Computers in Biology and Medicine 157 (2023) 106751

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Two-Branch network for brain tumor segmentation using attention mechanism and super-resolution reconstruction

Zhaohong Jia, Hongxin Zhu, Junan Zhu *, Ping Ma

School of Internet, Anhui University, Hefei 230039, China

ARTICLE INFO

Keywords:
Brain tumor segmentation
MRI
3D U-Net
Shared encoder
Super-resolution image reconstruction
Attention mechanism

ABSTRACT

Accurate segmentation of brain tumor plays an important role in MRI diagnosis and treatment monitoring of brain tumor. However, the degree of lesions in each patient's brain tumor region is usually inconsistent, with large structural differences, and brain tumor MR images are characterized by low contrast and blur, current deep learning algorithms often cannot achieve accurate segmentation. To address this problem, we propose a novel end-to-end brain tumor segmentation algorithm by integrating the improved 3D U-Net network and super-resolution image reconstruction into one framework. In addition, the coordinate attention module is embedded before the upsampling operation of the backbone network, which enhances the capture ability of local texture feature information and global location feature information. To demonstrate the segmentation results of the proposed algorithm in different brain tumor MR images, we have trained and evaluated the proposed algorithm on BraTS datasets, and compared with other deep learning algorithms by dice similarity scores. On the BraTS2021 dataset, the proposed algorithm achieves the dice similarity score of 89.61%, 88.30%, 91.05%, and the Hausdorff distance (95%) of 1.414 mm, 7.810 mm, 4.583 mm for the enhancing tumors, tumor cores and whole tumors, respectively. The experimental results illuminate that our method outperforms the baseline 3D U-Net method and yields good performance on different datasets. It indicated that it is robust to segmentation of brain tumor MR images with structures vary considerably.

1. Introduction

Gliomas are the most frequent primary brain tumor, generally caused by glial cells and surrounding tissue lesions with varying degrees of invasiveness and destructiveness [1]. According to the degree of spread of glioma, it is divided into low-grade glioma (LGG) and high-grade glioma (HGG). HGG is aggressive, develop rapidly, and has a high mortality rate. LGG is generally benign and develops slowly, but it is possible to develop into HGG. Magnetic resonance imaging (MRI) is a typical non-invasive imaging technology, which is widely used in the research and clinical observation of brain tumors. In the diagnosis of brain tumors, it can clearly depict the structural images of the brain tumor without injury in the human brain, and provide comprehensive brain tumor information. There are four common modalities in MR images: (1) T1-Weighted (T1); (2) contrast enhanced T1-Weighted (T1-ce); (3) T2-Weighted (T2); (4) Fluid Attenuation Inversion Recovery (FLAIR). The four modalities can provide complementary information for different brain tumor regions.

To achieve accurate segmentation, we need to have a clear recognition of the brain tumor area and location before the surgery. Although great progress has been made in the study of related gliomas segmentation, brain tumor MR images of patients are generally visual observed, manually detected and tracked by radiologists with strong expertise in clinical diagnosis. It is a very time-consuming, tedious and subjective task. Therefore, accurate segmentation of brain tumor sub-region by automated or semi-automated segmentation to assist physicians is crucial for clinical diagnosis, treatment planning and follow-up disease tracking of brain tumor patients. In conclusion, brain tumor segmentation based on magnetic resonance imaging data is a key step in patient treatment and prognosis judgment. It is of great significance for the development of medicine to use computer related technology to assist the accurate segmentation of glioma.

Over a long period in the past, traditional machine learning methods have made great progress in the field of image segmentation, such as support vector machines, random forests, conditional random fields and probability theory, etc. However, in brain tumor image segmentation tasks, brain tumor images often have the characteristics of blurred boundaries, artifacts, low contrast, and different shapes. Machine learning methods are difficult to achieve these accurate segmentation tasks, especially to actively learn meaningful feature information in MR images.

* Corresponding author.
E-mail address: zhujunan@ahu.edu.cn (J. Zhu).

https://doi.org/10.1016/j.compbiomed.2023.106751
Received 16 October 2022; Received in revised form 12 February 2023; Accepted 6 March 2023
Available online 15 March 2023

With the development of deep learning technology and the improvement of GPU computing capability, deep learning methods are playing an increasingly important role in the field of computer vision, showing state-of-the-art performance in tasks, such as object detection and tracking, image segmentation and classification [2]. Owing to the deep learning methods have inherent and unique advantages in learning image feature information, more and more researchers have invested a lot of time and energy in deep learning methods and techniques. Pereira et al. [3] used a convolutional neural network with 3×3 convolution kernel for MR image segmentation. To fully exploit the tumor structure and accurately classify each voxel. Chen et al. [4] designed a novel densely connected convolutional block for hierarchical segmentation from different lesion regions to obtain multi-scale contextual information. To overcome the computational burden of processing 3D medical images, Kamnitsas et al. [5] proposed a two pathway and 11-layer 3D convolutional neural network that combines local and global contextual information to process the multi-scales images simultaneously. Zhou et al. [6] designed a lightweight multi-task model OM-Net that can decompose brain tumor segmentation into three different tasks and train together to exploit the underlying correlation for alleviate the class imbalance problem. Dolz et al. [7] designed a network HyperDenseNet that applies a dense connection module to achieve multi-modal segmentation, where each modality corresponds to a pathway, and it can effectively learn complex relationships between MR modalities. Liew et al. [8] proposed a CASPIANNET++ network, it introduced a channel and spatial wise asymmetric attention by leveraging the structure of brain tumors, and the multi-scale and multi-plane attention branches are designed to increase the spatial context information. To improve the representation ability of the model, Qin et al. [9] designed an autofocus convolutional layers for semantic segmentation by parallelize the convolutional layers with different dilation rates, and they combine the attention mechanism to generate more powerful features in an adaptive manner. Kamnitsas et al. [10] explored ensembles of multiple models and architectures (EMMA) for brain tumor segmentation, this approach reduces the influence of a single model in the segmentation task. To make up for the scarcity of medical image data scale, Zhang et al. [11] proposed a novel cross-modality deep learning framework, including cross-modality feature transition (CMFT) process and the cross-modality feature fusion (CMFF) process to mine rich patterns in multi-modality data. Wang et al. [12] introduced a fully convolutional neural network named WRN-PPNet by improving the pyramid pooling module for brain tumor regions segmentation. Havaei et al. [13] presented a convolutional neural network with a novel two pathway structure that learns about both the local detail features and global contextual features simultaneously, which significantly improves the segmentation performance. To efficiently extract multimodal MR image features, Peng et al. [14] proposed an automatic weighted dilated convolutional network (AD-Net), the proposed auto-weight dilated convolutional unit utilizes dual-scale convolutional feature maps to acquire channel separation features. In addition, the training technique of deep supervision is used to achieve fast fitting in his method. Hu et al. [15] proposed a novel brain tumor segmentation method based on multi-cascaded convolutional neural network (MCCNN) and fully connected conditional random fields (CRFs). Ma et al. [16] proposed a 3D lightweight CNN using dilated convolutions and residual connections to extract brain tumor substructures. Wang et al. [17] investigated how test-time augmentation can improve CNNs' performance for brain tumor segmentation. A range of methods are also used to enhance the image in his work. Zhou et al. [18] designed a novel 3D dense connectivity network to realize feature reuse, and added a new feature pyramid module to fuse multi-scale contexts for brain tumor segmentation.

Although the above algorithms have made some progress, they still do not achieve satisfactory segmentation accuracy. The accuracy of brain tumor segmentation algorithm based on deep learning needs to be further improved. There are three main key limitations: (1) deep convolutional neural network (DCNN) simply increasing or decreasing the number of layers cannot effectively improve the accuracy of brain tumor segmentation. (2) the size and shape of brain tumors in each patient are inconsistent, small-scale tumor areas are present, which requires algorithms to actively analyze the features of brain tumor MR images and achieve accurate brain tumor segmentation by enhancing the feature representation ability. (3) Due to the heterogeneity and highly class imbalance of brain tumors, common dense-prediction model cannot effectively solve these problems.

Based on encoder–decoder network, such as our baseline 3D U-Net, good segmentation accuracy is achieved in the medical image segmentation, but neglects the importance of positional feature information, which is critical to capture brain tumor structures. Besides, for class imbalance problem in brain tumor segmentation tasks, many recent works adopt the Model Cascade (MC) strategy to alleviate this matter, but the model usually is enormous. In the paper, we propose a novel end-to-end brain tumor segmentation method to better address the above problems. In general, the main contributions of this paper are in three aspects:

1. We propose a novel two-branch network, including segmentation branch and reconstruction branch, which has advantages in the segmentation of MR images with large shape differences.
2. To effectively alleviate the problem of class imbalance and influence of image artifacts, as well as better segmentation of brain tumor boundaries, a super-resolution image reconstruction method based on GAN is introduced into our model to assist the training of the network.
3. We expand coordinate attention mechanism from 2D to 3D and introduced it into our backbone network, which can effectively capture local feature information and global spatial location information in MR images. It can enhance the feature expression ability and improve the segmentation accuracy.

2. Related works

In this section, firstly, the encoder–decoder model based on full convolutional neural network and its application in medical image segmentation are described. Then, the related works of generative adversarial network and super-resolution reconstruction network are discussed. Finally, we describe the application of attention mechanism in deep learning methods. The description and limitations of related works are shown in Table 1.

2.1. Encoder-decoder network

In recent years, fully convolutional networks [12,21,33–35] are widely concerned because they can address many pixel-wise tasks. Ronneberger et al. [36] proposed a symmetric fully convolutional network based on the encoder–decoder structure, the so-called U-Net model for image segmentation. The network architecture consists of two paths, one is a contraction path used to capture contextual information, the other is an expansion path used to locate and segment, and use a skip connection method between the two paths, which greatly increases the accuracy of image segmentation. Therefore, more and more U-Net variant models are designed by researchers [19–21]. Owing to the simple encoder–decoder network structure has limited ability to learn image feature information, it needs to optimize the network architecture. Kong et al. [33] integrated the feature pyramid module into the U-Net architecture to achieve semantic feature capture of multi-scale feature information. Zhou et al. [37,38] proposed a multiple modalities network based on attention mechanism for brain tumor segmentation, using four independent encoding paths to extract features from the four MRI modalities respectively, and using a single decoding path for feature fusion, achieved a good segmentation result in brain tumor segmentation with missing modalities. Bukhari et al. [39] proposed

Table 1
This table summarizes different segmentation methods, the description, the limitations and the application of these methods for brain tumor segmentation tasks.

No	Segmentation methods	Description	Limitations	References
1	Based on Encoder Decoder architecture	The architecture consists of a contracting path and a symmetric expanding path. skip connections in the contracting path provide the essential high-resolution features to the expanding path.	Ordinary encoder–decoder network cannot take full advantage of context information, which is easy to cause the loss of effective information of image.	[19–21]
2	Based on GAN of SR Image Reconstruction	GAN is composed of generator and discriminator, which can generate any image, which can effectively relieve the problem of class imbalance in brain tumor datasets.	Directly train GAN needs to reach nash equilibrium, which will lead to instability in the training.	[22–26]
3	Based on Attention Mechanism	The attention-based networks can extract abundant multi-scale semantic information and enhance feature learning ability by using attention blocks.	Existing attention mechanisms only considers encoding channel and spatial information but neglects the importance of positional information, which is critical to capturing object structures in vision tasks.	[27–32]
4	Ours	By improving the encoder–decoder network, our method extends the coordinate attention block from 2D to 3D into our network, which can effectively extract multi-scale brain tumor image information, capture positional feature information and spatial feature information. Furthermore, we introduced the improved SRGAN into our network as a branch, which can effectively alleviate the problems of low contrast, artifact and class imbalance in brain tumor images.	To pursue the lightweight structure, a simple GAN module was used as our branch, which may lead to the segmentation details not good enough in a certain brain tumor area.	–

a novel network, called E1D3 U-Net, is a one-encoder, three-decoder fully convolutional neural network architecture where each decoder segments one of the hierarchical regions of interest. Fidon et al. [40] explored the inclusion of a transformer in the bottleneck of the U-Net architecture, Further, they adopted an efficient TTA strategy for faster and robust inference. To reduce memory consumption and decrease the effect of unbalanced data, Ballestar et al. [41] designed a 3D encoder–decoder architectures with patch-based techniques. Allah et al. [42] proposed a deep convolutional neural network, named the Edge U-Net, which can more precisely locate tumors by merging boundary-related MRI data with the main brain MRI data.

To alleviate the class imbalance problem, the two-stage network [35,43,44] adopted a cascade strategy for coarse-to-fine medical image segmentation, it has been favored by many researchers. Jiang et al. [35] designed a network that performs coarse prediction in the first stage to extract relevant features information to generates a segmentation feature map, and performs fine prediction in the second stage to refine the segmentation results, it won the first place in the BraTS2019 challenge. Lyu et al. [44] proposed a two-stage region segmentation model based on encoder–decoder architecture, incorporating a variational autoencoder into the network to regularize the model and prevent network overfitting.

Considering that the brain tumor data is a volumetric medical image, we improved the original encoder–decoder structure. Compared with two-dimensional networks, our 3D architecture can extract features more efficiently by voxel-by-voxel dense prediction.

2.2. Super-resolution image reconstruction

Since GAN was proposed by Goodfellow et al. [45], generative adversarial networks have become increasingly popular today because of their ability to synthesize any image, including the field of medical image analysis [22,23,46]. Cirillo et al. [22] proposed a 3D volume-to-volume generative adversarial network, called Vox2Vox model for brain tumor segmentation. Chen et al. [23] used generative adversarial networks instead of conditional random fields as a high-order smoothing method to improve the performance of the model for brain tumor segmentation. To achieve better segmentation performance, Zhu et al. [24] proposed a DualMMP-GAN network,it introduced dilated residual blocks and constructed a dual-scale discriminator to increase the receptive field, preserving context information of images.

Image super-resolution is an important processing technology used to improve image resolution in computer vision [47]. Generally, a high-resolution image is recovered from a degraded image or image sequence of low-resolution. Traditional super-resolution image reconstruction usually has the following three methods: (1) super-resolution image reconstruction based on interpolation; (2) super-resolution image reconstruction based on degradation model; (3) super-resolution image reconstruction based on traditional machine learning methods. With the development of deep learning, most researchers use deep learning techniques for super-resolution image reconstruction currently [48–50]. The technology also has a significant application in the field of medical images. Delannoy et al. [25] proposed a generative adversarial network that uses interpolation and single image super-resolution resampling methods for image segmentation.

The above works is to directly use GAN for brain tumor image segmentation or super resolution network for image reconstruction, which often leads to long training time and unstable training [51]. In our method, we improved the SRGAN [52] and applied the GAN-based super resolution image reconstruction as a branch to our network to assist training, which can effectively alleviate the problem of class imbalance and the influence of MR image artifact of brain tumor. In addition, using GAN as a branch to assist training can prevent the matter of gradient disappearing in the training process of network.

2.3. Attention mechanism

In recent years, attention mechanism plays an increasingly important role in computer vision tasks [27–29]. The attention mechanism mainly includes channel attention and spatial attention. Most of the existing works is to combine these two kinds of attention to enhance feature representation [30–32]. To better segment brain tumor region, Rehman et al. [53] designed a residual spatial pyramid pooling (RASPP) module and an attention gate (AG) module to acquire rich feature representations and retaining the local information. Xu et al. [54] proposed a deep supervised U-Attention network for pixel-wise brain tumor segmentation, which combines the U-Net, attention network and a deep supervised multistage layer. Jun et al. [55] used an improved attention block to refine the feature graph representation along a skipping join bridge consisting of parallel connected spatial and channel attention blocks.

However, these attention mechanism can only capture local relations but fail in modeling long-range dependencies that are essential for brain tumor segmentation tasks [30,32,55]. Different from some of the above methods, we also consider the spatial location information of

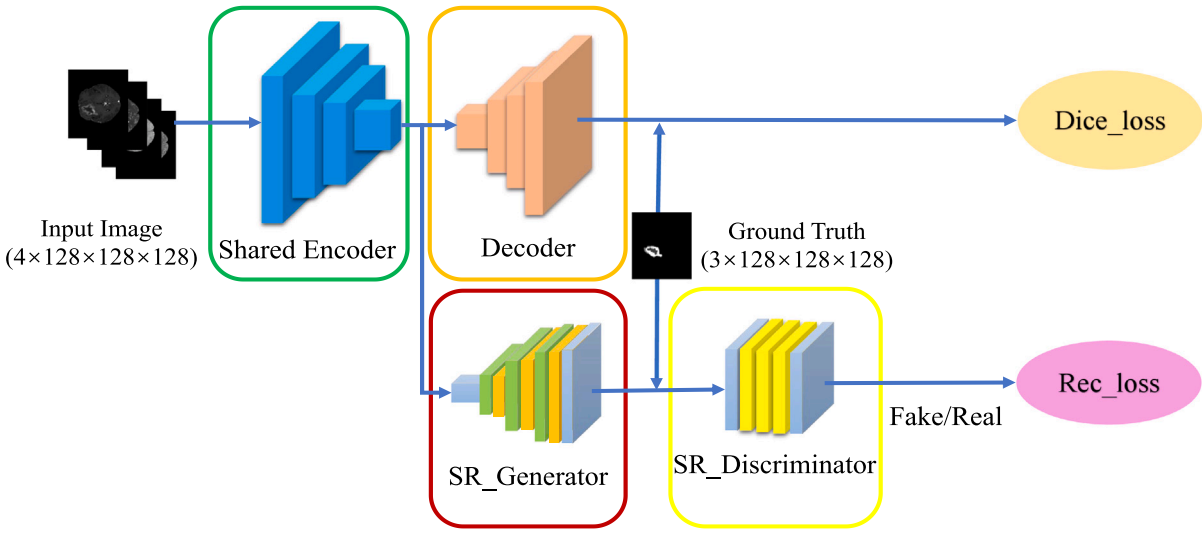

Fig. 1. The overview of our proposed network architecture. The input multimodal images are encoded through a shared encoder to extract relevant features, then one branch is the decoder for accurate localization and segmentation, and the other branch is the super-resolution image reconstruction module for image reconstruction.

the image. Inspired by coordinate attention [56], we extend the coordinate attention mechanism from 2D to 3D for voxel segmentation, and aggregate features from three directions (axial, sagittal, and coronal) of brain tumor image to enhance feature representation.

3. Methods

3.1. Overview

We propose a two-branch network based on shared encoder module, as shown in Fig. 1. It consists of the following four modules: shared encoder, decoder, super-resolution generator network, super-resolution discriminator network. Firstly, the multimodal MR image feature maps is extracted from the input images through the shared encoder module, and the high-resolution image features are encoded into low-resolution semantic features. Then there are two branches, one is the decoder branch for the backbone network, which is used to precisely locate the position information of the MRI volumetric images, and generate a predicted image for comparison with the ground-truth label. The other branch is the generator network for super-resolution image reconstruction. It generates a corresponding high-resolution image from a low-resolution image, then put the high-resolution image and ground-truth label are fed into the discriminator network for training. The discriminator network is used to discriminate between true and false to assist the training of backbone segmentation network. Finally, the segmentation loss and super-resolution reconstruction loss are summed as the total loss of our network to optimize our segmentation results.

3.2. Network architecture based improved 3D U-Net

In the brain tumor segmentation task, our training datasets is volumetric medical images. If we train on two-dimensional slices, we could lose a part of the spatial information in the MR images. Therefore, our backbone network is based on an improved version of the 3D U-Net [19] network structure. The network is illustrated Fig. 2, the model is mainly composed of two paths, we usually call them encoder and decoder. The encoder part consists of four typical convolution modules and four downsampling modules. Each convolution module has two $3 \times 3 \times 3$ convolution operations, where padding and dilation are set to 1. Each convolution operation is followed by an InstanceNorm3d and a nonlinear activation function leakyReLU layer. To preserve more details of voxels in brain tumor images, we set the dropout layer parameter to 0. The spatial downsampling operation is performed by max pooling layer with a kernel size of $2 \times 2 \times 2$ with stride 2. With each downsampling operation, the number of channels is doubled, and then reaches the network bottleneck layer. The size of the end point of the encoder is $512 \times 16 \times 16 \times 16$, which is 1/8 of the input MR image size. In the decoder, we add an attention fusion module before upsampling operation to capture and emphasize the image feature information. The traditional trilinear interpolation is used instead of the transposed convolution operation for upsampling, which reduces the computational complexity. The network uses the skip connection method between the module after upsampling and the encoder module. Symmetrical with the encoder, the decoder also has 4 typical convolution modules. Finally use a $1 \times 1 \times 1$ convolution operation to reduce the number of channels to the number of labels.

3.3. Image reconstruction branch

As another branch of the shared encoder, we propose an image super-resolution reconstruction method to restore low-resolution images to high-resolution images. Based on the Generative Adversarial Networks [45], we define a generator network G and a discriminator network D to solve the adversarial min–max problem. As shown in Eq. (1), X^{lr} represents the low-resolution image, and Y^{hr} represents the high-resolution image of the ground-truth. Our training is to expect the discriminator model to maximize the objective function $V(D, G)$ make $D(Y^{hr})$ approach 1 for real data and $D(G(X^{lr}))$ approach 0 for generated fake data. The generator model is expected to minimize the objective function $V(D, G)$, make the $D(G(X^{lr}))$ approach 1 to fool our discriminator network D. The generator network is optimized by making the discriminator network unable to recognize real or fake images.

$$\min_G \max_D V(D, G) = E_{Y^{hr}}[log\, D(Y^{hr})] + E_{X^{lr}}[log(1 - D(G(X^{lr})))] \tag{1}$$

After the image reconstruction branch receives the output of the encoder network, the generator of super-resolution image reconstruction generates a high-resolution image. Then the generated high-resolution image and the original ground-truth label image are trained by the discriminator network. As shown in Fig. 3, influenced by SRGAN [52], in the generator network, use a series of residual modules for channel transformation. Each residual module uses two $3 \times 3 \times 3$ convolution operations with padding 1, two instance normalization layers for regularization and one LeakyReLU layer for constraining the model. Residual connections are used to reduce the complexity of the model

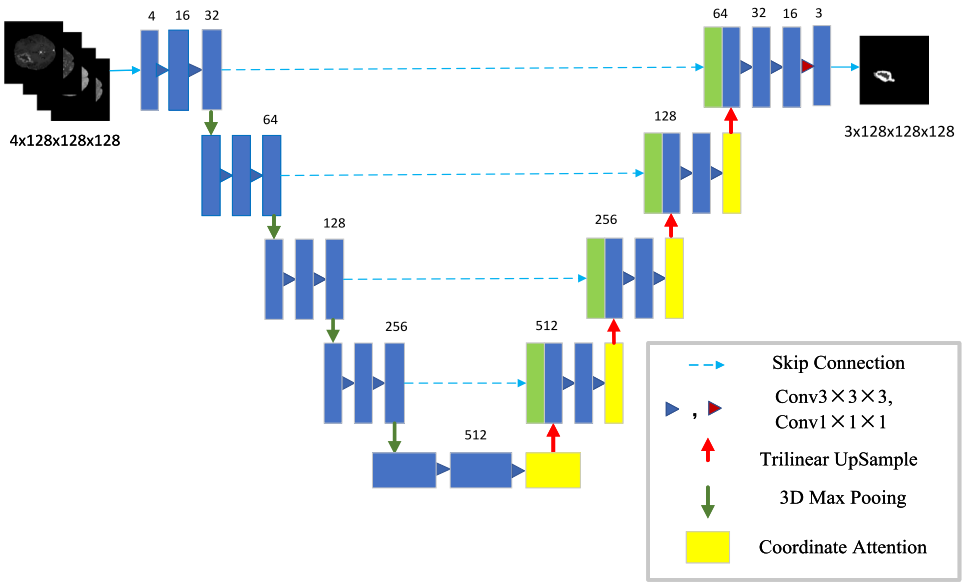

Fig. 2. The network structure of the improved 3D U-Net. The blue block represents feature maps. The green block represents copied feature maps. The yellow block represents the coordinate attention fusion module.

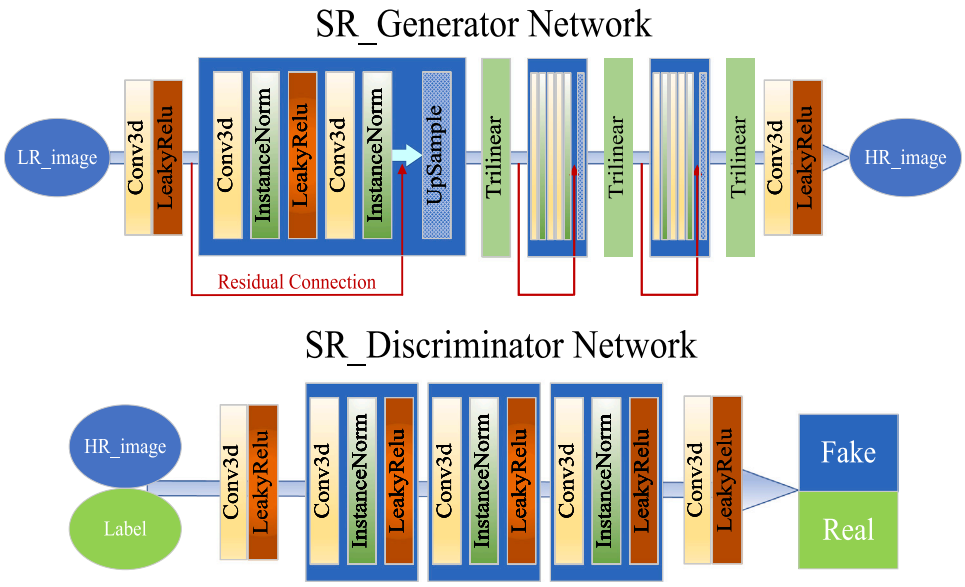

Fig. 3. Architecture of the generator network (top) and discriminator network (bottom) for super-resolution image reconstruction.

and prevent the vanishing gradient. At the end of the residual module, upsampling is performed using the $1 \times 1 \times 1$ convolution operation to reduce the channel by half. Owing to the input images are volumetric MR images, the network uses trilinear interpolation to improve the resolution of the input images. Using three times trilinear interpolation to expand the image resolution, and each time the image resolution is doubled, restore a $16 \times 16 \times 16$ size MR image to a $128 \times 128 \times 128$ size MR image. A convolution operation and a nonlinear activation function LeakyReLU layer are used at the beginning and end of the generator network to change the number of input and output channels.

Inspired by PixelGAN [57] in Conditional Adversarial Networks, we adjust the receptive field in the discriminator to the size of one pixel, use a convolution operation with a kernel of $1 \times 1 \times 1$ to extract features from MR images. The discriminator network uses three times of convolution operation, InstanceNorm3d and LeakyReLU layers. Similar to the generator network, a $1 \times 1 \times 1$ convolution operation and LeakyReLU layer is also used at the beginning and end of the discriminator network.

3.4. Coordinate attention mechanism

The attention mechanism generally includes channel attention mechanism and spatial attention mechanism. As the name suggests, channel attention is to focus on which features on the channel are meaningful, such as SE block [58] it includes squeeze module and excitation module. In voxel segmentation tasks, channel attention is usually represented as Eqs. (2) and (3). Given the input F, the squeeze step for the cth channel can be formulated as Eq. (2), the excitation step can be formulated as Eq. (3):

$$g_c = \frac{1}{D \times H \times W} \sum_{i}^{D} \sum_{j}^{H} \sum_{k}^{W} F_c(i, j, k) \tag{2}$$

$$G = F \cdot \sigma(L_2(Relu(L_1(g_c)))) \tag{3}$$

where g_c is the output associated with the cth channel, the D, H, W denotes depth, height and width directions, σ is the sigmoid function,

L_1 and L_2 are two linear transformation functions. $\cdot$ represents the channel-wise multiplication, G is the final output result.

However, channel attention only pays attention to the information on the channel and ignores the importance of spatial location information. In visual tasks, spatial context feature information is crucial for accurate segmentation, especially for brain tumor segmentation challenges. Spatial attention modules mainly use and emphasize spatial dimension information flow to capture local feature information. For example, CBAM [30] and scSE [32] both contain spatial attention modules. Global location feature information needs to be acquired to accurately segment volumetric images such as MR images. Therefore, inspired by the coordinate attention mechanism [56] embedding spatial location information of MR images into channel attention. As shown in Fig. 4, starting from three directions of axial D, sagittal H and coronal W of MR images, respectively, global average pooling is performed in parallel to form three independent directional feature maps, each feature map can capture long-range dependent feature information in one direction. Feature fusion in three directions and then use a $1 \times 1 \times 1$ convolution operation. The feature map is separated from each direction, and the convolution operation is performed again to obtain the local texture feature information of each direction. Finally, use the sigmoid function for nonlinear processing and then perform the residual connection to reduce the complexity of the model. The coordinate attention mechanism can be represented by Eq. (4).

$$\begin{aligned} f^d &= \varphi(F_1^d([g_d, g_h, g_w])) \\ f^h &= \varphi(F_1^h([g_d, g_h, g_w])) \\ f^w &= \varphi(F_1^w([g_d, g_h, g_w])) \\ Z^d &= \sigma(F_2^d(f^d)) \\ Z^h &= \sigma(F_2^h(f^h)) \\ Z^w &= \sigma(F_2^w(f^w)) \\ Y &= X \cdot Z^d \cdot Z^h \cdot Z^w \end{aligned} \tag{4}$$

where, g_d, g_h, g_w denotes global average pooling operation in the three directions. $[\cdot, \cdot, \cdot]$ represents the concatenation operation along the spatial dimension. F_1, F_2 denotes the first and second convolution operations. $\varphi(*)$ is a regularization operation and non-linear activation function. σ is the sigmoid function. X is the input image. Finally, the Y is the output result of our coordinate attention block.

In this way, not only the global position information can be captured, but also the local texture feature information can be captured. This local and global combination method assist our network to more accurately locate and identify the features of interest. After adding the attention module, our proposed network not only the accuracy can be improved, but also the convergence speed of network training can be accelerated.

4. Experiments

4.1. Data description

The datasets used in our experiments are BraTS2018, BraTS2020 and BraTS2021 training datasets and online validation datasets for the Brain Tumor Segmentation Challenge Task. The details of datasets are shown in Table 2. The BraTS2018 dataset contain a training dataset of 285 cases and an online validation dataset of 66 cases with hidden ground-truth, and the training dataset consists of 210 HGG images and 75 LGG images. The BraTS2020 dataset contain a training dataset of 369 cases and an online validation dataset of 125 cases with hidden ground-truth, the training dataset consists of 259 HGG images and 110 LGG images. The BraTS2021 dataset contain a training dataset with 1251 cases and an online validation dataset with 219 cases with hidden ground-truth. Fig. 5 show a typical case of MR brain tumor image along with the ground-truth. In these training datasets, each case has four image modalities including T1, T1-ce, T2 and FLAIR. The

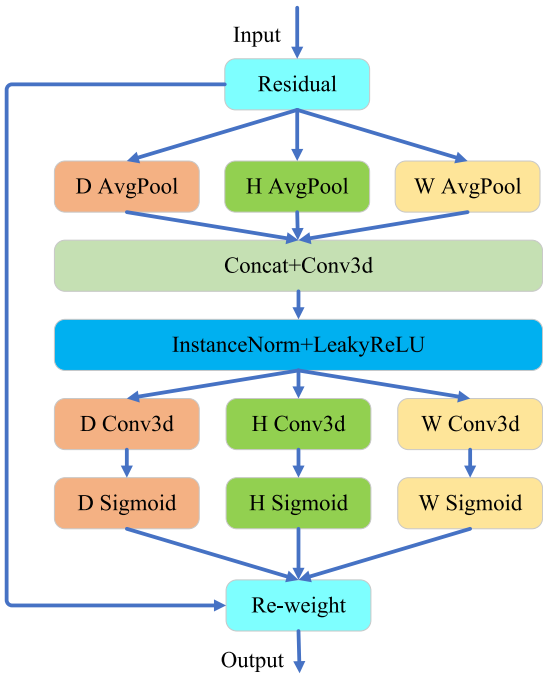

Fig. 4. Overview of the improved coordinate attention mechanism. D represents the axial direction, H represents the sagittal direction, and W represents the coronal direction.

Table 2
Experiment data partitioning.

Dataset	BraTS2018	BraTS2020	BraTS2021
Training set	285	369	1251
Validation set	66	125	219
Total	351	494	1470

labels are divided into four classes, namely the healthy tissues (label 0), the necrotic tumor core (label 1), the peritumoral edematous/invaded tissue (label 2), and the Gd-enhancing tumor (label 4). Our task is to segment three mutually inclusive tumor sub-regions: (1) The enhancing tumor region (ET); (2) The tumor core region (TC); (3) The whole tumor region (WT).

4.2. Implementation details

The proposed method is implemented by the Pytorch framework in a PC with an NVIDIA 3090 GPU and an AMD Ryzen 9 5900X CPU 3.70 GHz with 32 GB RAM. To save computing resources and speed up network training, we use automatic mixed precision (AMP) for training. The experiment adopts five-fold cross-validation for training. The training dataset samples are divided into five parts, one of which is used as the validation set to adjust the parameters and monitor the performance of the model, and the other four parts are used as the training set to train the parameters of the network. After split the BraTS2021 training set into five folds, the number of samples in the training set is 1000 cases (80%), and the number of samples in the validation set is 251 cases (20%). To demonstrate the robustness of the proposed network, the BraTS2018 and BraTS2020 datasets are also used for training and evaluation. After adopting the five-fold cross-validation method, the BraTS2020 training set samples are 295 cases (80%), and validation set samples are 74 cases (20%); the BraTS2018 training set samples are 228 cases (80%), and validation set samples are 57 cases (20%). Our segmentation model was built with five-fold cross validation upon the training data, then take the best model as our final model. Since the testing data of the dataset did not provide

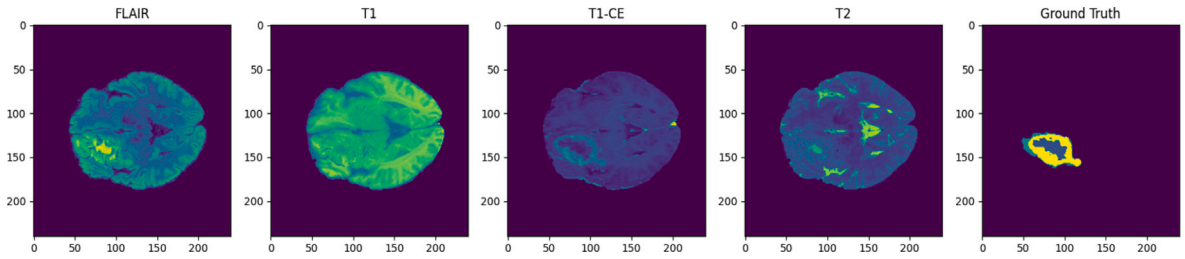

Fig. 5. Example of the brain MR images from the BraTS2021 training dataset. From left to right: FLAIR, T1, T1-ce, T2 and ground-truth.

ground truth segmentation result, the final segmentation results were evaluated by the BraTS online evaluation platform.

4.2.1. Image preprocessing

Owing to the MR image datasets used in the experiments were collected by many different institutions using different equipment and different imaging protocols, the image quality in each case was widely different, with characteristics such as low contrast and different intensities. In addition, the structure of brain organs is relatively complex, and the tumor lesion region is also inconsistent. Brain tumor segmentation has always been a difficult challenge in the field of medical images. Therefore, additional normalization is required for the MR images in the training dataset. Before feeding the input images into the network, the min–max normalization method is used to scale the input MR images. Since brain tumor images are volumetric data, spatial context information is important, some of the pixels labeled 0 were removed and the image was adjusted from $155 \times 240 \times 240$ to a volume size of $128 \times 128 \times 128$ to adapt our segmentation task.

4.2.2. Optimization

We use the AdamW optimizer with initial learning rate of $lr_{init} = 1e-4$ for weights updating in the backbone network and use the Adam optimizer with initial learning rate of $lr_{init} = 1e-4$ in the super-resolution reconstruction network. We progressively decay the learning rate according to the Eq. (5):

$$lr = lr_{init}(1 - \frac{e}{N_e})^{\mu} \tag{5}$$

where e is an epoch counter. N_e is the total number of the epochs during training, we set N_e is to 100 in our experiments. μ is a parameter and we set μ is to 0.9. Our network adopts the He initialization method to initialize our weights and biases.

4.2.3. Loss functions

The loss function of brain tumor segmentation network includes two terms, expecting to minimize L_{total} to optimize the network parameters. The network is trained by the overall loss function as Eq. (6):

$$L_{total} = L_{dice} + \alpha L_{rec} \tag{6}$$

where α is the trade-off parameters that weighting the importance of the super-resolution image reconstruction components, which is set to 1 in our experiments.

For segmentation network, dice loss is used to evaluate the overlap rate of prediction results and ground-truth for three classes (enhancing tumor, tumor core and whole tumor).

$$L_{dice} = 1 - \frac{2\Sigma(X * Y) + \epsilon}{\Sigma X^2 + \Sigma Y^2 + \epsilon} \tag{7}$$

where X represents the predicted probability value, Y represents the ground-truth value. ϵ is a smoothing factor to avoid dividing by 0. We set the ϵ to 1 in our experiments.

For the super-resolution image reconstruction network, inspired by Generative Adversarial Networks [45], the purpose of training is to allow the discriminator network to distinguish which is real data and which is fake data by the generator network. The Binary Cross Entropy (BCE) loss function is used to match each reconstructed image and ground-truth image.

$$L_{rec} = -\Sigma[(1 - y)ln(1 - \sigma(x)) + ylN\sigma(x)] \tag{8}$$

where x denotes prediction results for each class, σ is a sigmoid activation function, and y is the label (value is 0 or 1)

4.2.4. Post processing

Owing to the different sizes and shapes of brain tumor data, it can be observed that when the predicted volume of ET is particularly small, our proposed method tends to incorrectly predict TC voxels as ET. Therefore, if the prediction probability of the ET region is less than 0.5, we use the TC instead of the ET region. This may result in a dice score of 0 in the ET region, but overall, we can obtain a better segmentation result during training phase.

4.2.5. Evaluation metrics

To verify the effectiveness of our proposed method, dice are used to evaluate the performance of the proposed algorithm, which stands for the overlap rate between ground truth and model prediction results. The larger the dice score value, the higher the overlap rate, which means the better the segmentation result. Dice ranges from 0~1. Dice can be calculated by Eq. (9).

$$Dice\ Score = \frac{2TP}{2TP + FP + FN} \tag{9}$$

Hausdorff Distance 95% (HD95) is computed between boundaries of the ground truth and the prediction results. The smaller the HD95 value, the better the prediction result. It is formulated as Eq. (10).

$$\text{Hausdorff Distance} = max\{d(A, B), d(B, A)\} \tag{10}$$

where A denotes the pixel set of prediction result, and B denotes the pixel set of ground-truth. For $A = \{a_1, a_2, \cdots, a_n\}$ and $B = \{b_1, b_2, \cdots, b_n\}$ in Euclidean space, the $d(A, B) = max_{a \in A} min_{b \in B} \|a - b\|$.

Sensitivity measures the predictive power for the positive voxel predictions. Specificity measures the predictive power for negative voxels predictions. Precision measures the predict correct power for the predicted voxels. Accordingly, they are defined as Eq. (11)~Eq. (13).

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

where TP represents the number of true positive voxels, FN represent the number of false negative voxels, FP represents the number of false positive voxels, and TN represents the number of true negative voxels.

Table 3
Ablation results on the BraTS2021 training dataset. The evaluation metrics is the dice score of each sub-region of the prediction result, and the best results in **bold**.

Method	Dice score		
	ET	TC	WT
3D U-Net	0.8452	0.8871	0.9154
3D U-Net+ATT	0.8495	0.8947	0.9169
3D U-Net+REC	0.8420	0.8901	0.9159
3D U-Net+REC+ATT	**0.8513**	**0.9009**	**0.9211**

Table 4
The segmentation performance metrics on BraTS2021 online validation dataset.

Class	Dice	HD95	Sensitivity	Specificity	Precision
ET	0.8961	1.4142	0.8456	0.9997	0.9531
TC	0.8830	7.8103	0.9618	0.9977	0.8161
WT	0.9105	4.5826	0.9638	0.9976	0.8629

5. Results

5.1. Quantitative results

5.1.1. Ablation results

In this section, to demonstrate the effectiveness of the proposed network, we conduct ablation experiments on the BraTS2021 training dataset to evaluate our method and verify the role of our proposed components. It includes super-resolution reconstruction component and coordinate attention fusion component. As shown in Table 3, The 3D U-Net network is used as our baseline model, and the dice scores of baseline network on the BraTS2021 training set are 0.8452, 0.8871 and 0.9154 for enhanced tumor (ET), tumor core (TC) and whole tumor (WT), respectively. By comparing the baseline model of 3D U-Net and the model of 3D U-Net + ATT method, it can be found that adding attention fusion mechanism can improve the segmentation accuracy of brain tumor sub-regions, which indicates that the feature learning ability of the model is enhanced with the help of the attention mechanism. By comparing the baseline model of 3D U-Net and the model of 3D U-Net + REC method, it can be seen that the dice scores of the whole tumor (WT) and the tumor core (TC) region increased after adding the super-resolution image reconstruction, but for the enhanced tumor (ET), the dice scores decreased slightly, which may be mainly caused by the simple generator network structure in the super-resolution image reconstruction component, which fails to produce clear high-resolution images. When these two components are included, the dice score was significantly improved and the best evaluation results are achieved, with enhanced tumor (ET), tumor core (TC) and whole tumor (WT) dice scores of 0.8513, 0.9009 and 0.9211 on the BraTS2021 training set, respectively. This shows that the attention mechanism enhanced the ability of the network to learn features, and the super-resolution reconstruction component helps the backbone network to obtain more powerful feature streams. As shown in Fig. 6, the boxplot shows the minimum, lower quartile, median, upper quartile and maximum for each tumor class.

5.1.2. Challenge results

We submit the predicted results of the BraTS2021 online validation dataset to the website, obtain evaluation results are shown in Table 4. It shows the dice score, Hausdorff Distance (95%) and sensitivity, specificity and precision of each sub-region in our prediction results.

5.1.3. Comparative results

In this section, we compare different trade-off parameter to determine the optimal coefficients of our algorithm. In addition, we compare the proposed method with the latest deep learning methods on the BraTS2018, BraTS2020 and BraTS2021 online validation datasets.

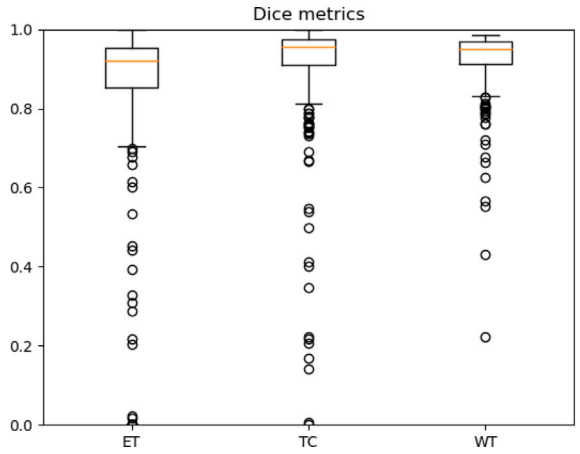

Fig. 6. The box-plot of BraTS2021 training result. The horizontal axis is each tumor segmentation area (ET, TC and WT), and the vertical axis is the dice similarity score.

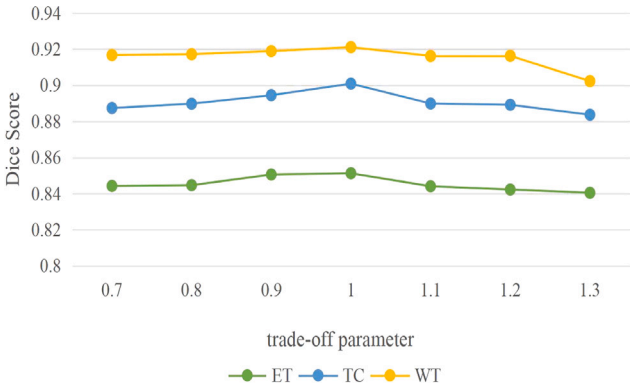

Fig. 7. For the super-resolution reconstruction network, we compare the trade-off parameter α. The parameters are in the range of 0.7∼1.3.

Inspired by [38], our initialization trade-off parameter α is set to 1. To prove that the trade-off parameters setting in the loss function is optimal, we use adjacent different parameters for comparison as shown in Fig. 7. When the trade-off parameter α is set to 1.0, the segmentation results of each brain tumor sub-region are the best.

We also train and evaluate on BraTS2018 and BraTS2020 datasets and compare them with other algorithms. As shown in Table 5, our proposed method achieves good segmentation results in each sub-region on the three years datasets. On the BraTS2021 dataset, we achieved the best evaluation in the enhanced tumor (ET) and tumor core (TC), outperforming Peiris et al. [46] by 8.22%, 2.91%, and 0.28% on the Dice scores for ET, TC and WT, respectively. Fidon et al. [40] achieved the best evaluation in the whole tumor (WT), outperforming ours by 1.65%. The main reason for this result is that the algorithm adds the transformer module to the bottleneck of the U-Net architecture. The transformer module can accumulate the global context information of the image and learn an anatomically consistent representation of the tumor classes, so the segmentation result of the whole tumor is better than our algorithm. However, we achieved the best average accuracy of 89.65%. On the BraTS2020 dataset, we achieved the best results in the enhanced tumor (ET) and whole tumor (WT), it reached 81.92% and 88.62%, respectively. Ballestar et al. [41] achieved the best evaluation in the tumor core (TC), which reach 79%. He used a patch-based 3D encoder–decoder structure to train the model, the attributes of each model are used for ensemble training to improve performance. However, the algorithm did not perform well in other brain tumor subregions, the accuracy is almost 5%∼10% lower than

Table 5
In BraTS2021, BraTS2020 and BraTS2018 online validation datasets, we use dice scores in each sub-region to compare with other algorithms. The best dice similarity score results for each tumor regions in **bold**.

Datasets	Methods	Metrics			Avg
		ET	TC	WT	
BraTS2021	E1D3 [39]	0.822	0.865	0.924	0.8703
	Fidon et al. [40]	0.84	0.87	**0.927**	0.879
	Peiris et al. [46]	0.8139	0.8539	0.9077	0.8585
	Our Method	**0.8961**	**0.883**	0.9105	**0.8965**
BraTS2020	Ballestar et al. [41]	0.72	**0.79**	0.84	0.7833
	Dual-Path [55]	0.752	0.779	0.879	0.8033
	U-Attention [54]	0.67	0.7	0.86	0.7433
	Our Method	**0.8192**	0.7434	**0.8862**	**0.8163**
BraTS2018	Two-stage [43]	0.7229	0.7675	0.8623	0.7842
	multi-cascaded [15]	0.7178	0.7481	0.8824	0.7828
	Ma et al. [16]	**0.743**	0.773	0.872	0.796
	Wang et al. [17]	0.7344	0.7658	0.8638	0.788
	Our Method	0.7085	**0.8161**	**0.8896**	**0.8047**

ours. In addition, the average dice accuracy is 3.3% better than his method. On the BraTS2018 dataset, we achieved the best evaluation in the tumor core (TC) and whole tumor (WT), it reached 81.61% and 88.96%, respectively. Ma et al. [16] is 3.45% better than ours for dice accuracy of enhance tumor (ET). The main reason is because he proposed a 3D lightweight convolutional neural network to extract brain tumor sub-region, and used atrous convolution and residual connections in the model to reduce the amounts of parameters, enhanced feature extraction ability of small tumor area. Therefore, the algorithm can produce better segmentation in the ET region. But for the average dice accuracy, we still reach the best accuracy of 80.47%.

To relieve the class imbalance problem in MR brain tumor datasets, many different strategies and methods are adopted for brain tumor image segmentation, its results are very encouraging. However, these studies have achieved good accuracy in segmentation class such as ET, TC or WT, but the average accuracy is not satisfactory. Inspired by SRGAN's successful experience, deep residual network is able to recover image details from downsampling images. As one of the branches of our network, the generator is used to generate a high-resolution image from the downsampling image, then the discriminator compares the high-resolution image to the ground truth and calculates the reconstruction loss to train our model. It can effectively alleviate the class imbalance problem. GAN is a minimax two-player game problem, training GAN needs to reach nash equilibrium, when the balance is not reached, it will lead to unstable training or unsatisfactory segmentation accuracy of a certain subregion, but the average dice score is still optimal in our network. Overall, our algorithm is better than most of the existing algorithms.

5.2. Qualitative result

To intuitively show the effectiveness of our proposed algorithm in different brain tumor datasets, as shown in Fig. 8, we randomly select an example to display the visual segmentation results on the BraTS2018 (first row), BraTS2020 (second row) and BraTS2021 (third row) datasets, respectively. The *A* option (the first column) shows the images of the FLAIR modality in the datasets, the *B* option (the second column) shows the visual segmentation images of the ground-truth, and the *C* option (the third row) shows the visual segmentation results of our proposed method. As can be seen from the figure, the prediction results of our algorithm in BraTS2021 and BraTS2020 datasets are good, this proves that our segmentation algorithm is effective. However, in the BraTS2018 dataset, the prediction results of brain tumor subregions are not satisfactory, which may be caused by the simple structure of super-resolution image reconstruction components, which limits the generalization performance of the model. In summary, our proposed segmentation algorithm is feasible.

6. Discussion

Brain tumor is a great threat to human health, it is of great clinical significance to use computer aided artificial segmentation. However, in the brain tumor datasets, the shape and size of brain tumor image data are generally uneven, the structure is different, the boundary is blurred, and there are often artifacts, low contrast. Although most of the existing algorithms can achieve good segmentation results, for brain tumor images with large differences and uneven distribution, the segmentation results will be unstable and easily affected by the data distribution of MR images. Therefore, robust and accurate segmentation of brain tumors is an extremely challenging task. In this study, we propose a two-branch network based on shared encoder to address this challenge, one is the super-resolution image reconstruction branch, and the other is a coordinate attention mechanism branch. Compared with the existing networks, our model can effectively capture the feature stream and has better results of segmentation on different brain tumor MR image datasets. For data distribution of different images, our generator network can fit the distribution of these MR images, and it can learn the unique feature information of brain tumor images. The coordinate attention mechanism can obtain feature streams from different directions, considering both spatial global location feature information and local texture feature information, which is of great help to accurately segment individual subregions of brain tumors. To verify the effectiveness of our proposed algorithm, as shown in Table 3, we can observe that the proposed method improves the dice similarity scores compared to the 3D U-Net baseline model. For adding individual component, we integrate the components together to achieve the best result of segmentation. As shown in Table 5, to prove the robustness of the proposed model for image segmentation with large differences, we evaluated and compared the BraTS datasets with other deep learning algorithms. Our visual segmentation rendering is shown in Fig. 8. Experimental results show that the proposed method achieves better results of segmentation on different datasets. However, the proposed method has a little drawback. Firstly, due to the limitation of computing resources, we use the traditional trilinear interpolation method for super-resolution image reconstruction, which leads to the generator network cannot to learn more parameters. Furthermore, the segmentation details of brain tumor sub-regions are not good enough, mainly due to the simple GAN branch, which may lead to unstable behavior of GAN training. Last but not least, the more network blocks or layers in the model, the higher the computational cost. To address these drawbacks, in the future, we can perfect the super-resolution image reconstruction component and add more learnable parameters to adapt to accurate segmentation of brain tumors. In addition, we can perform appropriate pruning operations in the backbone network or consider other lightweight models to segment brain tumors.

7. Conclusion

In this paper, we propose a novel two-branch brain tumor segmentation network based on shared encoder. To enhance the capture ability of local texture features and global location features, we introduce a lightweight attention mechanism, named coordinate attention. Furthermore, to better segment the MR image boundary of brain tumor, we combine the super-resolution image reconstruction with the improved 3D U-Net backbone network, and discriminate the ground-truth and the generated high-resolution images through the adversarial thought to achieve accurate segmentation. Finally, we conduct qualitative and quantitative comparative experiments on our methods and compare it with other similar deep learning algorithms on different datasets. Experiments show that our end-to-end segmentation method has good robustness in different brain tumor datasets. In short, it indicates that our proposed method has good segmentation potential for magnetic resonance imaging data with large data distribution differences.

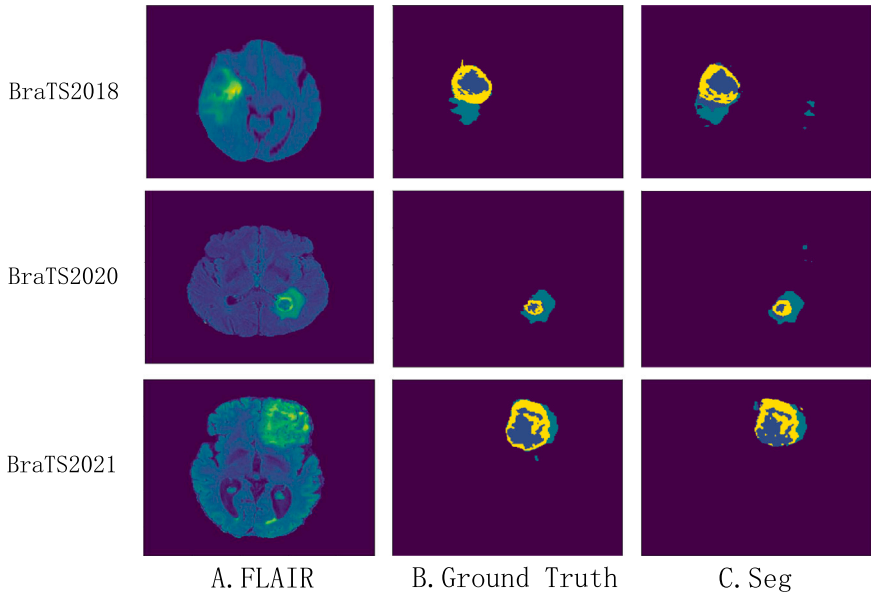

Fig. 8. Visual segmentation renderings on three datasets: BraTS2018, BraTS2020 and BraTS2021. The FLAIR modality images, the corresponding ground-truth labels and the visual segmentation images predicted by our model are shown respectively.

CRediT authorship contribution statement

Zhaohong Jia: Conceptualization, Methodology, Writing – review & editing. **Hongxin Zhu:** Software, Writing – original draft. **Junan Zhu:** Investigation, Methodology, Writing – review & editing. **Ping Ma:** Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China No. 71971002.

References

[1] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2014) 1993–2024, http://dx.doi.org/10.1109/TMI.2014.2377694.

[2] T. Zhou, S. Ruan, S. Canu, A review: Deep learning for medical image segmentation using multi-modality fusion, Array 3 (2019) 100004, http://dx.doi.org/10.1016/j.array.2019.100004.

[3] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE Trans. Med. Imaging 35 (5) (2016) 1240–1251, http://dx.doi.org/10.1109/TMI.2016.2538465.

[4] L. Chen, Y. Wu, A.M. DSouza, A.Z. Abidin, A. Wismüller, C. Xu, MRI tumor segmentation with densely connected 3D CNN, in: Medical Imaging 2018: Image Processing, Vol. 10574, SPIE, 2018, pp. 357–364, http://dx.doi.org/10.1117/12.2293394.

[5] K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, et al., Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78, http://dx.doi.org/10.1016/j.media.2016.10.004.

[6] C. Zhou, C. Ding, Z. Lu, X. Wang, D. Tao, One-pass multi-task convolutional neural networks for efficient brain tumor segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 637–645, http://dx.doi.org/10.1007/978-3-030-00931-1_73.

[7] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, I.B. Ayed, HyperDenseNet: a hyper-densely connected CNN for multi-modal image segmentation, IEEE Trans. Med. Imaging 38 (5) (2018) 1116–1126, http://dx.doi.org/10.1109/TMI.2018.2878669.

[8] A. Liew, C.C. Lee, B.L. Lan, M. Tan, CASPIANET++: a multidimensional channel-spatial asymmetric attention network with noisy student curriculum learning paradigm for brain tumor segmentation, Comput. Biol. Med. 136 (2021) 104690, http://dx.doi.org/10.1016/j.compbiomed.2021.104690.

[9] Y. Qin, K. Kamnitsas, S. Ancha, J. Nanavati, G. Cottrell, A. Criminisi, et al., Autofocus layer for semantic segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 603–611, http://dx.doi.org/10.1007/978-3-030-00931-1_69.

[10] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, et al., Ensembles of multiple models and architectures for robust brain tumour segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2017, pp. 450–462, http://dx.doi.org/10.1007/978-3-319-75238-9_38.

[11] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, Y. Yu, Cross-modality deep feature learning for brain tumor segmentation, Pattern Recognit. 110 (2021) 107562, http://dx.doi.org/10.1016/j.patcog.2020.107562.

[12] Y. Wang, C. Li, T. Zhu, J. Zhang, Multimodal brain tumor image segmentation using WRN-PPNet, Comput. Med. Imaging Graph. 75 (2019) 56–65, http://dx.doi.org/10.1016/j.compmedimag.2019.04.001.

[13] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, et al., Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (2017) 18–31, http://dx.doi.org/10.1016/j.media.2016.05.004.

[14] Y. Peng, J. Sun, The multimodal MRI brain tumor segmentation based on AD-Net, Biomed. Signal Process. Control 80 (2023) 104336, http://dx.doi.org/10.1016/j.bspc.2022.104336.

[15] K. Hu, Q. Gan, Y. Zhang, S. Deng, F. Xiao, W. Huang, et al., Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field, IEEE Access 7 (2019) 92615–92629, http://dx.doi.org/10.1109/ACCESS.2019.2927433.

[16] J. Ma, X. Yang, Automatic brain tumor segmentation by exploring the multi-modality complementary information and cascaded 3D lightweight CNNs, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 25–36, http://dx.doi.org/10.1007/978-3-030-11726-9_3.

[17] G. Wang, W. Li, S. Ourselin, T. Vercauteren, Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 61–72, http://dx.doi.org/10.1007/978-3-030-11726-9_6.

[18] Z. Zhou, Z. He, M. Shi, J. Du, D. Chen, 3D dense connectivity network with atrous convolutional feature pyramid for brain tumor segmentation in magnetic resonance imaging of human heads, Comput. Biol. Med. 121 (2020) 103766, http://dx.doi.org/10.1016/j.compbiomed.2020.103766.

[19] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432, http://dx.doi.org/10.1007/978-3-319-46723-8_49.

[20] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2018, pp. 3–11, http://dx.doi.org/10.1007/978-3-030-00889-5_1.

[21] F. Milletari, N. Navab, S.A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV, IEEE, 2016, pp. 565–571, http://dx.doi.org/10.1109/3DV.2016.79.

[22] M.D. Cirillo, D. Abramian, A. Eklund, Vox2Vox: 3D-GAN for brain tumour segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 274–284, http://dx.doi.org/10.1007/978-3-030-72084-1_25.

[23] H. Chen, Z. Qin, Y. Ding, T. Lan, Brain tumor segmentation with generative adversarial nets, in: 2019 2nd International Conference on Artificial Intelligence and Big Data, ICAIBD, IEEE, 2019, pp. 301–305, http://dx.doi.org/10.1109/ICAIBD.2019.8836968.

[24] L. Zhu, Q. He, Y. Huang, Z. Zhang, J. Zeng, L. Lu, et al., DualMMP-GAN: Dual-scale multi-modality perceptual generative adversarial network for medical image segmentation, Comput. Biol. Med. 144 (2022) 105387, http://dx.doi.org/10.1016/j.compbiomed.2022.105387.

[25] Q. Delannoy, C.H. Pham, C. Cazorla, C. Tor-Díez, G. Dollé, H. Meunier, et al., SegSRGAN: Super-resolution and segmentation using generative adversarial networks—Application to neonatal brain MRI, Comput. Biol. Med. 120 (2020) 103755, http://dx.doi.org/10.1016/j.compbiomed.2020.103755.

[26] M. Ebner, G. Wang, W. Li, M. Aertsen, P.A. Patel, R. Aughwane, et al., An automated localization, segmentation and reconstruction framework for fetal brain MRI, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 313–320, http://dx.doi.org/10.1007/978-3-030-00928-1_36.

[27] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, et al., Dual attention network for scene segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154, http://dx.doi.org/10.48550/arXiv.1809.02983.

[28] Q. Hou, L. Zhang, M.M. Cheng, J. Feng, Strip pooling: Rethinking spatial pooling for scene parsing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4003–4012, http://dx.doi.org/10.1109/CVPR42600.2020.00406.

[29] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, W. Liu, Ccnet: Criss-cross attention for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 603–612.

[30] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19, http://dx.doi.org/10.48550/arXiv.1807.06521.

[31] J. Park, S. Woo, J.Y. Lee, I.S. Kweon, Bam: Bottleneck attention module, 2018, http://dx.doi.org/10.48550/arXiv.1807.06514, arXiv preprint arXiv:1807.06514.

[32] A.G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 421–429, http://dx.doi.org/10.1007/978-3-030-00928-1_48.

[33] X. Kong, G. Sun, Q. Wu, J. Liu, F. Lin, Hybrid pyramid u-net model for brain tumor segmentation, in: International Conference on Intelligent Information Processing, Springer, 2018, pp. 346–355, http://dx.doi.org/10.1007/978-3-030-00828-4_35.

[34] T. Henry, A. Carré, M. Lerousseau, T. Estienne, C. Robert, N. Paragios, et al., Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a BraTS 2020 challenge solution, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 327–339, http://dx.doi.org/10.1007/978-3-030-72084-1_30.

[35] Z. Jiang, C. Ding, M. Liu, D. Tao, Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task, in: International MICCAI Brainlesion Workshop, Springer, 2019, pp. 231–241, http://dx.doi.org/10.1007/978-3-030-46640-4_22.

[36] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241, http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[37] T. Zhou, S. Ruan, Y. Guo, S. Canu, A multi-modality fusion network based on attention mechanism for brain tumor segmentation, in: 2020 IEEE 17th International Symposium on Biomedical Imaging, ISBI, IEEE, 2020, pp. 377–380, http://dx.doi.org/10.1109/ISBI45749.2020.9098392.

[38] T. Zhou, S. Canu, P. Vera, S. Ruan, Latent correlation representation learning for brain tumor segmentation with missing MRI modalities, IEEE Trans. Image Process. 30 (2021) 4263–4274, http://dx.doi.org/10.1109/TIP.2021.3070752.

[39] S.T. Bukhari, H. Mohy-ud Din, E1D3 U-Net for brain tumor segmentation: Submission to the RSNA-ASNR-MICCAI BraTS 2021 challenge, 2021, http://dx.doi.org/10.48550/arXiv.2110.02519, arXiv preprint arXiv:2110.02519.

[40] L. Fidon, S. Shit, I. Ezhov, J.C. Paetzold, S. Ourselin, T. Vercauteren, Generalized wasserstein dice loss, test-time augmentation, and transformers for the BraTS 2021 challenge, 2021, http://dx.doi.org/10.1007/978-3-031-09002-8_17, arXiv preprint arXiv:2112.13054.

[41] L.M. Ballestar, V. Vilaplana, MRI brain tumor segmentation and uncertainty estimation using 3D-UNet architectures, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 376–390, http://dx.doi.org/10.1007/978-3-030-72084-1_34.

[42] A.M.G. Allah, A.M. Sarhan, N.M. Elshennawy, Edge U-Net: Brain tumor segmentation using MRI based on deep U-Net model with boundary information, Expert Syst. Appl. 213 (2023) 118833, http://dx.doi.org/10.1016/j.eswa.2022.118833.

[43] M. Marcinkiewicz, J. Nalepa, P. Lorenzo, W. Dudzik, G. Mrukwa, Automatic brain tumor segmentation using a two-stage multi-modal fcnn, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 2018, pp. 13–24.

[44] C. Lyu, H. Shu, A two-stage cascade model with variational autoencoders and attention gates for MRI brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 435–447, http://dx.doi.org/10.1007/978-3-030-72084-1_39.

[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144, http://dx.doi.org/10.1145/3422622.

[46] H. Peiris, Z. Chen, G. Egan, M. Harandi, Reciprocal adversarial learning for brain tumor segmentation: a solution to BraTS challenge 2021 segmentation task, 2022, http://dx.doi.org/10.48550/arXiv.2201.03777, arXiv preprint arXiv:2201.03777.

[47] Z. Wang, J. Chen, S.C. Hoi, Deep learning for image super-resolution: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (10) (2020) 3365–3387, http://dx.doi.org/10.1109/TPAMI.2020.2982166.

[48] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307, http://dx.doi.org/10.1109/TPAMI.2015.2439281.

[49] J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645, http://dx.doi.org/10.1109/CVPR.2016.181.

[50] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, et al., Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883, http://dx.doi.org/10.1109/CVPR.2016.207.

[51] M. Arjovsky, L. Bottou, Towards principled methods for training generative adversarial networks, 2017, arXiv preprint arXiv:1701.04862.

[52] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690, http://dx.doi.org/10.1109/CVPR.2017.19.

[53] M.U. Rehman, J. Ryu, I.F. Nizami, K.T. Chong, RAAGR2-Net: A brain tumor segmentation network using parallel processing of multiple spatial frames, Comput. Biol. Med. (2022) 106426, http://dx.doi.org/10.1016/j.compbiomed.2022.106426.

[54] J.H. Xu, W.P.K. Teng, X.J. Wang, A. Nürnberger, A deep supervised U-attention net for pixel-wise brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 278–289, http://dx.doi.org/10.1007/978-3-030-72087-2_24.

[55] W. Jun, X. Haoxiang, Z. Wang, Brain tumor segmentation using dual-path attention U-net in 3D MRI images, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 183–193, http://dx.doi.org/10.1007/978-3-030-72084-1_17.

[56] Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13713–13722, http://dx.doi.org/10.1109/CVPR46437.2021.01350.

[57] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134, http://dx.doi.org/10.1109/CVPR.2017.632.

[58] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141, http://dx.doi.org/10.1109/CVPR.2018.00745.