module is introduced to enhance the effectiveness of feature selection. This module facilitates the exchange of features extracted by DeepLabV1 and the Vision Transformer (ViT), enabling more effective fusion of salient features. As a result, the feature maps learned by the model can better represent the information of the target regions, thereby improving the overall performance of the model.

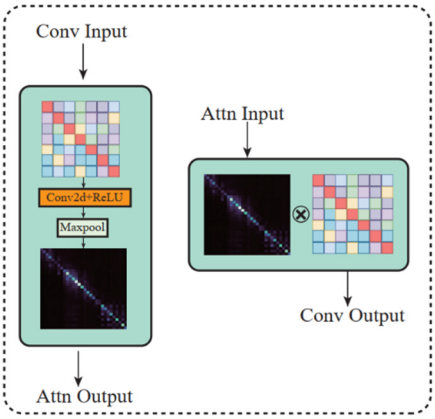

Fig 4. Feature Information Exchange Module

2.2.3. GSM(Gated Selection Module)

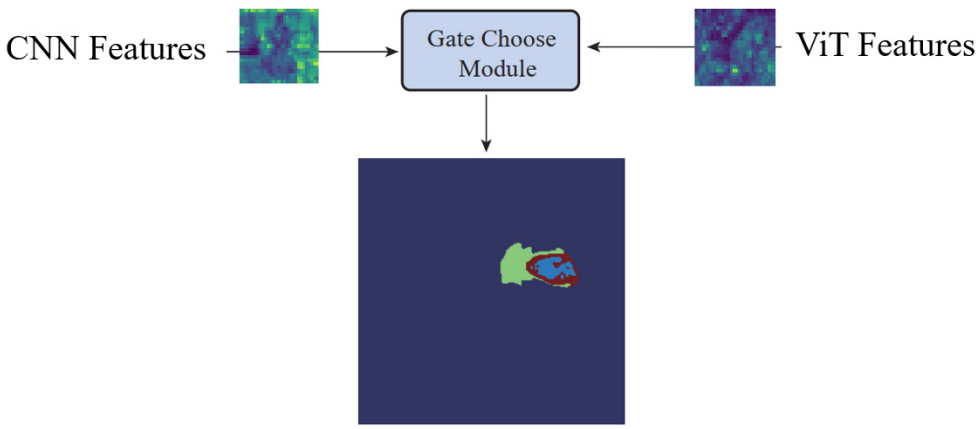

Fig 5. Gated Selection Module

ISSN: 2688-8645

To more effectively extract the most representative features, a trainable Gated Selection Module is introduced. This module automatically selects the optimal features from the outputs of the Convolutional Neural Network (CNN) and the Vision Transformer (ViT), thereby enhancing the model's ability to capture and represent critical information.

3. Experiments

3.1. Experimental Environment

The experiments were conducted on a system equipped with an **Intel(R) Core(TM) i5-10200H CPU** and an **NVIDIA GeForce RTX 3060 Laptop GPU**, running **Windows**. The implementation was based on the **PyTorch** framework, utilizing **CUDA 11.1** and **cuDNN**, combined with **Python** for algorithm development. The system configuration is summarized in **Table 1**.

Table 1. Hardware Configuration

Component	Model
CPU	Intel(R)_Core (TM)_i5-10200H_CPU
Memory	16GB
GPU	NVIDIA GeForce RTX 3060 Laptop GPU
GPU Memory	6GB

3.2. Evaluation Metrics

To evaluate the performance of **TransDeepLab** on brain tumor segmentation tasks, we adopt several commonly used metrics, including **Dice (DSC)**, **Area Under the Curve (AUC)**, **Jaccard Index (Jac, also known as IoU)**, **Precision (Prec)**, **Recall (Rec)**, and **Average Symmetric Surface Distance (ASSD)**.

Dice Coefficient (DSC) measures the similarity between two samples, ranging from 0 to 1, where a higher value indicates greater similarity. **Jaccard Index (Jac/IoU)** quantifies the consistency between the segmentation result and the ground truth by computing the ratio of the intersection over the union of the two, with values closer to 1 indicating more accurate segmentation.

Precision (Prec) evaluates the proportion of true positive pixels (correctly predicted tumor pixels) among all pixels predicted as tumor, reflecting the "purity" of the segmentation result. Values range from 0 to 1, with higher values indicating fewer false positives (normal tissue misclassified as tumor). **Recall (Rec)** measures the proportion of true positive pixels among all actual tumor pixels, reflecting the "completeness" of the segmentation. Higher values indicate fewer false negatives (tumor pixels misclassified as normal tissue).

Average Symmetric Surface Distance (ASSD) calculates the mean distance between corresponding points on the surfaces of the predicted segmentation and the ground truth, expressed in pixels. Lower values indicate that the predicted boundaries are closer to the true boundaries, representing higher boundary segmentation accuracy.

Area Under the Curve (AUC) is used in binary classification tasks. The **Receiver Operating Characteristic (ROC)** curve is plotted with the false positive rate (FPR) on the x-axis and the true positive rate (TPR, i.e., Recall) on the y-axis. The AUC represents the area under the ROC curve, ranging from 0.5 to 1, with values closer to 1 indicating stronger discrimination between tumor and normal tissue.

ISSN: 2688-8645

3.3. Experimental Results

3.3.1. Ablation Study

The dataset used in this study is the BraTS 2021 MRI dataset for brain gliomas, provided by the Medical Image Computing and Computer Assisted Intervention (MICCAI) society. To evaluate the effectiveness and scientific validity of the different modules, we conducted comparative experiments. The results are summarized in Table 2.

Table 2. Comparative Experiments of Different Modules in TransDeepLab

Model	Dataset	DICE	AUC	ASSD	Jac
Basline	BraTS 2021	0.4617	0.8718	6.584	0.3291
Baseline+FIEM	BraTS 2021	0.5233	0.8803	6.575	0.3826
Baseline+FIEM+GSM	BraTS 2021	0.6662	0.9367	7.681	0.4172

3.3.2. Comparative Experiments

To more intuitively demonstrate the performance of **TransDeepLab**, we trained **DeepLabV1** and the **Vision Transformer (ViT)** separately on the **BraTS 2021** dataset. Their results across different evaluation metrics are presented in **Table 3**.

Table 3. Model Comparison Experiments

Model	Dataset	DICE	AUC	Jac	Prec	Rec	ASSD
Deeplab V1	BraTS 2021	0.4606	0.8997	0.3269	0.6033	0.5412	5.9654
VIT	BraTS 2021	0.5173	0.9289	0.3883	0.4204	0.5137	7.5324
TransDeeplab	BraTS 2021	0.6662	0.9399	0.4548	0.6109	0.6012	7.6821

4. Summary

This study is based on the **BraTS 2021** dataset and proposes the **TransDeepLab** model. The model adopts an **encoder-decoder U-shaped architecture**, where the encoder first partitions the MRI images into 4 × 4 patches and transforms them via a linear embedding. The **Vision Transformer (ViT)** is then employed to capture global contextual features, followed by **DeepLabV1** with atrous convolution to extract multi-scale local spatial features. In the decoder, skip connections are used to fuse multi-source features, and the resolution is progressively restored through a patch expansion layer, ultimately producing pixel-level segmentation outputs.

Furthermore, the model innovatively incorporates a **Feature Information Exchange Module**, which enhances the fusion of salient features, and a trainable **Gated Selection Module**, which automatically selects the optimal features from CNN and ViT outputs. These modules improve feature utilization efficiency and the representation of critical information.

The model is implemented using the **PyTorch** framework and trained on a high-performance server to ensure computational efficiency. To evaluate the performance of TransDeepLab, comparative experiments are conducted against traditional segmentation methods and mainstream deep learning models, such as single DeepLab, U-Net, and ViT. Performance is assessed in terms of **segmentation accuracy** (e.g., Dice, Jaccard), **robustness** (noise and artifact resistance), and **computational efficiency**. Experimental results on the BraTS 2021 dataset demonstrate that the proposed TransDeepLab network achieves superior performance in brain tumor segmentation.

ISSN: 2688-8645

References

[1] Li G, Muller M, Thabet A, et al. DeepGCNs: Can GCNs go as deep as CNNs?[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9267-9276.

[2] Bagheri F, Tarokh M J, Ziaratban M. Skin lesion segmentation from dermoscopic images by using Mask R-CNN, Retina-Deeplab, and graph-based methods[J]. Biomedical Signal Processing and Control, 2021, 67: 102533.Information on: www.ISSSconf.org.

[3] Han K, Wang Y, Chen H, et al. A survey on vision transformer[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 45(1): 87-110.

[4] Khan S, Naseer M, Hayat M, et al. Transformers in vision: A survey[J]. ACM computing surveys (CSUR), 2022, 54(10s): 1-41.

[5] 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.

[6] Işın A, Direkoğlu C, Şah M. Review of MRI-based brain tumor image segmentation using deep learning methods[J]. Procedia computer science, 2016, 102: 317-324.

[7] Gao S H, Cheng M M, Zhao K, et al. Res2net: A new multi-scale backbone architecture[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(2): 652-662.