frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 March 2022
doi: 10.3389/fnins.2022.832824

Residual Block Based Nested U-Type Architecture for Multi-Modal Brain Tumor Image Segmentation

Sirui Chen[1], *Shengjie Zhao*[1]* *and Quan Lan*[2]*

[1] School of Software Engineering, Tongji University, Shanghai, China, [2] Department of Neurology, First Affiliated Hospital of Xiamen University, Xiamen, China

Multi-modal magnetic resonance imaging (MRI) segmentation of brain tumors is a hot topic in brain tumor processing research in recent years, which can make full use of the feature information of different modalities in MRI images, so that tumors can be segmented more effectively. In this article, convolutional neural networks (CNN) is used as a tool to improve the efficiency and effectiveness of segmentation. Based on this, Dense-ResUNet, a multi-modal MRI image segmentation model for brain tumors is created. The Dense-ResUNet consists of a series of nested dense convolutional blocks and a U-Net shaped model with residual connections. The nested dense convolutional blocks can bridge the semantic disparity between the feature maps of the encoder and decoder before fusion and make full use of different levels of features. The residual blocks and skip connection can extract pixel information from the image and skip the link to solve the traditional deep traditional CNN network problem. The experiment results show that our Dense-ResUNet can effectively help to extract the brain tumor and has great clinical research and application value.

Keywords: brain tumor, multi-modal image segmentation, MRI, CNN, ResNet, UNet

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jing Xue,
Wuxi People's Hospital Affiliated to Nanjing Medical University, China
Zhifan Gao,
Sun Yat-sen University, China

***Correspondence:**
Shengjie Zhao
shengjiezhao@tongji.edu.cn
Quan Lan
xmdylanquan@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *10 December 2021*
Accepted: *03 February 2022*
Published: *09 March 2022*

Citation:
Chen S, Zhao S and Lan Q (2022) Residual Block Based Nested U-Type Architecture for Multi-Modal Brain Tumor Image Segmentation. Front. Neurosci. 16:832824. doi: 10.3389/fnins.2022.832824

1. INTRODUCTION

Tumors are localized cell growths in the body that constitute a tumor for various reasons. There are two classifications of tumors, benign and malignant. Cancer is usually a malignant tumor, and is a common malignant diseases that threaten human health. Therefore, accurate segmentation and subsequent quantitative analysis of tumor images is a routine and critical step in treatment.

MRI is based on the principle that the electromagnetic signal released by the instrument has different attenuation in different structures in the body to obtain the electromagnetic signal from the body and reconstruct the body information (Vaughan et al., 2006). Compared with other traditional medical imaging techniques such as X-ray and CT imaging, MRI can provide early detection of smaller and more microscopic lesions (Nandpuru et al., 2014). Due to the complexity of human organ tissues, the same imaging technique usually results in images of different modalities, where different modalities also reveal different pathological information (Legg et al., 2015). MRI provides four different imaging sequences obtained by several different imaging display techniques, and these four different imaging sequences can provide complementary information for clinical diagnosis.

At present, the segmentation of MRI brain tumor images is mainly based on the experience of expert doctors, but it is difficult to accurately segment MRI brain tumor images preoperatively

because of the time-consuming and repetitive work, and the subjective judgment of the physician can interfere with it. In recent years, many researchers have improved natural image segmentation algorithms and then migrated them to brain tumor segmentation tasks, resulting in a large number of effective research methods and findings. As shown in **Figure 1**, we can classify them into two categories: (i) Fully automated segmentation methods based on deep learning and (ii) Traditional semi-automatic medical image segmentation methods.

Edge-based detection: The basic idea of the edge detection-based image segmentation method is to first find its edge pixels in the image and then aggregate these pixels to form the desired region edges and segment the image with the region edges. Aslam et al. (2015) proposed an improved edge detection algorithm for brain tumor segmentation based on the operator (Sobel, 1970) of the first-order derivatives, which combined the Sobel method and the image dependent thresholding method and used a closed contour algorithm to find the different regions. Finally, the intensity information within the closed contours was used to extract the tumor from the image.

Threshold-based: The idea of the algorithm is that first, the grayscale features of the image are computed to obtain one or more grayscale thresholds, and then the grayscale values of each pixel in the image are compared with each other and with the computed thresholds. Finally, based on the comparison results, the pixels can be classified into different classes. Kittler and Illingworth (1986) proposed a computationally efficient solution to the minimum error thresholding problem applicable to multiple threshold selection. Saad et al. (2010) used the histogram thresholding technique in segmentation to detect pixels with high or low intensity.

Region-based: Region-based image segmentation algorithms can be divided into two basic forms: (i) The global departure form, where segmentation is performed gradually until the desired region and (ii) The region growth form, where individual image pixels are started and gradually merged, which in turn forms the segmented region we need. Kaus et al. (2001) used a region-growing approach to segment MRI brain tumor images based on signal intensity values.

Since the main idea of traditional segmentation algorithms is to start from features such as texture and grayscale values between different tissues in medical images, the main challenges of this type of algorithms are the large grayscale similarity and uneven distribution between brain tissues in MR images, while the grayscale values contain too little information and the variability between different cases, which can affect the final segmentation accuracy.

Machine learning-based: Machine learning-based image segmentation algorithms require training a model from a certain number of images with well-defined mapping relationships between image features and labels to learn the segmentation laws. Weakly supervised learning methods and semi-supervised learning methods were first developed, mainly random forests (Shi and Horvath, 2005), Adaboost (Rtsch et al., 2001), K-Means clustering (Hartigan and Wong, 2013), support vector machines (SVMs) (Zhang et al., 2006), etc. The method proposed by Abdelmaksoud et al. (2015) can not only take advantage of the fuzzy C-means algorithm to obtain high accuracy, but also make full use of K-Means clustering to obtain the minimum computation time, which decreased the segmentation time while improving the segmentation accuracy. Tustison et al. (2014) combined the random forest model with regularized probability and used the probability map generated by this model for Markov random field, which finally obtained better results for brain tumor segmentation. Mahmood and Basit (2016) proposed an automatic segmentation framework for ischemic stroke lesion segmentation in multi-spectral MRI based on random forest. However, the problem with the machine learning-based approach is that the selection and labeling of image features require the use of specialized medical knowledge, which limits the development of such methods.

CNNs were originally used in areas such as handwritten alphabet classification, but the input of such images was already fixed in size, so when migrating such networks to the field of image segmentation, it also started by classifying blocks of

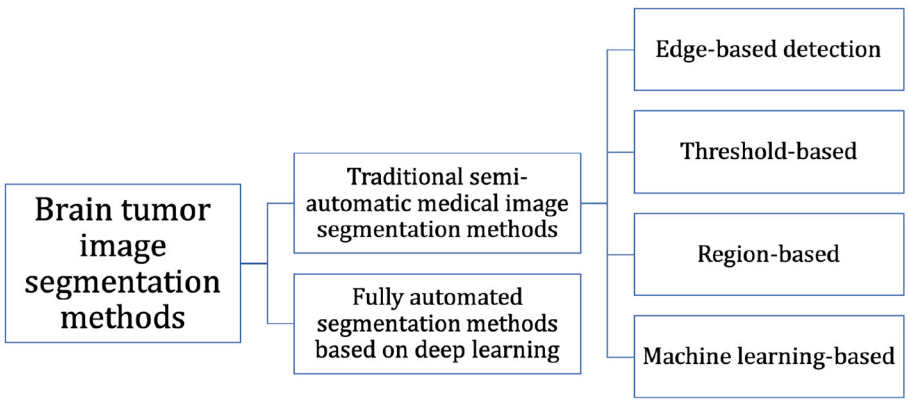

FIGURE 1 | Classification of brain tumor image segmentation methods.

images of fixed size. Moeskops et al. (2016) came up with a method to automatically segment MRI brain images. Their network obtained multi-scale information for each voxel, using multiple convolutional kernels of different sizes. The method did not rely on explicit features but learns to identify important information about the classification based on the training data. However, this method was used only for a single modality MRI image. Hou et al. (2016) proposed a segmentation network that cut a complete image into fixed-size image blocks, after which they were convolved separately to extract features and finally classified for the central pixel points to obtain the final segmentation results. Their method focused the classification from the image level to the pixel level. Brosch (2015) proposed a 3D deep neural network, this model had both convolutional and deconvolutional layers and combined feature extraction and segmentation prediction, this method has high accuracy but was too computationally intensive.

The high cost of acquiring medical images and the relatively small dataset for medical image segmentation also pose a problem for deep learning that requires large amounts of data. And due to the layer-by-layer convolution of neural networks, the accuracy of segmentation is reduced while increasing the computational effort. Multi-modal MRI brain tumor image segmentation can make full use of the feature information of different modalities in MRI thus improve the effectiveness of segmentation, which is one of the recent research hotspots in the field of brain tumor image processing. Using artificial intelligence technology, a fully automated method for brain tumor image segmentation can be designed so that it can support physicians' analysis and diagnosis, which is the best way to solve the above problem.

In general, this article focuses attention on how to fully and efficiently utilize multi-modal MRI image information, proposes a multi-modal nested dense ResU-Net segmentation network, and validates the effectiveness of the algorithm on an open-source dataset of brain tumors.

2. MATERIALS AND METHODS

The model proposed in this section, given the abbreviation Dense-ResUNet, is a multi-modal MRI brain tumor segmentation model based on the traditional U-Net (Ronneberger et al., 2015) and U-Net++ (Zhou et al., 2018), and the network structure is shown in **Figure 2**. The left is the downsampling path, which effectively extracts the image features through a series of successive convolution and downsampling operations, and reduces the image size while increasing the number of channels; the right is the upsampling path, which successfully recovers the image size, improves the image segmentation accuracy and allows better reconstruction of details through a series of successive transposition convolution operations; the middle is a series of nested dense convolutional blocks. The structure bridges the semantic disparity between the feature maps of the encoder and decoder before fusion. Based on this modular design, our model can train the network by adding a small number of convolutional layers to extract good medical image features.

The shalllow convolution structure cannot fully capture the complex structure of the image, but the deep convolution and redundant structure of the stack lead to the gradient vanishing and tearing problem. Therefore, we use a residual unit to extract pixel information from the image and skip the link to solve the traditional deep CNN network problem. The structure of the residual unit is shown in **Figure 3** and the specific embedding method is shown in **Figure 5**. Each layer in the constructed Dense-ResUNet model needs to use convolutional blocks so that feature extraction can be performed. Each of these convolutional blocks consists of 2 convolutional units. In each of these 2 convolutional units, a BatchNorm (BN) layer is included to improve model convergence. The activation is then performed with the ReLU function, which is used to improve the non-linearity of the function.

2.1. ResNet Architecture Overview

Each of the convolutional layers contains feature maps. When a pixel in the image is scanned by the convolution kernel and makes full use of the complex content information in the environment to generate semantic image features, the image features with content and space are expressed by the activation function (ReLU) as Equation (1):

$$X_j^{l+1} = f(t_i^{l+1} + \sum_{i \in I_j} X_i^l \divideontimes k_{ij}^{l+1}) \tag{1}$$

Where $X_j{}^{l+1}$ represents the feature map after the $(l+1)-th$ layer, t represents the offset term, $X_i{}^l$ represents the input feature in the $(l+1)-th$ layer, f is the activation function (rectifier linear unit, ReLU), I_j represents a series of input eigenmatrices, $\divideontimes$ represents the convolution operation, k represents the convolution kernel.

By reducing the dimension of the image features, the pooling layer can represent the high-level content and semantics:

$$X_j^{l+1} = t_j^{l+1} + X_j^l \circledast k_j^{l+1} \tag{2}$$

Where $\circledast$ refers to pooling operations in the convolutional structure. Finally, the fully connected layer takes the maximum likelihood function as the prediction layer to carry out the image classification task.

Figure 4 illustrates how a building block works in ResNet. A quick connection mechanism between each initial input X and the module output $H(X)$ is introduced, so that the input can learn the residual expression $F(X) = H(X) - x$ directly to model the target output $[F(X) + X]$. With such a mechanism, the precision problem and performance degradation due to too many stacked convolution structures can be avoided, and such a quick connection mechanism can perform identity mapping in multi-layer structures. It serves as a reference for the input elements in each layer and learns to form the corresponding residual function instead of some function blocks with no practical value $[F'(X)]$. It is easier to optimize the propagation using such a mapping method and thus can significantly increase the number of layers

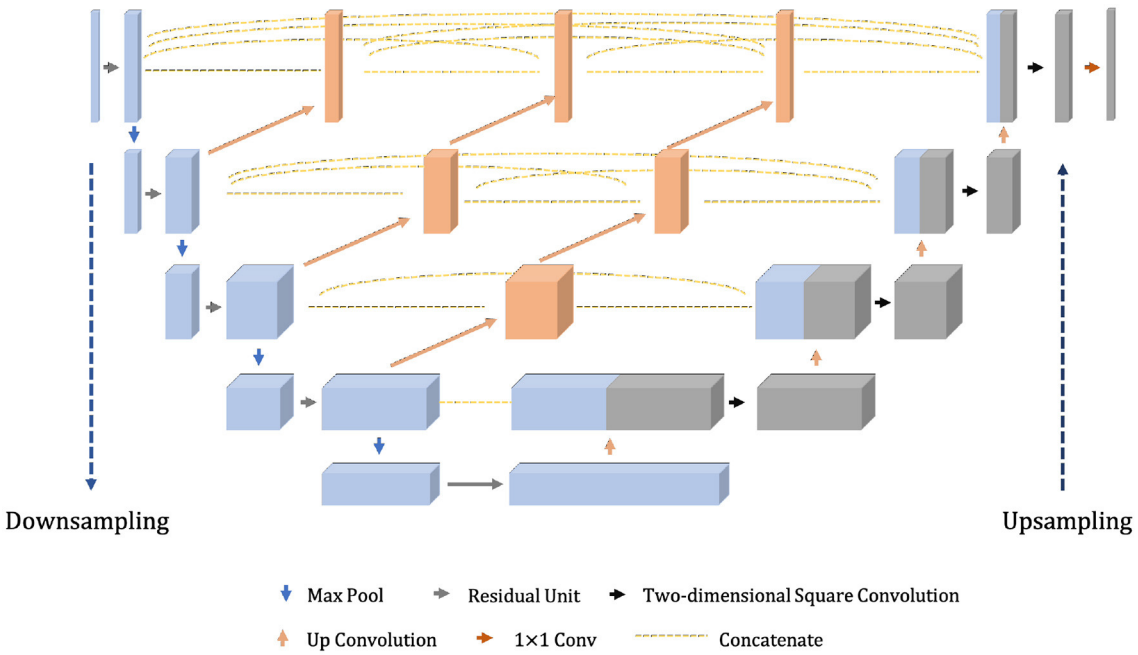

FIGURE 2 | The overall structure of our proposed Dense-ResUNet.

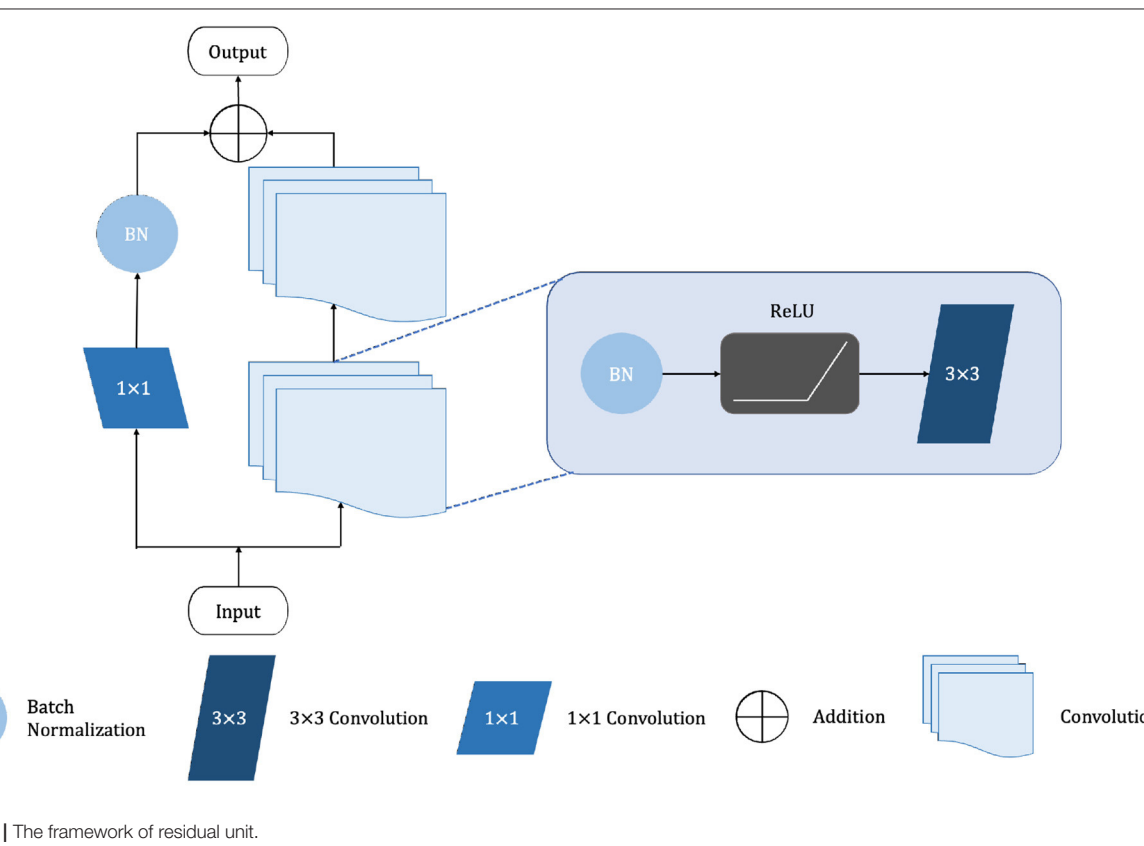

FIGURE 3 | The framework of residual unit.

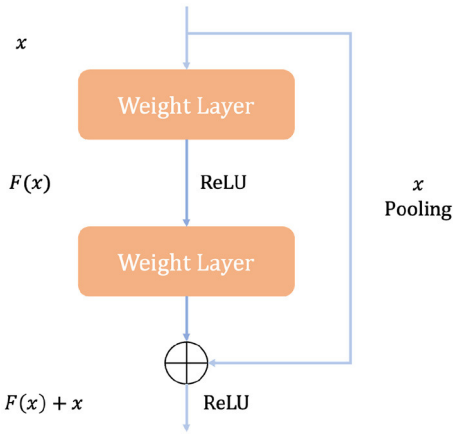

FIGURE 4 | The basic block of ResNet.

in the network structure. The formula for the residual mapping function is as follows:

$$F = W_2 f(W_1 x) \tag{3}$$

$$y = F(x, W_i) + W_s x \tag{4}$$

Where $f(\cdot)$ represents the ReLu activation function and then generates the corresponding output y through a Shortcut Connection and the second ReLU function. When we switch the dimensions of the input and output representations, we can use W_s for linear transformation.

2.2. UNet Architecture Overview

Medical images usually have small samples and most medical images have complex modalities, and the U-Net model proposed by Ronneberger et al. (2015) has an excellent performance in processing medical images with small samples due to its excellent learning ability. The main idea of U-Net is to connect a network with a similar structure to the previous layer behind the downsampling stage, and then recover the resolution of the image output by applying upsampling and combining the output of upsampling with downsampling having high-resolution features combined in these ways to better extract the edge features of the image. **Figure 5** demonstrates the network framework.

Suppose F_{mn} denotes the convolution operation performed at position (m, n), the size of the two-dimensional convolution kernel is $a \times b$, and the two-dimensional convolution is calculated as Equation (5).

$$F_{mn} = f(s_{mn} + \sum_{i=0}^{a-1} \sum_{j=0}^{b-1} w_{ij} \cdot X_{(m+i)(n+j)}) \tag{5}$$

Where s refers to stride, w refers to convolution kernel and X is the input.

From **Figure 2**, we can see that the feature maps of the encoder go through a dense convolution block. The convolution layers' number of the blocks depends on the pyramid level. We assume that X^{l_d,l_c} is a node in the model, where l_d refers to the downsampling layer along the encoder and l_c refers to the convolution layer of the dense block along the skip connection. Meanwhile, we define x^{l_d,l_c} as the output of X^{l_d,l_c}, then the x^{l_d,l_c} can represent the feature maps as Equation (6).

$$x^{l_d,l_c} = \begin{cases} \Delta(x^{l_d-1,l_c}), & where \quad l_c = 0 \\ \Delta([[x^{l_d,l_k}]_{l_k=0}^{l_c-1}, \mathcal{F}(x^{l_d+1,l_c-1})]), & where \quad l_c > 0 \end{cases} \tag{6}$$

Where $\Delta(\cdot)$ refers to the convolution operation followed by ReLU, $\mathcal{F}(\cdot)$ refers to the upsampling operation and [] refers to the concatenate operation.

The role of the pooling layer is to perform merge operations on the input data. We conducted the maximum pooling approach in this article. We calculate the height and weight of the output according to Equations (7) and (8).

$$H_{out} = \lfloor \frac{H_{in} + 2 \times p_i - d_i \times (k_i - 1) - 1}{s_i} + 1 \rfloor \tag{7}$$

$$W_{out} = \lfloor \frac{W_{in} + 2 \times p_j - d_j \times (k_j - 1) - 1}{s_j} + 1 \rfloor \tag{8}$$

Where p refers to the padding, d refers to the dilation, H refers to height and W refers to weight.

The objective function used in the network proposed in this article is dice loss, which is a concept first proposed by Milletari et al. (2016). Its functional expression is shown in Equation (9) and the function takes values in the range [0,1].

$$D = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{9}$$

where p_i is the pixel value of point i in the prediction result and g_i is the pixel value of point i in the label, resulting in the gradient equation as in Equation (10).

$$\frac{\partial D}{\partial p_i} = 2[\frac{g_i \sum_i^N p_i^2 + \sum_i^N g_i^2 - 2p_i \sum_i^N p_i g_i}{(\sum_i^N p_i^2 + \sum_i^N g_i^2)^2}] \tag{10}$$

2.3. Multi-Modal MRI

MRI is performed by varying the direction and intensity of the magnetic field thus obtaining different imaging sequences, which is called multi-modal. Four main imaging forms of multi-modal usually exist, namely: T1-weighted form, T1c-weighted form, T2-weighted form, and FLAIR-weighted form. The four modalities are shown in order from left to right in **Figure 6**, and these pictures also show that the brain tumor image has complex texture and obvious structural information. The imaging of different modalities can highlight different characteristic information of the tumor. By deeply analyzing and thinking about multi-modal images, we can obtain more comprehensive information of brain tumor images and provide help to the research related to brain tumor lesion regions.

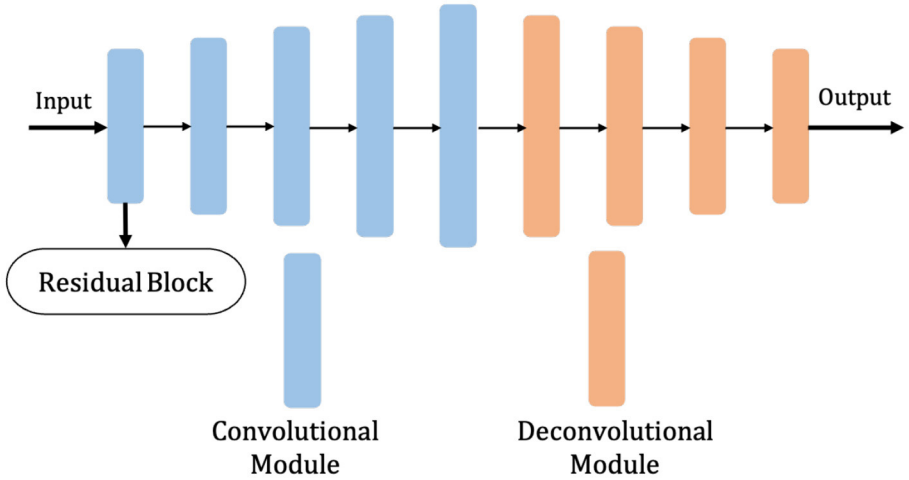

FIGURE 5 | Dense-ResUNet model framework diagram.

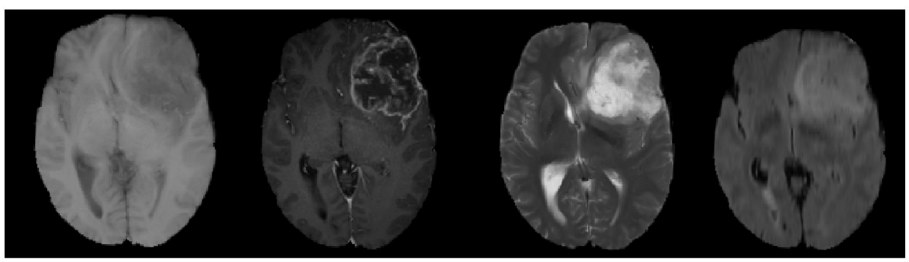

FIGURE 6 | MRI images of four different modalities.

2.4. Datasets

We use BraTS 2018 dataset to train the modal, and it consists of high-grade gliomas (HGG), low-grade gliomas (LGG), and real known tumor segments repeatedly hand-drawn by several medical experts. Four MRI sequences were performed in each case. All imaging sequences of the same subject were preprocessed and stripped from the skull, and all 3D MRI images were volume sized 240 × 240 × 155 and they were aligned, interpolated, and rescaled to achieve the same resolution ($1mm^3$). A total of five types of intra-tumor structures were available in the dataset: (i) normal tissue; (ii) edematous regions (Peritumoral Edema, ED); (iii) necrosis (NCR); (iv) Non-enhancing Tumor (NET); and (v) Enhancing Tumor (ET).

2.5. Multi-Modal Fusion of Brain Tumor Images

Multi-modal brain tumor image fusion is the synthesis of multiple images into a new image. The fusion result is more beneficial to human recognition and automatic machine detection due to the ability to exploit the spatio-temporal correlation and information complementarity of multiple brain tumor images, and to make the fused image obtained a more comprehensive and clear description of the scene.

Our proposed method performs pixel-level image fusion of images and constructs multi-modal image channels, as show in Equation (11). The image obtained after pixel-level image fusion preserves the original information to the maximum extent, which is conducive to further analysis, processing, and understanding of the image, and is also able to expose potential targets, facilitating the operation of judgment to identify potential target pixel points, which is the only way to preserve as much information as possible in the source image, making the fused image increased in both content and detail, an advantage that is unique and exists only in pixel-level fusion.

$$F_{COMB}^{(i)} = F_{Flair}^{(i)} + F_{T1}^{(i)} + F_{T1c}^{(i)} + F_{T2}^{(i)} \tag{11}$$

3. EXPERIMENTAL CONFIGURATIONS

3.1. Evaluation Metrics

In this article, the evaluation metrics proposed by MICCAI BraTS are used to evaluate the performance of the proposed segmentation method. The dataset MICCAI BraTS provides five types of labels for training the network. Different structures in the five types of labels are divided into three regions to meet the medical clinical applications in real situations, and the evaluation

method described in this chapter also focuses on these three tumor regions.

(i) **Enhancing tumor (ET)**, which is present only in high-grade gliomas and is a characteristic enhancing core structure; (ii) **Tumor core (TC)**, a collection of three types of tumors: necrotic, non-enhancing and enhancing tumors; (iii) **Whole-area tumor (WT)**, which is a collection of all intra-tumor structures.

T_1 denotes the real brain tumor area (Ground Truth). T_0 denotes other parts, i.e., normal brain regions. P_1 denotes the predicted brain tumor area, P_0 denotes the other parts are the predicted normal brain regions. It is assumed that brain tumor is a positive sample and normal brain tissue is a negative sample. In the actual evaluation, (i) TP: the part of brain tumor correctly segmented by the model; (ii) TN: the part of normal tissue correctly segmented by the model; (iii) FP: the part of normal tissue but segmented as brain tumor; and (iv) FN: the part of brain tumor but segmented as normal tissue, which have also been shown in **Table 1**. Thus, we utilize the following evaluation metrics to evaluate our image classification model:

(i) Dice Similarity Coefficient (DSC) is a kind of set similarity measure, usually used to calculate the similarity of two samples, the higher the Dice coefficient means that the segmentation result of the algorithm is closer to the real result.

$$Dice(P,T) = \frac{|P_1 \cap T_1|}{\frac{(|P_1|+|T_2|)}{2}} = \frac{2TP}{FP+2TP+FN} \tag{12}$$

(ii) Sensitivity indicates the intersection of the results of the segmentation algorithm and the real results than the value of the real results, the larger the value of this evaluation index means that the segmentation results are closer to the real results.

$$Sensitivity = \frac{|P_1 \cap T_1|}{|T_1|} = \frac{TP}{TP+FN} \tag{13}$$

(iii) Positive Predictive Value (PPV) refers to the proportion of cases that are tumors among the number of positive samples of tumor segmentation results to be evaluated.

$$PPV = \frac{TP}{TP+FP} \tag{14}$$

(iv) Hausdorff distance, which represents the maximum distance between the edge points segmented by the algorithm and the actual edge points.

$$Hausdorff(P,T) = \max\{h(P,T), h(T,P)\} \tag{15}$$

3.2. Data Pre-processing

The dataset has completed the basic steps of brain MRI image processing (image alignment, cranial separation, etc.), but due to the specificity of medical images, there are some differences in MRI images obtained from the same patient machine trial in the same scanner at different time points. Considering the variability of MRI brain tumor image data distribution, to make the contrast and intensity range of patient MRI images similar, this section standardizes and normalizes the pixel intensity to balance the difference of gray value between images, and the process is as follows (also shown in **Figure 7**):

(1) Removal of 1% of the lowest and highest intensity pixel values.

(2) For each case, the mean value of the image is subtracted and divided by the standard deviation of the image, and data normalization is performed on the data within each input channel to obtain the standardized image.

$$F = \frac{X-\mu}{std} \tag{16}$$

In Equation (16), the X denotes the image matrix, and μ represents the pixel-average value of the image.

$$std = \max(\sigma, \frac{1}{\sqrt{N}}) \tag{17}$$

TABLE 1 | Pixel point classification labels.

Classification A[a] \ Classification B[b]	Brain tumor area	Normal tissue area
High-grade glioma	TP	FN
Tumor Tag	FP	TN

[a] *Classification of pixel dots in the standard.*
[b] *Classification of pixel points in prediction.*

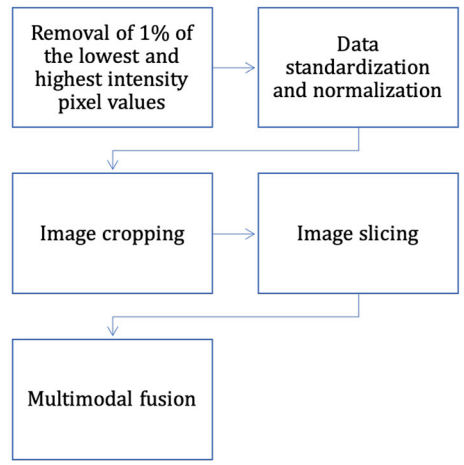

FIGURE 7 | The process of data pre-processing.

TABLE 2 | Training parameters setting.

Parameter	Value
Learning rate	0.0003
Batch_size	10
Early_stop	20
Epochs	1000
Optimizer	Adam

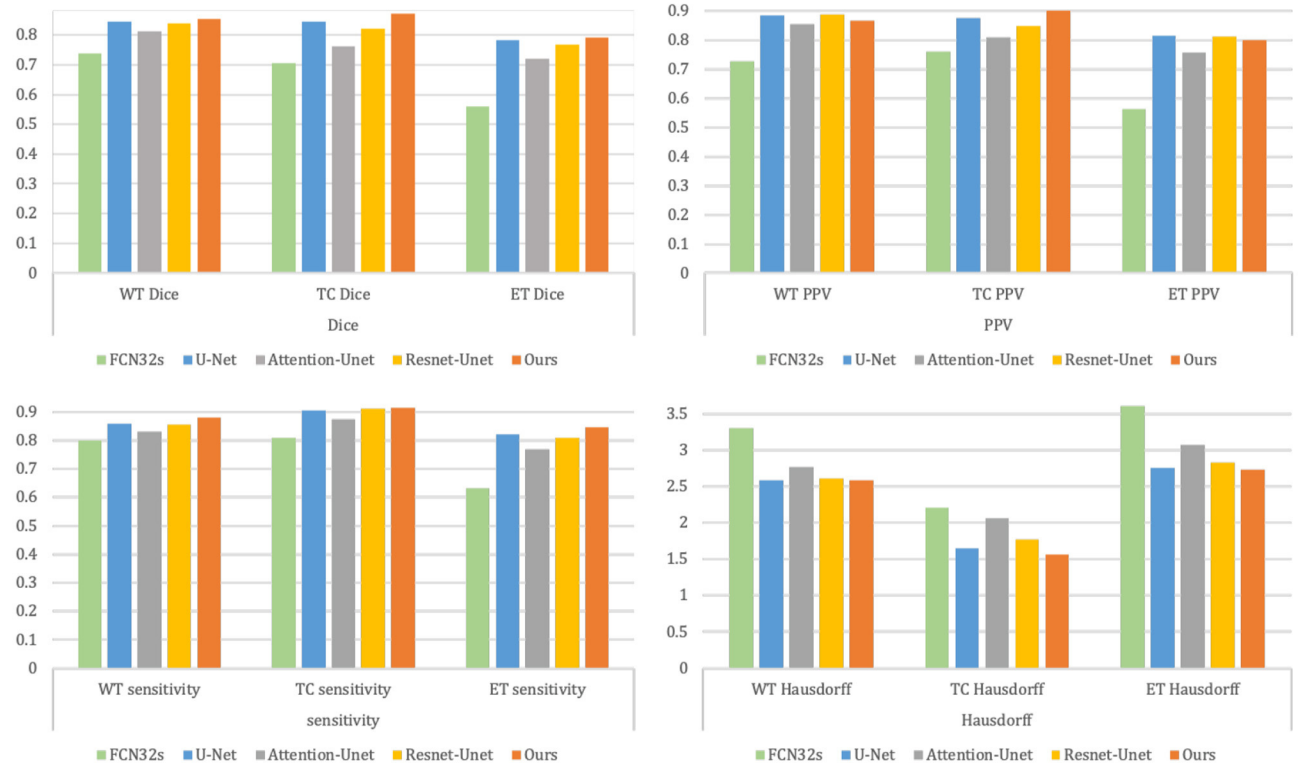

FIGURE 8 | Results comparison between the baseline methods and our proposed methodology.

In Equation (17), the σ denotes the standard variance of the image, and N denotes the total number of pixels of the matrix X represented by the image.

(3) Normalization operation is performed for image slices to compress their pixel intensities.

The intensity of the MRI images of the brain after the three processing steps described above is in the range of [0, 1]. Most medical images are 3D data, but the expert annotations in the BraTS dataset are for 2D axial slices rather than 3D images. In this article, two-dimensional brain tumor images were used for analysis as using 2D slices as network input can also provide enough features to identify each tumor region; therefore, the only way to fit the two-dimensional network was to slice the three-dimensional data into two-dimensional data. The problem of segmenting brain tumors in MRI images is associated with a large data imbalance. To reduce the imbalance of categories, we crop the image slices from the axial plane extraction into 2D image blocks of 128 × 128. The cropped sampled image blocks not only well avoid the zero-intensity pixel blocks, but also reduce the size of normal tissue regions, which are helpful to alleviate the data imbalance problem. In addition, this also reduces the computational effort and the training time.

3.3. Implementation Details

The Pytorch framework is used to implement the model and the experiments are done on Ubuntu 20.04 system with an Intel Core i7 3.5GHz processor and an NVIDIA TITAN 2080Ti graphics card. Eighty percent (i.e., 233 patients) was used to train the model There are four hyperparameters set for Adam in Pytorch: lr (learning rate), smoothing constant betas, eps and weight_decay. The parameters of our proposed model are shown in **Table 2**.

TABLE 3 | Results comparison.

Evaluation metrics	FCN32s	UNet	Attention-UNet	ResNet-UNet	Ours
WT Dice	0.7376	0.8450	0.813	0.8384	**0.8529**
TC Dice	0.7070	0.8454	0.7633	0.8228	**0.8705**
ET Dice	0.5610	0.7817	0.7218	0.7685	**0.7908**
WT PPV	0.7287	0.8859	0.8567	0.8887	0.8686
TC PPV	0.7599	0.8775	0.8105	0.8478	**0.9050**
ET PPV	0.5634	0.8154	0.7583	0.8127	0.7991
WT Sensitivity	0.7986	0.8595	0.8296	0.8553	**0.8818**
TC Sensitivity	0.8082	0.9069	0.8739	0.9102	**0.9145**
ET Sensitivity	0.6329	0.8203	0.7698	0.8099	**0.8453**
WT Hausdorff	3.3011	2.5787	2.7640	2.6126	2.5804
TC Hausdorff	2.2013	1.6516	2.0579	1.7696	**1.5660**
ET Hausdorff	3.6041	2.7487	3.0706	2.8314	**2.7331**

3.4. Baseline Methods

We compare our model with the following baseline methods:

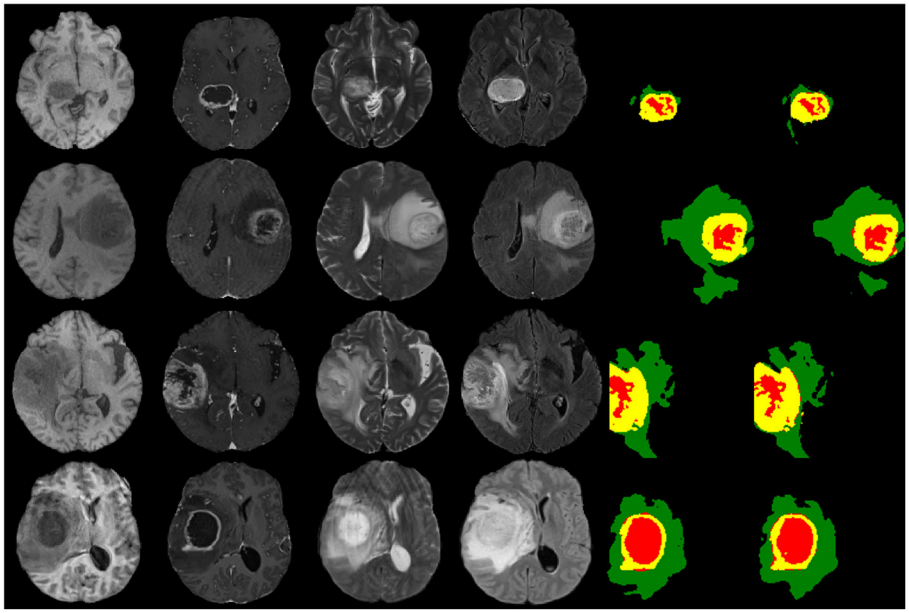

FIGURE 9 | Segmentation results for four cases. C_1 are the images represent one axial MRI slice obtained in T1-weighted, C_2 are the T1c-weighted, C_3 are the T2-weighted, C_4 are the FLAIR-weighted, C_5 are the the standard tumor segmentation results manually labeled by experts, and the C_6 show the segmentation results obtained in this section.

FCN32s (Shelhamer et al., 2016): It can accept input image of arbitrary size, and use a deconvolution layer to upsample the feature map of the last convolution layer. It recovers images to the same size as the input image, which can produce a prediction for each pixel while preserving the spatial information in the original input image.

UNet (Ronneberger et al., 2015): The main idea is to connect a network with a similar structure to the previous layer behind the downsampling stage, and then recover the resolution of the image output by applying upsampling. Finally combine the upsampled output with the downsampled output which has high-resolution features. By these steps the edge features of the image can be better extracted.

Attention-UNet (Oktay et al., 2018): The attention mechanism is introduced to learn the importance of each element with respect to the target. It limits the activation to the region with segmentation and reduces the activation value of the background to optimize the result.

ResNet-UNet: UNet with residual block helps to solve the gradient vanishing and gradient exploding problems and train deeper networks while ensuring good performance.

3.5. Performance Comparison

Table 3 and **Figure 8** summarize the baseline results compared to our proposed methodology. It is apparent from **Table 3** that our proposed framework outperforms the other four approaches across 9 metrics and achieves the highest Dice and Sensitivity, and the results are shown in bold values.

Figure 9 shows the segmentation results for the four cases obtained from the segmentation network of our proposed model. In these graphs, each row represents a real clinical case. From left to right, the images represent one axial MRI slice obtained in T1-weighted, T1c-weighted, T2-weighted, and FLAIR-weighted, the standard tumor segmentation results manually labeled by experts, and the last column shows the segmentation results obtained in this section. The tumor categories are highlighted with different colors: enhancing tumor regions (yellow), peritumor edema regions (green), and necrotic and non-enhancing tumors (red). It can be seen that the tumor size, shape, location, and category differ in the four brain MRIs. As can be seen from the comparison graph, the segmentation network of the Dense-ResUNet model is close to the standard segmentation results of cancer tumors manually labeled by experts, which proves the effectiveness of the Dense-ResUNet model.

4. CONCLUSION

In this article, we design a model based on 2D UNet brain tumor segmentation model and named it as Dense-ResUNet. By combining the features of multi-modal brain tumor MRI images, the attributes of each phase in MRI images were improved. The

Dense-ResUNet makes full use of the nested dense convolutional blocks to fill in the gaps of traditional UNet, which capture different levels of features. Then it uses a residual unit to extract pixel information from the image and skip the link to solve the traditional deep CNN network problem. Finally the multi-scale feature maps are fused to obtain the segmentation results. Experiment results show that our proposed method is better than other comparative approaches. The result can prove the effectiveness of our framework and show that the Dense-ResUNet can help extract the fusion of multi-modal of brain tumor images through the image convolutional network, which has clinical research and application value.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://aistudio.baidu.com/aistudio/datasetdetail/64660.

AUTHOR CONTRIBUTIONS

SC wrote the main manuscript and conducted the experiments. SZ and QL wrote, reviewed, edited, and supervised the manuscript. All authors read and approved the submitted manuscript.

REFERENCES

Abdelmaksoud, E., Elmogy, M., and Al-Awadi, R. (2015). Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Inf. J.* 58, 71–81. doi: 10.1016/J.EIJ.2015.01.003

Aslam, A., Khan, E., and Beg, M. (2015). Improved edge detection algorithm for brain tumor segmentation. *Procedia Compu. Sci.* 58, 430–437. doi: 10.1016/j.procs.2015.08.057

Brosch, T., Yoo, Y., Tang, L.Y.W., Li, D.K.B., Traboulsee, A., and Tam, R. (2015)."Deep convolutional encoder networks for multiple sclerosis lesion segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, Vol. 9351*, eds N. Navab, J. Hornegger, W. Wells, and A. Frangi (Cham: Springer). doi: 10.1007/978-3-319-24574-4_1

Hartigan, J. A., and Wong, M. A. (2013). A k-means clustering algorithm. *Appl. Stat.* 28, 100–108. doi: 10.2307/2346830

Hou, L., Samaras, D., Kurc, T., Gao, Y., Davis, J., and Saltz, J. (2016). "Patch-based convolutional neural network for whole slide tissue image classification," in *Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016 (Las Vegas, NV).

Kaus, M., Warfield, S., Nabavi, A., Black, P., Jolesz, F., and Kikinis, R. (2001). Automated segmentation of mr images of brain tumors. *Radiology* 218, 586–591. doi: 10.1148/radiology.218.2.r01fe44586

Kittler, J., and Illingworth, J. (1986). Minimum error threshold. *Pattern Recognit.* 19, 41–47.

Legg, P., Rosin, P., Marshall, D., and Morgan, J. (2015). Feature neighbourhood mutual information for multi-modal image registration: an application to eye fundus imaging. *Pattern Recognit.* 48, 1937–1946. doi: 10.1016/j.patcog.2014.12.014

Mahmood, Q., and Basit, A. (2016). "Automatic ischemic stroke lesion segmentation in multi-spectral mri images using random forests classifier," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, eds A. Crimi, B. Menze, O. Maier, M. Reyes, and H. Handels (Cham: Springer). doi: 10.1007/978-3-319-30858-6_23

Milletari, F., Navab, N., and Ahmadi, S. (2016). "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, 565–571. doi: 10.1109/3DV.2016.79

Moeskops, P., Viergever, M., Mendrik, A., Vries, L., Benders, M., and Isgum, I. (2016). Automatic segmentation of mr brain images with a convolutional neural network. *IEEE Trans. Med. Imag.* 35, 1252–1261. doi: 10.1109/TMI.2016.2548501

Nandpuru, H., Salankar, S., and Bora, V. (2014). "MRI brain cancer classification using support vector machine," in *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science* (IEEE), 1–6. doi: 10.1109/SCEECS.2014.6804439

Oktay, O., Schlemper, J., Folgoc, L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). Attention u-net: Learning where to look for the pancreas. arXiv [Preprint]. arXiv: 1804.03999. Available online at: https://arxiv.org/pdf/1804.03999.pdf (accessed May 20, 2018).

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, Vol. 9351*, eds N. Navab, J. Hornegger, W. Wells, and A. Frangi. doi: 10.1007/978-3-319-24574-4_28

Rtsch, G., Onoda, T., and Mller, K.-R. (2001). Soft margins for adaboost. *Mach. Learn.* 42, 287–320. doi: 10.1023/A:1007618119488

Saad, N. M., Bakar, S. A. R. A., Muda, S., and Mokji, M. (2010). "Automated segmentation of brain lesion based on diffusion-weighted mri using a split and merge approach," *Proceedings of 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2010* (Kuala Lumpur).

Shelhamer, E., Long, J., and Darrell, T. (2016). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1. doi: 10.1109/TPAMI.2016.2572683

Shi, T., and Horvath, S. (2005). Unsupervised learning with random forest predictors. *J. Comput. Graph. Stat.* 15, 118–138. doi: 10.1198/106186006X94072

Sobel, I. (1970). *Camera models and machine perception* (Ph.D. dissertation). Stanford University, United States.

Tustison, N., Shrinidhi, K., Wintermark, M., Durst, C., Kandel, B., Gee, J., et al. (2014). Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with antsr. *Neuroinformatics* 13, 209–225. doi: 10.1007/s12021-014-9245-2

Vaughan, J., Snyder, C., DelaBarre, L., Tian, J., Akgun, C., Ugurbil, K., et al. (2006). "Current and future trends in magnetic resonance imaging (MRI)," in *2006 IEEE MTT-S International Microwave Symposium Digest*, 2006, pp. 211–212, doi: 10.1109/MWSYM.2006.249451

Zhang, H., Berg, A., Maire, M., and Malik, J. (2006). "SVM-KNN: discriminative nearest neighbor classification for visual category recognition," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE), 2126–2136. doi: 10.1109/CVPR.2006.301

Zhou, Z., Siddiquee, Mmr, Tajbakhsh, N., and Liang, J. (2018). "UNet++: a nested U-net architecture for medical image segmentation," in *4th Deep Learning in Medical Image Analysis (DLMIA) Workshop* (Granada).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.