Multi-input Unet model based on the integrated block and the aggregation connection for MRI brain tumor segmentation

Lingling Fang（✉ fanglingling1985@163.com ）

Liaoning Normal University   https://orcid.org/0000-0002-4397-7212

Xin Wang

Liaoning Normal University

Multi-input Unet model based on the integrated block and the aggregation connection for MRI brain tumor segmentation

Lingling Fang[a,*], Xin Wang[a]

ª Department of Computing and Information Technology, Liaoning Normal University, Dalian City,

Liaoning Province, China

(Corresponding e-mail: fanglingling@lnnu.edu.cn).

Abstract: Due to the unresectability of normal brain tissue and the extensive invasive growth of the malignant tumor, the boundary between the tumor and surrounding healthy brain tissue or blood vessels is blurred, which greatly affects the accuracy of diagnosis and treatment. Meanwhile, with the growth of data information and the development of computer equipment, it is extremely time-consuming and laborious to rely on the traditional manual segmentation of brain medical images. To solve the above problems, this paper proposes a multi-input Unet model based on the integrated block and the aggregation connection to achieve efficient and accurate segmentation of tumor structure. Besides, this paper studies two-dimensional (2D) transverse brain tumor slices to meet the needs of doctors in the actual diagnosis. It solves the low-resolution problem in sagittal and coronal planes, which can effectively improve memory efficiency. The proposed algorithm is innovative in three aspects. Firstly, by inputting the mask images which can effectively represent the tumor location characteristics, it can provide more information about the spatial relationship to alleviate the problems of fuzzy boundary and low contrast between the lesion region and healthy brain tissue. Then, the integrated block extracts the tumor local information in different receptive domains by a multi-scale convolution kernel. The aggregation connection realizes the implicit deep connection of context information, which combines the shallow and deep information of the brain with strong geometric

spatial relationships. Meanwhile, to effectively alleviate the waste of memory resources caused by redundant and background information in medical images, the amount of calculation in model training is reduced by dimension reduction of the feature map. It can also overcome the gradient vanishing problem caused by network deepening. In this paper, an ablation experiment is used to verify the innovation of the proposed algorithm on the BraTS dataset, which compares with the state-of-the-art brain tumor segmentation methods. The accuracy of the proposed multi-input Unet model for the whole tumor and core lesion is 0.92 and 0.90, respectively.

1 Introduction

Glioma [1, 2] is the most common disease in the central nervous system and its incidence rate is about 40% [3, 4]. In terms of pathological diagnosis and prognosis, glioma is a disease with the highest recurrence rate and relatively short survival cycle in primary brain tumors. The World Health Organization divides glioma into four grades. Here, grades I and II are low-grade glioma (LGG); grades III and IV are high-grade glioma (HGG) [5, 6]. Relatively, the growth rate of the LGG is slow and the treatment can effectively improve the survival cycle of patients. The HGG has the tendency of faster recurrence and development to a higher grade tumor, which poses a great threat to the health. Therefore, how to effectively determine the tumor grade and timely diagnosis plays an important role in improving the life cycle of patients. Recently, with the development of computed tomography (CT) [7-9] and magnetic resonance imaging (MRI) [10-13], the detection rate of brain tumors has been greatly improved, which plays an increasingly important role in cancer diagnosis [14-16], surgical

planning [17, 18], and treatment outcome evaluation [19, 20]. Compared with the CT images, the MRI

images [21, 22] have higher soft-tissue resolution and can detect more tiny lesions in early time, which

is of great clinical significance to improve the survival rate of patients. Besides, because the MRI

images can obtain non-invasive vascular and water imaging, it can analyze more tissues and

pathological metabolism without causing damage [23]. Considering the generation need of the MRI

images, the human body needs to be placed in a strong magnetic field and the hydrogen protons are

excited by radio-frequency pulses [24, 25]. Then, the nuclear magnetic resonance is received signals

from protons to generate multi-angle and multi-modal images, which can reflect brain diseases more

effectively.

Fig. 1.   The multi-modal MRI medical images of brain tumors. (a) HGG; (b) LGG.

The multi-modal MRI brain medical images [26-28] include the T1 weighted images (T1), the T2

weighted images (T2), the T1 contrast images (T1c, The contrast agent is used in the T1 weighted

images), and the fluid attention inversion recovery images (FLAIR) [29, 30]. Figure 1 shows the

imaging status of the HGG and the LGG in multi-modal images. The T1 and the T2 images are used to

observe the anatomical structure and lesion structure of the brain, respectively. From Figure 1, the color

contrast of gray matter, white matter, and cerebrospinal fluid in the T1 images is obvious, which can

effectively reflect the difference between brain tumor and other healthy tissues. The T1c images can

distinguish lesion from the non-tumor region by blood contrast agent, which makes the tumor with rich

blood flow display as the high-intensity signal. As T2 imaging is related to water content, the signal

intensity of the lesion is significantly stronger than that of the surrounding normal brain tissue.

Therefore, it can effectively identify tumor with obvious edema occupying effect and obtain more clear

boundary information. The FLAIR images can suppress the high-intensity signal in cerebrospinal fluid

to make the lesions adjacent display more clearly. However, due to the invasive growth of brain tumors and the increasing amount of medical data, the traditional manual segmentation method relying on doctors has some limitations.

Recently, with the development of deep learning, how to simulate the abstract cognition of the human brain and process the complex mass of data information accurately and efficiently, using computer-aided medical disease diagnosis has become one of the key problems. The performance of the deep learning model is mainly related to four aspects: network depth [31-34], feature map width [35,], kernel scale [36, 37], and kernel step size [38]. Here, with the increase of network depth, the non-linear transformation of convolution in each layer can enhance the non-linear fitting ability. However, when the depth arrives at a certain extent, the model to an over-fitting problem, which leads to the problem of gradient vanishing or gradient exploding. To solve this problem, researchers have proposed activation function, batch standardization operation, and auxiliary loss function, which can slow down the network degradation, such as convolutional neural network (CNN) [39-43] and GoogLeNet model [44-46].

Kao et al. [47] propose a CNN model based on the block location information. By fusing the tumor information extracted from different advanced networks, the uncertainty can be reduced and the accuracy can be effectively improved. Russo et al. [48] propose a depth CNN model for brain tumor segmentation based on the spherical coordinate transformation pre-processing operation. This algorithm works independently of resolution and images settings by assuming three-dimensional (3D) spherical coordinate transformation, which can effectively improve the learning accuracy. Zhang et al. [49] propose a deep network structure of cross pattern learning features. By extracting information between different modal data and converting into features, this model can effectively improve the

performance of brain tumor cell segmentation. The above algorithms use the way of deepening the network depth to improve the feature extraction ability, which can overcome the problem of a feature missing caused by convolution operation to a certain extent. However, this operation may provide more information for the model by adding additional eigenvalues, which can result in a lot of redundant computation. On this basis, Chandra et al. [50] propose an end-to-end brain tumor segmentation model to optimize the deep learning network by calculating the generalized dice overlap loss. Sharif et al. [51] propose a depth neural network based on the pixel enhancement along with the limit, which has good segmentation performance by active feature selection of entropy filling. The algorithm proposed by Chandra et al. [50] and Sharif et al. [51] can reduce the amount of unnecessary information by optimizing the loss function. Thus, it can effectively improve the computational efficiency, but the model training still needs a large number of datasets. To solve this problem, Nassar et al. [52] input the narrow-band information in the integrated MRI images into the CNN model, which can obtain more accurate brain tumor anatomical information. Kayalıbay et al. [53] propose an improved CNN model based on the filter algorithm, which effectively alleviates the problem of high-class imbalance by combining the feature maps of long skip connected lesions. However, the problem of high memory and resource consumption caused by high-dimensional medical images has not been greatly improved.

Fig. 2. The structure of the traditional Unet model.

Besides, by adding the network feature maps, more features can be obtained and the generalization ability can be effectively enhanced. However, as the scale of the feature map increases, a large number of redundant data will appear in the model. Therefore, many researchers propose to replace the performance of large-scale convolution by continuous multiplexing of small-scale convolution, such as the AlexNet model [66-68] and the VGG model [72, 73]. Adding the network feature maps can

effectively enhance the non-linear approximation ability, increase the width of the network, and reduce the cost of memory resources, but more training data are needed to train the deep network. However, due to the variety of brain tumors and the small difference between the gray amplitude of healthy brain tissue, the accuracy of gray value can not be guaranteed. Because the Unet model [54-60] constructs the encoder-decoder structure through the skip connection, which ensures the maximum extraction of lesion features in the case of a small amount of data, as shown in Figure 2.

In the traditional Unet model, skip connection is used to link low-level and high-level features in one stage, instead of direct monitoring and loss back propagation, which ensures that more local features are fused into the feature maps. However, the convolution of the traditional Unet model will cause the loss of image details. Although an up-sampling operation can recover part of the lost fine boundary information, it still does not meet the higher requirements of medical image segmentation accuracy. To solve the above problems, Zhang et al. [61] improve the combination of high-level and low-level features by adding an attention mechanism. It can enhance the expression of local features in the Unet model and further reduce the loss of the network in the feature extraction process. However, due to the high requirements of equipment and memory, Nitnaware et al. [62] propose an improved Unet model for brain tumor segmentation, which converts the 3D brain tumor images into two-dimensional (2D) slice images through the DICOM conversion tool. McHugh et al. [63] propose an automatic glioma segmentation algorithm based on a 2D density unit. By dividing the 3D images into 2D slices, the above algorithms can better alleviate the problem of large memory resources required by the depth network. On this basis, Yang et al. [64] propose an improved brain tumor segmentation model based on Unet model. By deepening the network depth and using the compression operator to control the parameters in the model training process, the feature extraction ability is improved and the network

degradation phenomenon is better overcome. Ahmad et al. [65] extract multi-context information from the concept of feature reusability by using densely connected blocks. Besides, it can extract local and global information by combining features of different kernel scales. Thus, the non-linear feature extraction ability is effectively enhanced and the segmentation performance is improved. However, in the process of network training, there is still the problem that useful features are lost or not used. Therefore, this paper proposes a multi-input Unet model based on the integrated block and the aggregation connection for MRI brain tumor segmentation.

To improve the efficiency and meet the actual needs of physician diagnosis, this paper uses multi-input 2D slices to train the proposed Unet model. Meanwhile, due to the particularity of MRI medical image sampling, the contrast between brain tumor and surrounding healthy tissue is often low, which affects the judgment of the dynamic relationship. To effectively obtain the feature information of atypical brain tumors, the proposed algorithm inputs the mask images, which can correctly reflect the location and structure characteristics of brain tumors into the multi-input Unet model. Besides, through the integrated block and the aggregation connection, the large receptive domain feature maps are connected with the small-scale feature maps that have strong geometric detail expression ability. In the integrated block, the bottleneck convolution is used to effectively reduce the dimension of current eigenvectors of the convolution process. It can not only keep the sparsity of network structure, but also improve the computing performance by using a dense matrix. Meanwhile, the proposed multi-input Unet model combines the multi-scale local information in different stages with the global information, which can effectively reflect the structural characteristics of tumor location. The feature results of the upper layer are directly connected to the lower layer and used as supplementary information in the process of up-sampling operation. It also makes up for the problem of a feature missing caused by the

down-sampling operation. This proposed algorithm can not only effectively overcome the gradient vanishing problem, but also obtain more tumor information. The main contributions of the proposed algorithm are as follows:

- **Data processing**: To alleviate the problem of fuzzy boundary and low contrast between the tumor and the healthy brain tissue, the algorithm inputs the mask images that can effectively reflect the location and structure of brain tumors into the proposed multi-input Unet model.

- **Model design**: The proposed algorithm can effectively combine the shallow information with the deep information, which improves learning ability through the combination of the integrated block and the aggregation connection. The local information is obtained by the integrated block, which provides more tumor features for model training. The context information is connected by the aggregation connection to enhance the feature strength of the proposed model implicitly and deeply.

- **Optimization process**: By expansion convolution and bottleneck convolution, the multi-scale feature analysis ability of the proposed model is enhanced. The expansion convolution uses continuous convolution to improve memory efficiency, which achieves large kernel utility. The bottleneck convolution reduces the number of calculation parameters, which alleviates the gradient disappearance phenomenon caused by the deepening of the network.

The paper is organized as follows. Section Ⅱ introduces the related algorithms in the training deep learning model, including the integrated block and the aggregation connection. Section Ⅲ introduces the experimental process of the proposed multi-input Unet model, including reconstruction of the multi-input dataset, training of the proposed Unet model based on the integrated block and the aggregation connection, and fusion of context information. Section Ⅳ shows the experimental results

and analysis to verify the effectiveness of each innovation point through ablation research. Besides, it is compared with the state-of-the-art brain tumor segmentation algorithm. Finally, the conclusion is given in section Ⅴ.

2 The related algorithm

Since the AlexNet model [66-68] won the ImageNet competition in 2012, the DNN [69-71] model has rapidly become one of the hot research topics. Then, the VGG model [72, 73], the GoogleNet model [44-46], the ResNet model [74, 75], and so on are improved from the following three aspects:

● Through the design of a more complex deep network structure, the ability of extracting deeper feature information is lifted, i.e., the GoogleNet model.

● By adding pre-processing and post-processing operations, the calculation parameters are simplified and the segmentation performance is improved, i.e., training model using multi-modal images.

● To alleviate the problem of a feature missing caused by convolution operations, special loss functions or links that are more suitable for deep networks are explored, i.e., the Unet model.

On this basis, the proposed algorithm overcomes the network degradation problem through the integrated block and the aggregation connection, which effectively improves the segmentation performance by extracting deeper features.

2.1 The integrated block

The traditional deep learning model directly uses continuous convolution to increase the dimension of the feature maps, which will aggravate the over-fitting phenomenon and lead to an exponential increase in the calculation. To solve this problem, the inception unit proposed in the GoogleNet model [44-46] transforms sparse matrix clustering into dense sub-matrix, which improves the performance of

computation. The corresponding parameter is $\sum_i K_i * K_i * F_{Ki} * N * N * M_i \quad (i = 1,3,5,6)$, where K_i

is the size of the i-th convolution kernel, F_{Ki} is the corresponding dimension. N and M are the

size and dimension of the obtained feature map, respectively, as shown in Figure 3. Although the

problem of redundant information is alleviated to a certain extent, the high-dimensional convolution

operation will lead to the loss of feature information.

Inspired by the inception unit, this paper improves the performance by building the integrated block

to cluster the sparse matrix into a dense sub-matrix. To effectively obtain the tumor local features, this

paper uses convolution linkage with different scales, which obtains the deep features in multi-scale

kernels. Besides, the proposed algorithm uses the bottleneck convolution to reduce the dimension of

the current eigenvector. It can debase the amount of computation effectively and improve the utilization

of computing resources. The integrated block can not only keep the sparsity of network structure, but

also improve the computing performance by using a dense matrix, as shown in Figure 3. Through 1*1

bottleneck convolution, the dimension reduction operation of large-scale feature information (That is,

the K_1, K_2, K_3, K_4, K_5, and K_6 values are 1, respectively) is carried out. Then, it will reduce

the dimension of certain strong feature signals, which is conducive to accelerated training. Therefore,

the calculation cost of the proposed algorithm is $\sum_{i=1}^{6} F_{Ki} * N * N * M_i$, which can balance the

relationship between network depth and density effectively. The number of tumor feature maps

extracted from each integrated block path is balanced by dimension reduction, which ensures that the

gradient vanishing problem can be solved while deepening the model depth.

Fig. 3. The integrated block.

By increasing the depth and width of the network, the segmentation performance of the proposed

model can be effectively improved. However, it will increase the redundant computation in network

training and reduce the segmentation efficiency. Therefore, the aggregation connection is used to link the deep and shallow features.

2.2 The aggregation connection

Although multi-modal MRI tumor images can make up for the lost features to a certain extent, the boundary between edema and healthy region is blurred because of the invasive growth of the tumor. Due to the insufficient information on brain tumor structure, more deep and shallow features are needed in lesion segmentation. To effectively solve the above problem, the Unet model [54-60] can obtain more global and local information with a small amount of data. Through U-shaped structure and skip connection operation, the deep information obtained by down-sampling operation can be restored to the original image scale. However, because part of the boundary information will be lost in the process of feature extraction, the corresponding larger feature maps obtained by the up-sampling operation are not complete. Even though the connection between shallow and deep features, the residual information still can not meet the needs of tumor segmentation. To solve the above problem, the ResNet model [74, 75] obtains the feature information of the layer L by adding the feature signal X_{L-1} and the non-linear transformation H_{L-1} value of the layer $L-1$:

$$X_L = H_{L-1}(X_{L-1}) + X_{L-1} \tag{1}$$

The ResNet model effectively enhances the non-linear mapping ability and alleviates the problem of a feature missing caused by convolution operation to a certain extent, as shown in Figure 4(a). However, this network requires a large amount of computation and takes up more resource space, which leads to low computational efficiency. Therefore, the proposed model adopts the aggregation connection and the corresponding structure is shown in Figure 4 (b).

Fig. 4.   The aggregation connection. (a) The ResNet model; (b) The proposed model.

The aggregation connection directly links the feature results of the upper layer to the lower layer, which is used as supplementary information in the up-sampling process to make up for the lost features caused by the down-sampling operation. The feature information of the layer L is obtained from all the non-linear changes H_{L-1} of the previous layer $L-1$:

$$X_L = H_{L-1}([X_0, X_1, ..., X_{L-1}]) \tag{2}$$

$$H_{L-1} = C_{b_1}^{L-1}(X_{L-2}) \oplus C_{b_2}^{L-1}(C_{b_1}^{L-1}(X_{L-2})) \oplus C_d^{L-1}(X_{L-2}) \oplus C_p^{L-1}(X_{L-2}) \tag{3}$$

where $C_{b_1}^{L-1}$, $C_{b_2}^{L-1}$, C_d^{L-1}, and C_p^{L-1} represent bottleneck convolution, convolution with batch normalization and ReLU, extended convolution after dimension reduction, and pooling operation, respectively. $\oplus$ is used to connect the pooling or convolution feature maps. The aggregation connection establishes the relationship between different layers. On the premise of making full use of the feature information, it narrows the network by bottleneck convolution, which greatly reduces the number of calculation parameters. Moreover, because the problem of network degradation is solved, it is possible to obtain more sampling information by building a deeper network.

3 The proposed method

In this paper, a multi-input Unet model for MRI brain tumor segmentation is proposed. Through feature analysis of different scales, the continuous 3*3 expansion convolution is used to obtain deep features with the large receptive domain. The 1*1 bottleneck convolution is used to decrease the dimension, which can effectively reduce the number of training parameters and greatly improve the utilization of memory resources. Meanwhile, with the increase of the depth and width of the proposed model, the feature maps obtained by the down-sampling operation become more and more abstract. Besides, its resolution decreases that leads to the loss of some tumor information and the corresponding spatial geometric features.

Fig. 5.   The flow chart of the proposed algorithm.

To overcome this problem, the integrated block is used to connect the local feature information of different scales. The large receptive domain is convenient to connect the location of the tumor with other brain structures, which can effectively reduce the error caused by the high signal feature identification of lesions. The small kernel can provide the necessary resolution information for identifying small lesions. By connecting feature maps of different scales, the ability of tumor geometric expression and lesion differentiation can be effectively improved. With the increase of down-sampling times, the range of receptive domain extracted by the model expands. However, the lost local tiny information leads to the inability to identify small lesion effectively. Therefore, in this paper, the aggregation connection method is used to merge the up-sampling and down-sampling feature maps of the same scale. To express the geometric details, the small and the large receptive domains are combined, which makes the strong ability to distinguish lesions. The experimental process is mainly divided into three steps: reconstruction of the multi-input dataset, training of the proposed Unet model based on the integrated block and the aggregation connection, and fusion of context information. The specific process is shown in Figure 5.

3.1 The multi-input images

Due to the small number of transverse slices in multi-modal MRI brain tumor images, there is anisotropy in sagittal and coronal planes, which are much lower than that in the transverse plane, as shown in Figure 6. Besides, using low-resolution 3D images for tumor segmentation can not better express the brain features with different shapes and positions. Generally, doctors use 2D slice images to observe when they actually determine the type of tumor and the treatment plan. Therefore, to meet the actual needs of brain tumor segmentation and improve the utilization of resources, the proposed

algorithm is used to segment brain tumors on 2D slice images of the transverse plane.

Fig. 6.   The MRI images of brain tumors from different angles.

Because the boundary between tumor and healthy tissue is fuzzy, it is difficult to segment brain tumor structure only by single-modal images. Therefore, how to effectively integrate the feature information of multi-modal images is a hot topic in tumor segmentation. The structure of the traditional multi-modal MRI images training deep learning model is shown in Figure 7(a). Due to the different density signals of lesions in different multi-modal images, high-performance segmentation results can be obtained. However, consider that the mutual exclusion of multi-modal images and the feature loss of the network, the segmentation accuracy needs to be improved further. To grasp the features of the lesion quickly, the mask images are used as the training dataset of the proposed model, which can reflect the spatial relationship between the brain tumor and the surrounding healthy tissues. Specifically, the mask images instead of the T1 images are used to overcome the mutual exclusion phenomenon. Besides, it can effectively combine tumor context information and local lesion features to predict the attribution of pixels. Therefore, the proposed algorithm provides more features for MRI brain tumor segmentation through multi-input operation, as shown in Figure 7(b).

Fig. 7.   The process of the proposed multi-input Unet model. (a) Traditional Unet model; (b)The proposed model.

Through the region growing algorithm, the connected information can be obtained, which can provide fine boundary features between the tumor and the healthy tissue. Then, the mask images I' is obtained by multi-modal integration, which can directly reflect the shape, scale, and appearance of a brain tumor. The calculation process is as follows:

$$I'(x, y) = \begin{cases} r_1 & \text{if} \quad \sum_{m=1}^{4} I'_{m,1}(x, y) \geq 3 \\ r_2 & otherwise \end{cases} \tag{4}$$

where $I'_{m,1}$ and $I'_{m,2}$ represent the lesion region (value 1) and non-lesion region (value 0) in

modality m, respectively. r_1 and r_2 indicate that the points (x, y) belong to the tumor region or healthy region. The region growing algorithm is used to obtain the single-modal lesion information, as shown in Algorithm 1.

Algorithm 1 Region growing algorithm

Input: Single-modal image I_m, seed points $S_{1,\dots,n}$, where n is the number of the seed points,

      height and *width* of the image I_m.

Output: Mask image I'_m

Initialization: GrowQue= $S_{1,\dots,n}$, GrowQue'=GrowQue

1   while GrowQue $\neq$ GrowQue'
2   while $i < height$
3   while $j < width$
4   if $\min(S_{1,\dots,n}) \leq I(i,j) \leq \max(S_{1,\dots,n})$
5   add $I(i,j) \rightarrow$ GrowQue
6   end
7   end
8   end
9   end
10 GrowQue'=GrowQue

Then, to effectively overcome the mutual exclusion phenomenon [76-78] (T1 images and T1c images), multi-input images (T1c images, T2 images, FLAIR images, and mask images) are used to train the proposed model, which can effectively alleviate the low contrast between the brain tumor and the healthy tissue.

Fig. 8.   The mask images by region growing algorithm. (a) 2D slices; (b) The corresponding mask images.

Figure 8 shows a set of mask images obtained by region growing algorithm from continuous 2D slices. By providing the mask images, the multi-input images can provide more information for the deep learning model, which assist to derive the different features of lesion structure and obtain more accurate segmentation results.

3.2 The proposed multi-input Unet model

The proposed model is composed of encoder and decoder, as shown in Figure 9. The encoder consists of two convolution layers and three integrated block layers. After each convolution layer, batch normalization and ReLU layers are used to tune up the linear unit loss. To speed up the training and reduce the sensitivity to network initialization, the proposed model uses a batch normalization layer between the convolution layer and the non-linear layer:

$$f(x) = \frac{x}{(\alpha + \frac{\beta * ChannelTotal}{ChannelSize})^\eta}$$ (5)

where x and $f(x)$ represent input and output elements, respectively. α, β, and η are the super parameters in normalization [79], where α, β, and η are set to 2, 0.0001, and 0.75, respectively. *ChannelSize* is the size of the feature channel. *ChannelTotal* is the sum of the feature channel. To improve the accuracy of brain tumor segmentation network, the non-linear activation layer ReLU $f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$ is used to perform threshold operation on each element x.

Fig. 9.   The encoder and decoder of the proposed model.

The integrated block is composed of four paths. Each path has a 1*1 bottleneck convolution to achieve dimension reduction, which improves the computational efficiency. Meanwhile, the local feature maps in different scales are obtained by three paths. It can improve the ability of abstract feature extraction effectively. The fourth path is pooling operation, which reduces the number of computing parameters by bottleneck convolution without losing the extracted information. It can also provide a simplified feature map of the same scale for the up-sampling operation.

The decoder consists of four up-sampling layers and two aggregation connection layers, which makes the tumor information flow along the multipath to achieve implicit depth supervision. Two convolution layers are added after each up-sampling layer to recover the resolution and generalization information of the feature maps. Besides, to prevent the information loss caused by the sharp

dimension reduction, the slow decrease is used to retain the brain tumor information by the convolution operation. Meanwhile, the aggregation connection realizes the feature connection between encoder and decoder with the same resolution, which can combine the shallow and deep information. Through the connection of local and global features, the proposed model effectively reduces the gradient vanishing phenomenon caused by the increase of network width and depth. To get the final result of brain tumor segmentation, 16 feature components are mapped into four kinds of labels by 1*1 bottleneck convolution, as shown in Tab. 1. To realize multi-class generalization, the softmax layer calculates the loss $LOSS$ of classification problem:

$$LOSS = LOSS_{CE} + LOSS_{dice} \qquad (6)$$

$$LOSS_{dice} = -\frac{2\sum_{n=1}^{N} P_n \mathsf{g} GT_n + \sigma}{\sum_{n=1}^{N} P_n + \sum_{n=1}^{N} GT_n + \sigma} \qquad (7)$$

$$LOSS_{CE} = -\frac{1}{N}\sum_{n=1}^{N}(GT_n \mathsf{g}\log(P_n) + (1-GT_n)\mathsf{g}\log(1-P_n)) \qquad (8)$$

where $LOSS_{CE}$ and $LOSS_{dice}$ represent cross entropy loss and dice loss, respectively. P_n and GT_n represent the classification results of pixel n in the proposed model and ground truth, respectively. To avoid the situation that the classification results in the proposed model and ground truth exist 0, the constant σ is set to 10^{-8}. N is the total number of samples, g represents scalar multiplication.

Tab. 1. The structure of the proposed Unet model.

3.3 Fusion of context information

Because the 3D MRI medical images are reconstructed by interpolation, the resolution of the coronal plane and sagittal plane is much lower than that of the transverse plane. According to the doctor's experience and MRI acquisition protocol, the proposed model can segment the transverse plane, which effectively represents the shape and size of the brain tumor. Compared with the segmentation of 3D

images, the proposed model can overcome the problem of low computational efficiency caused by a large amount of data. By reducing the dimension of brain tumor images, the complexity and memory requirement are significantly reduced. However, the brain context information of the functional and spatial geometric features is partially lost in the process of 2D slice acquisition. To solve this problem effectively, the proposed algorithm obtains the points (x_t, y_t) on the current slice T by fusing the points (x_{t-1}, y_{t-1}) and (x_{t+1}, y_{t+1}) on the corresponding positions of two adjacent slices $T-1$ and $T+1$. The concrete process is as follows:

Case 1: The region of the current point (x_t, y_t) is different from that of the corresponding points on the adjacent slices:

(1) If points (x_{t-1}, y_{t-1}) and (x_{t+1}, y_{t+1}) are consistent, the current point (x_t, y_t) is revised to region of the corresponding adjacent points;

(2) If points (x_{t-1}, y_{t-1}) and (x_{t+1}, y_{t+1}) are different and the current point (x_t, y_t) is the same as one of them, the current result will be retained.

(3) If the results of three points are not consistent, the context scope is expanded to judge the region of the corresponding points on $T-2$ and $T+2$ slices.

Case 2: The current point (x_t, y_t) and its adjacent points belong to the same region and the result of the current point is retained.

By fusing the context information, i.e., functional and spatial geometric features, the wrongly divided lesion points can be effectively corrected, which can tackle the partial loss in the process of 2D slice acquisition.

4 Experiment and analysis

4.1 Dataset and evaluation indexes

All the medical images used in this experiment are from the public dataset provided by the multi-modal brain tumor segmentation challenge (BraTS) [80, 81]. The pre-processing procedure includes co-registration of the same anatomical template and skull dissection. Each group of MRI brain tumor images contains four multi-modal information: T1 images, T1c images, T2 images, and FLAIR images. Meanwhile, it includes ground truth, which is segmented manually according to the same annotation protocol. By neuroradiologists with rich clinical experience, the results of ground truth labeling can effectively provide accurate information of healthy brain tissue, whole tumor, and core lesion. The BraTS 2018 dataset consists of 210 high-grade gliomas cases and 75 low-grade gliomas cases. The experimental training set consists of 20 groups of HGG and 5 groups of LGG samples. The testing set consists of 5 groups of HGG and LGG cases.

To effectively evaluate the performance of the proposed multi-input Unet model in brain tumor segmentation, this paper uses three indicators: accuracy (Acc), sensitivity (Sen), and specificity (Spe). To assess the classification results for healthy tissue, whole tumor, and core lesion, the relevant calculation formulas are as follows:

$$Acc(i) = \frac{|P_T(i) \cap GT_T(i)| + |P_F(i) \cap GT_F(i)|}{|P_T(i)| + |GT_T(i)| + |P_F(i)| + |GT_F(i)|} *100\% \qquad (9)$$

$$Sen(i) = \frac{|P_T(i) \cap GT_T(i)|}{|P_T(i) \cap GT_T(i)| + |P_T(i) \cap GT_F(i)|} *100\% \qquad (10)$$

$$Spe(i) = \frac{|P_F(i) \cap GT_F(i)|}{|P_F(i) \cap GT_F(i)| + |P_F(i) \cap GT_T(i)|} *100\% \qquad (11)$$

where $i = \{1, 2, 3, 4\}$ corresponds to the complete image, healthy region, whole tumor, and core lesion, respectively. P_T and P_F represent the pixel in the segmentation result whether belongs to or does not belong to the tumor region, respectively. Similarly, GT_T and GT_F represent whether the pixel is determined as the tumor region or healthy brain tissue in the ground truth, respectively.

4.2 Segmentation results of the proposed model

Fig. 10.   Segmentation results of the proposed model.

The segmentation results of the proposed multi-input Unet model are shown in Figure 10. Here, red,

green, and yellow represent core lesion, whole tumor, and healthy brain tissue, respectively. By using

the mask images with strong spatial features to replace the T1 images which conflicts with the T1c

images, the proposed model can obtain the segmentation results effectively. The accuracy of

segmenting the whole tumor and core lesion can reach 0.92 and 0.90, respectively. The curve of

accuracy and loss rate with the number of iterations is shown in Figure 11.

Fig. 11.   The accuracy and loss rate of the proposed model.

Figure 12 shows the feature maps of the proposed multi-input Unet model under different layers.

Figure 12(a) and (b) denote the integrated block in the encoder and the up-sampling operation in the

decoder, respectively. It can be seen that the shallow network has a strong ability to express geometric

information. However, the corresponding lesion features are relatively scarce. Through the integrated

block, the receptive domain is larger and its brain tumor features are more abundant. Yet, its spatial

geometric information is less. Therefore, through the aggregation connection, the proposed algorithm

links the rich local information with the global information, which can effectively represent the

location characteristics of the lesions.

Fig. 12.   The feature maps of the proposed model. (a) Part of the feature maps in the integrated blocks 1-3; (b) Part of the

feature maps in the up-sampling (Up-conv 1-4).

4.3 Analysis of the proposed method

The ablation experiment is adopted to prove the good performance of the proposed multi-input Unet

model. It compares the accuracy, sensitivity, and specificity from three aspects: multi-modal images,

learning rate, and multi-input images. Besides, by comparing with the state-of-the-art brain tumor

segmentation model, the performance of the proposed network is evaluated.

4.3.1 Analysis of the multi-modal images

Figures 13-14 show the performance of the proposed multi-input Unet model under different multi-modal images, where the numbers 1-4 represent single-modal images and 5-9 represent multi-modal images. Due to the pertinence of the single-modal images, the performance of the proposed network is more prominent in a certain index, but segmentation results of other parts are slightly insufficient. For example, the accuracy of the T1c images is relatively high. By adding a blood contrast agent before MRI examination, the T1c images can make the tumor with fast blood flow appear as high-intensity information, which further enriches the structural information. The FLAIR images and the T1c images are more sensitive for the whole tumor and core lesion. The FLAIR images can effectively provide healthy brain tissue and lesion information by suppressing normal cerebrospinal fluid signals. In general, the T1 images and T2 images are sensitive to brain anatomical structure and tissue lesions, respectively. Because most brain tumors are accompanied by tissue edema, the T2 images related to water content can effectively monitor the lesion structure. Conversely, the T1 images mainly provide the characteristic information to distinguish gray matter, white matter, and cerebrospinal fluid. Due to the limitations of single-modal images in the display of lesions on the brain surface and paraventricular, the multi-modal images are often used to judge the central nervous system diseases, as shown in Figure 13.

Fig. 13.   Segmentation results of the multi-modal images.

Through a large number of experiments, one can get that the performance of most multi-modal images in segmenting whole tumor and core lesion is better than that of single-modal images, as shown in Figure 14. However, only the T1+T1c multi-modal images (number 7) have performance

degradation. This is because there is mutual exclusion between the T1 images and the T1c images, which leads to the opposite results of tumor in the same location and affects the judgment of the proposed network. In this paper, the T1c images are chosen to distinguish tumor and non-tumor regions effectively. Here, the performance of FLAIR+T1c+T2+T1 (number 9) images is worse than that of the FLAIR+T1c+T2 (number 8) images. It is proved that the coexistence of the T1 and the T1c images will affect the segmentation results. The proposed model has the highest segmentation accuracy when inputting the FLAIR+T1c+T2 multi-modal images. The segmentation accuracy of the whole tumor and core lesion is 0.92 and 0.90, respectively.

Fig. 14.   The performance of the proposed multi-input Unet model in different multi-modal images. (The first and second rows represent the segmentation results of whole tumor and core lesion, respectively. The first to third columns represent three performance indicators: accuracy, sensitivity, and specificity, respectively. Here, numbers 1-4 represent the single-modal images :T2 images, T1 images, FLAIR images, and T1c images. Numbers 5-9 represent the multi-modal images, which are T2+T1c images, FLAIR+T1c images, T1+T1c images, FLAIR+T1c+T2 images, and FLAIR+T1c+T2+T1 images.)

4.3.2 Analysis of the learning rate

The influence on the performance of the proposed model mainly reflects two aspects: the size of the initial learning rate and the transformation scheme of the learning rate. The initial learning rate determines the step size of weight iteration during network training. The learning rate is transformed by a non-uniform reduction strategy and the reduced step interval is specified to improve the learning efficiency of the network. Figure 15 shows the curves of the accuracy and loss rate of the proposed multi-input Unet model under different learning rates. It can be seen that the convergence speed is faster when the initial setting value is large (e.g., $1 \times e^{-3}$), but the weights are changed along the gradient direction of each sample. This will lead to a serious oscillation phenomenon, which causes the

gradient exploding problem. Conversely, when the initial value is relatively small (e.g., $8 \times e^{-5}$), the convergence speed is slower. In the circumstances, the over-fitting phenomenon is easy to occur and the local optimal value may be obtained.

Fig. 15.   The accuracy and loss rate curves of the proposed model under different learning rates.

In general, the most ideal curve descent condition is slide descent (e.g., $1 \times e^{-4}$), which has stronger generalization performance and faster convergence speed. The accuracy of the proposed model will be improved with the enhancement of the learning rate and the corresponding training loss will also be decreased. However, when the learning rate reaches the critical value, the loss will increase. Besides, the proposed model may also change from under-fitting to over-fitting, especially for large datasets. The best learning rate can be selected from the region with the least loss. Thus, the initial learning rate set in this paper is set to $1 \times e^{-4}$ and the transformation scheme is to divide the learning rate by 10 after every 5 training sessions.

4.3.3 Analysis of the proposed multi-input images

Because multi-modal MRI images can reflect tumor necrosis and edema caused by cell proliferation, it is often used in the detection of brain tumor diseases. However, due to the invasive growth of brain tumor, the boundary between the edema region and the healthy brain tissue or capillaries is fuzzy. In these circumstances, the weak boundary information of the adhesion type can not be obtained well. Therefore, by inputting the mask images with strong spatial connectivity into the proposed multi-input Unet model, the geometric position relationship between the tumor and healthy brain tissue can be effectively obtained.

Figure 16 shows the results of the proposed model with and without the mask images under different multi-modal images. Here, rows 1-3 represent the results in single-modal, multi-modal, and multi-input

images, respectively. It can be seen that in the same group of experiments, the performance of the multi-input images with the mask images are significantly better than that without the mask images. By inputting the mask images, the results of glioma spreading morphology and pathological location can be directly reflected in the proposed algorithm, thus avoiding the weak boundary problem of uneven gray and density.

Fig. 16.   The performance of the proposed algorithm in different multi-input images.(where the numbers 1-9 represent T2 , T1, FLAIR, T1c, T2+T1c, FLAIR+T1c, T1+T1c, FLAIR+T1c+T2, and FLAIR+T1c+T2+T1 with the mask images, respectively. The numbers 1'-9' indicate the corresponding multi-input images without the mask images.)

4.3.4 Comparison with other algorithms

To further verify the performance of the proposed model, several state-of-the-art brain tumor segmentation models are compared, as shown in Tab. 2.

Tab. 2. The segmentation results of proposed multi-input Unet model and state-of-the-art models

From Tab. 2, the accuracy of the proposed multi-input Unet model for segmentation of whole tumor and core lesion is 0.92 and 0.90, respectively. Here, the algorithm proposed in references [48, 49, 52] inputs more lesion information into the deep learning network through pre-processing operation, which can improve the ability to obtain non-linear lesion features. Nassar et al. [52] input the integrated narrow-band information of brain tumor into the CNN model, which has good segmentation results, especially in the aspect of specificity (*Spe* as shown in row 12 of Tab. 2). However, this method does not overcome the problem of boundary feature loss caused by convolution, which leads to an incomplete boundary feature map obtained by the up-sampling operation. References [47, 53] combine tumor features extracted from different networks to effectively alleviate the high-class imbalance and memory requirements of the medical images. However, this network ignores the relationship between

the context information and the local receptive domain features, which leads to poor classification accuracy (*Acc* as shown in Tab. 2). References [51, 50] realize active feature selection by optimizing operation and entropy filling to segment lesion region from healthy tissue. Sharif et al. [51] propose a depth neural network based on pixel enhancement along with the limit and its complete segmentation accuracy can reach 0.98. It can not only accelerate in the correct gradient direction, but also restrain the occurrence of local extremum (as shown in row 6 of Tab. 2). However, there is a problem that the initial data weight is too low. Although it can be fixed by deviation correction, it is still easy to appear the gradient vanishing problem.

To solve this problem, references [61-65] improve the Unet model and make up for the loss of lesion boundary features in the process of down-sampling operation by the skip connection between low-level and high-level features. Here, Zhang et al. [61] introduce the attention mechanism to effectively reduce the loss of the network in the process of feature extraction. This method can enhance the ability of network feature extraction, but the performance is lower than that of the proposed algorithm (as shown in row 8 of Tab. 2). Nitnaware et al. [62] use 2D slices training network to obtain more compatible segmentation results and better alleviate the problem of large memory resources required. The accuracy and sensitivity of this network are 0.99 and 0.97, respectively. Yang et al. [64] use a compression operator to control the parameters in the training process of the improved Unet model. This algorithm uses a 3×3 kernel to obtain high-level features that can detect the location and texture of tumor features. Then, it accurately locates the tumor boundary and gray features by a 2×2 decoding operation. However, this algorithm still has a lot of computational redundancy and the efficiency of feature extraction needs to be improved (as shown in row 10 of Tab. 2). Ahmad et al. [65] use the dense connection to obtain the context information of lesions. By connecting feature maps of different widths

with a kernel of scale 3×3, this model can obtain more detailed features in brain tumors. If the input

image size is $m \times m @ c$, the calculation parameter is $3 \times 3 \times c \times c_1 + 3 \times 3 \times c \times c_2 + 3 \times 3 \times c \times c_3$, where

c_1, c_2, and c_3 are the width of the feature map in each channel (as shown in row 11 of Tab. 2). The

proposed algorithm uses bottleneck convolution in feature extraction of each scale, whose value is

$1 \times 1 \times c \times c_1{}' + 1 \times 1 \times c_1{}' \times c_1 + 1 \times 1 \times c \times c_2{}' + 1 \times 1 \times c_2{}' \times c_2 + 1 \times 1 \times c \times c_3{}' + 1 \times 1 \times c_3{}' \times c_3$, where $c_1{}'$, $c_2{}'$, and

$c_3{}'$ are the width of bottleneck convolution in each channel. It can be seen that the proposed model

can reduce the number of calculation parameters and recover the original information of lesions by

dimension reduction operation, which obtains good segmentation performance.

5 Conclusion

To obtain the lesion features efficiently, a brain tumor segmentation algorithm based on the

multi-input Unet model is proposed. By inputting the mask images that can quickly grasp the temporal

and spatial relationship between the lesion and the surrounding brain tissue, the problems such as

uneven data sampling density or fuzzy boundary of atypical tumor can be overcome. Then, the

proposed Unet model is optimized by the integrated block and the aggregation connection. Here, the

integrated block extracts the tumor local information in different receptive domains through a

multi-scale convolution kernel. The aggregation connection is used to realize the implicit depth

connection of context information, which can realize the combination of brain local and global

information of strong geometric spatial relationship. Meanwhile, to alleviate the problem of

overlapping redundancy and feature loss caused by convolution, the proposed algorithm uses

bottleneck convolution in each path of the integrated block. It not only reduces the amount of

calculation in model training by reducing the dimension of the feature map, but also alleviates the

problem of gradient vanishing caused by network deepening. In this paper, the ablation experiment is

used to verify the effectiveness of the innovations on the BraTS dataset. It can further prove that the proposed multi-input Unet model has obvious advantages over the state-of-the-art brain tumor segmentation methods.

Acknowledgements

References

[1] Touat M, Li Y Y, Boynton A N, et al., Mechanisms and therapeutic implications of hypermutation in gliomas, Nature, 580(7804) (2020) 517-523.

[2] Aslan K, Turco V, Blobner J, et al., Heterogeneity of response to immune checkpoint blockade in hypermutated experimental gliomas, Nature communications, 11(1) (2020) 1-14.

[3] Grabowski M M, Sankey E W, Ryan K J, et al, Immune suppression in gliomas. Journal of Neuro-oncology, 151(1) (2021) 3-12.

[4] Jiang T, Nam D H, Ram Z, et al., Clinical practice guidelines for the management of adult diffuse gliomas, Cancer Letters, 499 (2021) 60-72. doi: 10.1016/J.CANLET.2020.10.050

[5] Vallejo-Armenta P, Soto-Andonaegui J, Villanueva-Pérez R M, et al., [99mTc] Tc-iPSMA SPECT brain imaging as a potential specific diagnosis of metastatic brain tumors and high-grade gliomas, Nuclear Medicine and Biology, 96 (2021) 1-8. doi: 10.1016/J.NUCMEDBIO.2021.02.003

[6] Alis D, Bagcilar O, Senli Y D, et al., The diagnostic value of quantitative texture analysis of conventional MRI sequences using artificial neural networks in grading gliomas, Clinical radiology, 75(5) (2020) 351-357.

[7] Ramakrishnan T, Sankaragomathi B., A professional estimate on the computed tomography brain tumor images using SVM-SMO for classification and MRG-GWO for segmentation, Pattern Recognition Letters, 94 (2017) 163-171. doi:

10.1016/j.patrec.2017.03.026

[8] Jiang Y, Gu X, Wu D, et al., A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. IEEE/ACM transactions on computational biology and bioinformatics, 18(1) (2020) 40-52.

[9] Cui L, Zhang H., Study on threshold segmentation of multi-resolution 3D human brain CT image, Journal of Innovative Optical Health Sciences, 11(06) (2018) 1850037.

[10] Wadhwa A, Bhardwaj A, Verma V S., A review on brain tumor segmentation of MRI images. Magnetic resonance imaging, 61 (2019) 247-259. doi: 10.1016/j.mri.2019.05.043

[11] Shree N V, Kumar T N R., Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network, Brain informatics, 5(1) (2018) 23-30.

[12] Thaha M M, Kumar K P M, Murugan B S, et al., Brain tumor segmentation using convolutional neural networks in MRI images, Journal of medical systems, 43(9) (2019) 1-10.

[13] Abd-Ellah M K, Awad A I, Khalaf A A M, et al., A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned, Magnetic resonance imaging, 61 (2019) 300-318. doi: 10.1016/j.mri.2019.05.028

[14] Hollon T C, Pandian B, Adapa A R, et al., Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks, Nature medicine, 26(1) (2020) 52-58.

[15] Afshar P, Plataniotis K N, Mohammadi A., Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, (2019) 1368-1372. doi: 10.1109/ICASSP.2019.8683759

[16] Ostrom Q T, Cioffi G, Gittleman H, et al., CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016, Neuro oncology, 21 (2019) v1-v100. doi: 10.1093/neuonc/noz150

[17] Lu J, Zhang H, Hameed N U F, et al., An automated method for identifying an independent component analysis-based

language-related resting-state network in brain tumor subjects for surgical planning, Scientific reports, 7(1) (2017) 1-16.

[18] Wang G, Li W, Ourselin S, et al., Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation//International MICCAI Brainlesion Workshop. Springer, Cham, (2018) 61-72.

[19] Zhao X, Wu Y, Song G, et al., A deep learning model integrating FCNNs and CRFs for brain tumor segmentation, Medical image analysis, 43 (2018) 98-111. doi: 10.1016/j.media.2017.10.002

[20] van Linde M E, Brahm C G, de Witt Hamer P C, et al., Treatment outcome of patients with recurrent glioblastoma multiforme: a retrospective multicenter analysis, Journal of neuro-oncology, 135(1) (2017) 183-192.

[21] Mohan G, Subashini M M., MRI based medical image analysis: Survey on brain tumor grade classification, Biomedical Signal Processing and Control, 39 (2018) 139-161. doi: 10.1016/j.bspc.2017.07.007

[22] Amin J, Sharif M, Yasmin M, et al., A distinctive approach in brain tumor detection and classification using MRI, Pattern Recognition Letters, 139 (2017) 118-127. doi: 10.1016/j.patrec.2017.10.036

[23] Iqbal S, Ghani M U, Saba T, et al., Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN), Microscopy research and technique, 81(4) (2018) 419-427.

[24] Sun L, Zhang S, Chen H, et al., Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning, Frontiers in neuroscience, 13 (2019) 810. doi: 10.3389/fnins.2019.00810

[25] Arunkumar N, Mohammed M A, Mostafa S A, et al., Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks, Concurrency and Computation: Practice and Experience, 32(1) (2020) e4962.

[26] Wang Z, Cui Z, Zhu Y., Multi-modal medical image fusion by Laplacian pyramid and adaptive sparse representation, Computers in Biology and Medicine, 123 (2020) 103823. doi: 10.1016/j.compbiomed.2020.103823

[27] Aktar M N, Lambert A J, Pickering M., An automatic fusion algorithm for multi-modal medical images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6(5) (2018) 584-598.

[28] Srinivas B, Sasibhushana Rao G., Segmentation of Multi-Modal MRI Brain Tumor Sub-Regions Using Deep Learning, Journal of Electrical Engineering & Technology, 15 (2020) 1899-1909. doi: 10.1007/s42835-020-00448-z

[29] Gulzar Hameed Chagla, Reed F. Busse, Ryan Sydnor, Howard A. Rowley, Patrick A. Turski., Three-Dimensional Fluid Attenuated Inversion Recovery Imaging With Isotropic Resolution and Nonselective Adiabatic Inversion Provides Improved Three-Dimensional Visualization and Cerebrospinal Fluid Suppression Compared to Two-Dimensional Flair at 3 Tesla, Investigative Radiology, 43(8) (2008) 547–51.

[30] Hausmann D, Liu J, Budjan J, et al, Image quality assessment of 2D versus 3D T2WI and evaluation of ultra-high b-value (b= 2,000 mm/s2) DWI for response assessment in rectal cancer, Anticancer research, 38(2) (2018) 969-978.

[31] Zuo Y, Fang Y, Yang Y, et al., Residual dense network for intensity-guided depth map enhancement, Information Sciences, 495 (2019) 52-64. doi: 10.1016/j.ins.2019.05.003

[32] Guo C, Li C, Guo J, et al., Hierarchical features driven residual learning for depth map super-resolution, IEEE Transactions on Image Processing, 28(5) (2018) 2545-2557.

[33] Huang L, Zhang J, Zuo Y, et al., Pyramid-structured depth map super-resolution based on deep dense-residual network, IEEE Signal Processing Letters, 26(12) (2019) 1723-1727.

[34] Song X, Dai Y, Qin X., Deeply supervised depth map super-resolution as novel view synthesis, IEEE Transactions on circuits and systems for video technology, 29(8) (2018) 2323-2336.

[35] Khan A, Sohail A, Zahoora U, et al., A survey of the recent architectures of deep convolutional neural networks, Artificial Intelligence Review, 53(8) (2020) 5455-5516.

[36] Huang W, Cheng J, Yang Y, et al., An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, Neurocomputing, 359 (2019) 77-92. doi: 10.1016/j.neucom.2019.05.052

[37] Ricci E, Ouyang W, Wang X, et al, Monocular depth estimation using multi-scale continuous crfs as sequential deep networks, IEEE transactions on pattern analysis and machine intelligence, 41(6) (2019) 1426-1440. doi:

10.1109/TPAMI.2018.2839602

[38] Xu X, Pan J, Zhang Y J, et al., Motion blur kernel estimation via deep learning, IEEE Transactions on Image Processing, 27(1) (2017) 194-205.

[39] Iqbal S, Ghani M U, Saba T, et al., Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN), Microscopy research and technique, 81(4) (2018) 419-427.

[40] Chang J, Zhang L, Gu N, et al., A mix-pooling CNN architecture with FCRF for brain tumor segmentation, Journal of Visual Communication and Image Representation, 58(2019) 316-322. doi: 10.1016/j.jvcir.2018.11.047

[41] Yaqub M, Jinchao F, Zia M S, et al., State-of-the-Art CNN Optimizer for Brain Tumor Segmentation in Magnetic Resonance Images, Brain Sciences, 10(7) (2020) 427.

[42] Sajjad M, Khan S, Muhammad K, et al., Multi-grade brain tumor classification using deep CNN with extensive data augmentation, Journal of computational science, 30 (2019) 174-182. doi: 10.1016/j.jocs.2018.12.003

[43] Razzak M I, Imran M, Xu G., Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks, IEEE journal of biomedical and health informatics, 23(5) (2018) 1911-1919.

[44] Sun Henan, Xu Haowei, Liu Bin, He Dongjian, He Jinrong, Zhang Haixi, Geng Nan., MEAN-SSD: A novel real-time detector for apple leaf diseases using improved light-weight convolutional neural networks, Computers and Electronics in Agriculture, 189 (2021). doi: 10.1016/J.COMPAG.2021.106379

[45] Shivaprasad B J., Bidirectional ConvLSTMXNet for Brain Tumor Segmentation of MR Images, Tehnički glasnik, 15(1) (2021) 37-42.

[46] Amin J, Sharif M, Yasmin M, et al., A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning, Journal of medical systems, 43(11) (2019) 1-16.

[47] Kao P Y, Shailja S, Jiang J, et al., Improving Patch-Based Convolutional Neural Networks for MRI Brain Tumor Segmentation by Leveraging Location Information, Front. Neurosci. 13: 1449. doi: 10.3389/fnins, 2020.

[48] Russo C, Liu S, Di Ieva A., Spherical coordinates transformation pre-processing in Deep Convolution Neural Networks for brain tumor segmentation in MRI, arXiv preprint arXiv:2008.07090, 2020.

[49] Zhang D, Huang G, Zhang Q, et al., Cross-modality deep feature learning for brain tumor segmentation, Pattern Recognition, 110 (2021) 107562. doi: 10.1016/j.patcog.2020.107562

[50] Chandra S, Vakalopoulou M, Fidon L, et al., Context Aware 3D CNNs for Brain Tumor Segmentation// International MICCAI Brainlesion Workshop, Springer, Cham, (2018) 393–405.

[51] Sharif M I, Li J P, Khan M A, et al., Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images, Pattern Recognition Letters, 129 (2020) 181-189. doi:: 10.1016/j.patrec.2019.11.019

[52] Nassar S E, Mohamed M A, Elnakib A., MRI Brain Tumor Segmentation Using Deep Learning.(Dept. E). MEJ. Mansoura Engineering Journal, 45(4) (2021) 45-54.

[53] B. Kayalibay, G. Jensen, P. Smagt, CNN-based Segmentation of Medical Imaging while, (2017), arXiv: 1701.03056.

[54] Jin Q, Meng Z, Sun C, et al., RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans, Frontiers in Bioengineering and Biotechnology, 8 (2020) 1471.

[55] Alom M Z, Hasan M, Yakopcic C, et al., Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation, arXiv preprint arXiv:1802.06955, 2018.

[56] Li X, Chen H, Qi X, et al., H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE transactions on medical imaging, 37(12) (2018) 2663-2674.

[57] Welander P, Karlsson S, Eklund A., Generative adversarial networks for image-to-image translation on multi-contrast MR images-A comparison of CycleGAN and UNIT, (2018) arXiv preprint arXiv:1806.07777.

[58] Lin F, Wu Q, Liu J, et al., Path aggregation U-Net model for brain tumor segmentation, Multimedia Tools and Applications, (2020) 1-14.

[59] Zhang J, Lv X, Sun Q, et al., SDResU-Net: Separable and Dilated Residual U-Net for MRI Brain Tumor Segmentation,

Current Medical Imaging, 16(6) (2020) 720-728.

[60] Ding Huijun, Cen Qian, Si Xiaoyu, Pan Zhanpeng, Chen Xiangdong. Automatic glottis segmentation for laryngeal endoscopic images based on U-Net, Biomedical Signal Processing and Control, 71(PA) (2022). doi: 10.1016/J.BSPC.2021.103116

[61] Zhang J, Jiang Z, Dong J, et al., Attention Gate ResU-Net for automatic MRI brain tumor segmentation, IEEE Access, 8 (2020) 58533-58545. doi: 10.1109/ACCESS.2020.2983075

[62] Mishra Pradipta Kumar,Satapathy Suresh Chandra,Rout Minakhi., Segmentation of MRI Brain Tumor Image using Optimization based Deep Convolutional Neural networks (DCNN), Open Computer Science, 11(1) (2021) 380-390. doi: 10.1515/COMP-2020-0166

[63] McHugh H, Talou G M, Wang A., 2D Dense-UNet: A Clinically Valid Approach to Automated Glioma Segmentation//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II. Springer Nature, (2021) 69-80.

[64] Yang T, Song J, Li L, et al., Improving brain tumor segmentation on MRI based on the deep U-net and residual units, Journal of X-ray Science and Technology, 28(1) (2020) 95-110.

[65] Ahmad P, Qamar S, Shen L, et al., Context Aware 3D UNet for Brain Tumor Segmentation, (2020) arXiv preprint arXiv:2010.13082.

[66] Lu S, Lu Z, Zhang Y D., Pathological brain detection based on AlexNet and transfer learning, Journal of computational science, 30 (2019) 41-47. doi: 10.1016/j.jocs.2018.11.008

[67] Shakarami Ashkan and Menhaj Mohammad Bagher and Tarrah Hadis., Diagnosing COVID-19 disease using an efficient CAD system, Optik, 241 (2021) 167199-167199. doi:10.1016/J.IJLEO.2021.167199

[68] Shambhu Shankar, Koundal Deepika,Das Prasenjit, Sharma Chetan. Binary Classification of COVID-19 CT Images Using

CNN: COVID Diagnosis Using CT, International Journal of E-Health and Medical Communications (IJEHMC),13(2) (2021) 1-13.

[69] Qin P, Zhang J, Zeng J, et al., A framework combining DNN and level-set method to segment brain tumor in multi-modalities MR image, Soft Computing, 23(19) (2019) 9237-9251.

[70] Li Kai, Wang Juanle, Yao Jinyi. Effectiveness of machine learning methods for water segmentation with ROI as the label: A case study of the Tuul River in Mongolia, International Journal of Applied Earth Observation and Geoinformation, 103 (2021). doi: 10.1016/j.jag.2021.102497

[71] Yan K, Wang X, Kim J, et al., A propagation-DNN: Deep combination learning of multi-level features for MR prostate segmentation, Computer methods and programs in biomedicine, 170 (2019) 11-21. doi: 10.1016/j.cmpb.2018.12.031

[72] Singh V, Sharma S, Goel S, et al., Brain Tumor Prediction by Binary Classification Using VGG-16, Smart and Sustainable Intelligent Systems, (2021) 127-138. doi: https://doi.org/10.1002/9781119752134.ch9

[73] Saba T, Mohamed A S, El-Affendi M, et al., Brain tumor detection using fusion of hand crafted and deep learning features, Cognitive Systems Research, 59 (2020) 221-230. doi: 10.1016/j.cogsys.2019.09.007

[74] Lu S, Wang S H, Zhang Y D., Detecting pathological brain via ResNet and randomized neural networks, Heliyon, 6(12) (2020) e05625.

[75] Zhang F, Wang Q, Li H., Automatic Segmentation of the Gross Target Volume in Non-Small Cell Lung Cancer Using a Modified Version of ResNet, Technology in Cancer Research & Treatment, 19 (2020) 1533033820947484. doi: 10.1177/1533033820947484

[76] Wang Z, Liu C, Cheng D, et al., Automated detection of clinically significant prostate cancer in mp-MRI images based on an end-to-end deep neural network, IEEE transactions on medical imaging, 37(5) (2018) 1127-1139.

[77] Langner T, Östling A, Maldonis L, et al., Kidney segmentation in neck-to-knee body MRI of 40,000 UK Biobank participants, Scientific reports, 10(1) (2020) 1-10.

[78] Palmisano A, Benedetti G, Faletti R, et al., Early T1 myocardial MRI mapping: value in detecting myocardial hyperemia in acute myocarditis, Radiology, 295(2) (2020) 316-325.

[79] Krizhevsky A, Sutskever I, Hinton G E, ImageNet classification with deep convolutional neural networks, Communications of the ACM, 60(6) (2017) 84-90.

[80] Bakas S, Reyes M, Jakab A, et al., Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. (2019) arXiv preprint arXiv:1811.02629.

[81] Ghaffari M, Sowmya A, Oliver R., Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BraTS 2012–2018 challenges, IEEE reviews in biomedical engineering, 13 (2019) 156-168. doi: 10.1109/RBME.2019.2946868

Figures
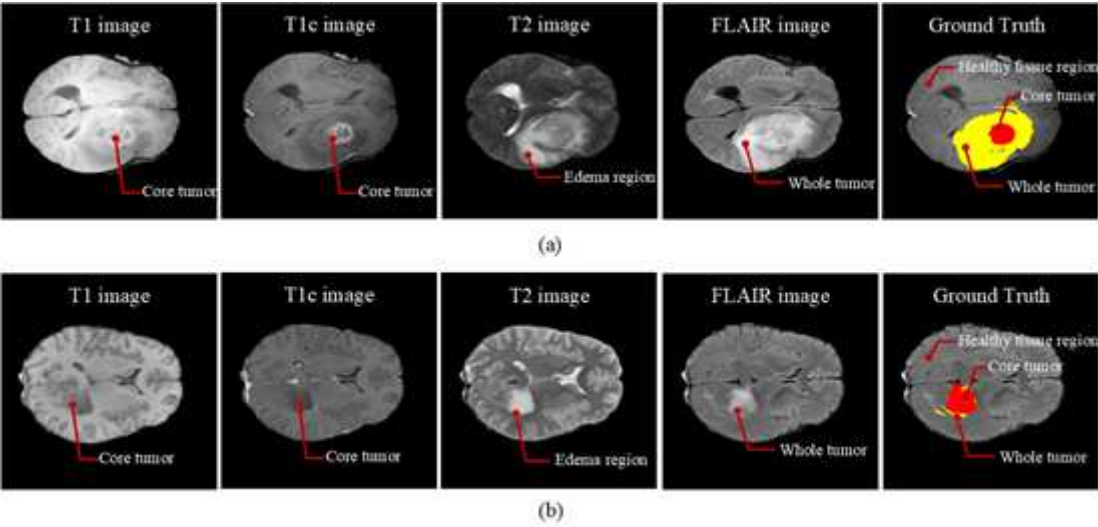
Figure 1

The multi-modal MRI medical images of brain tumors. (a) HGG; (b) LGG.
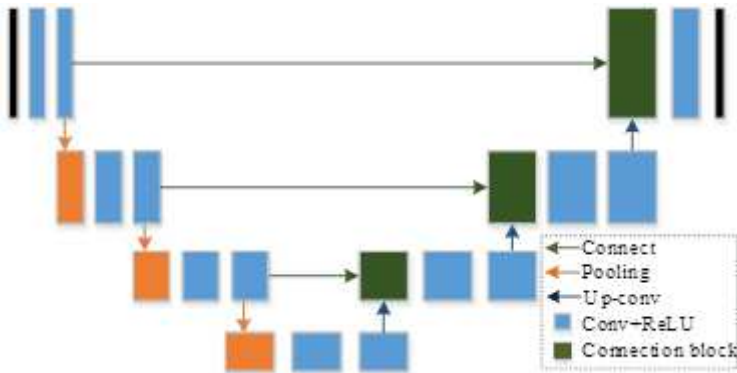
Figure 2

The structure of the traditional Unet model.
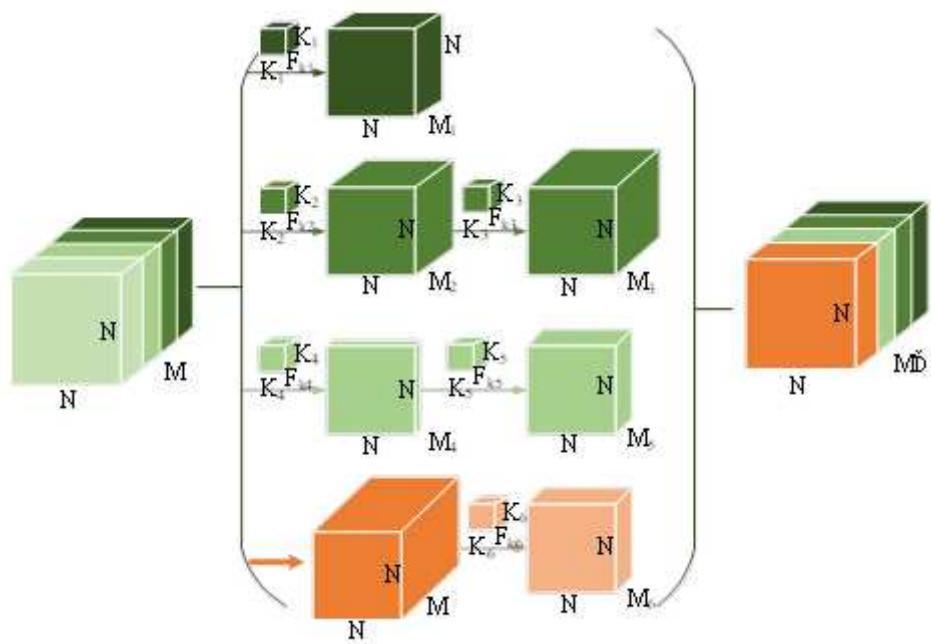
Figure 3

The integrated block.
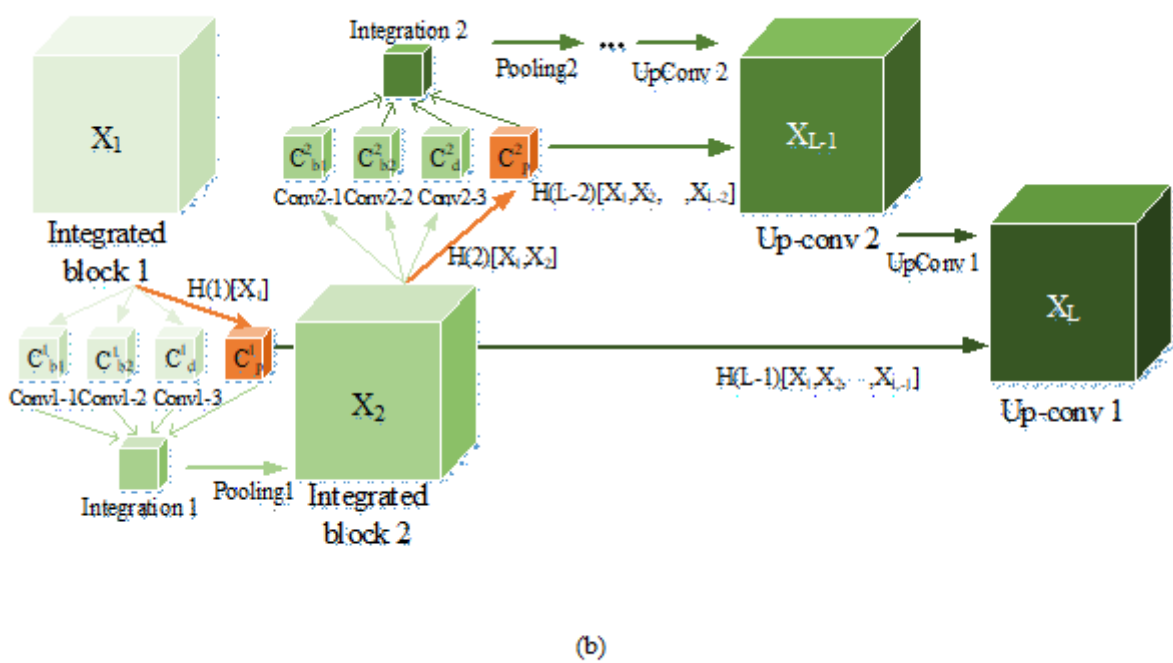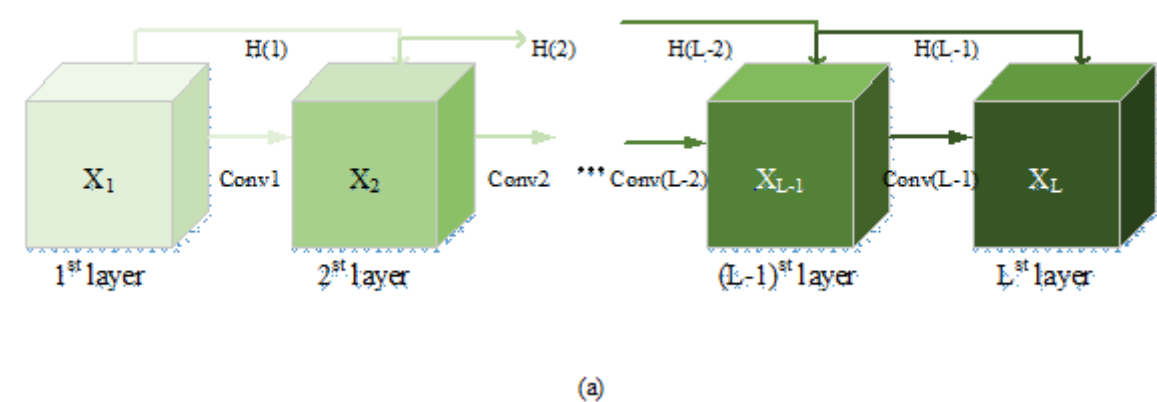
Figure 4

The aggregation connection. (a) The ResNet model; (b) The proposed model.
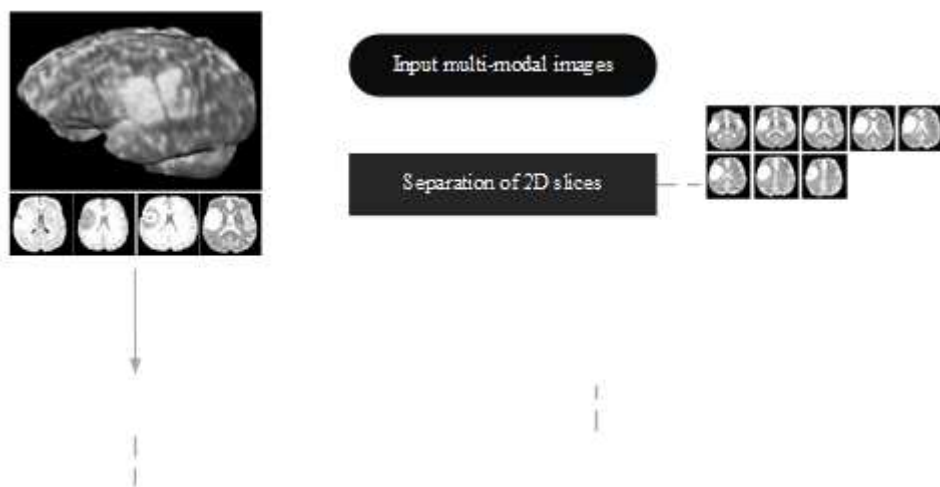
Figure 5

The flow chart of the proposed algorithm.
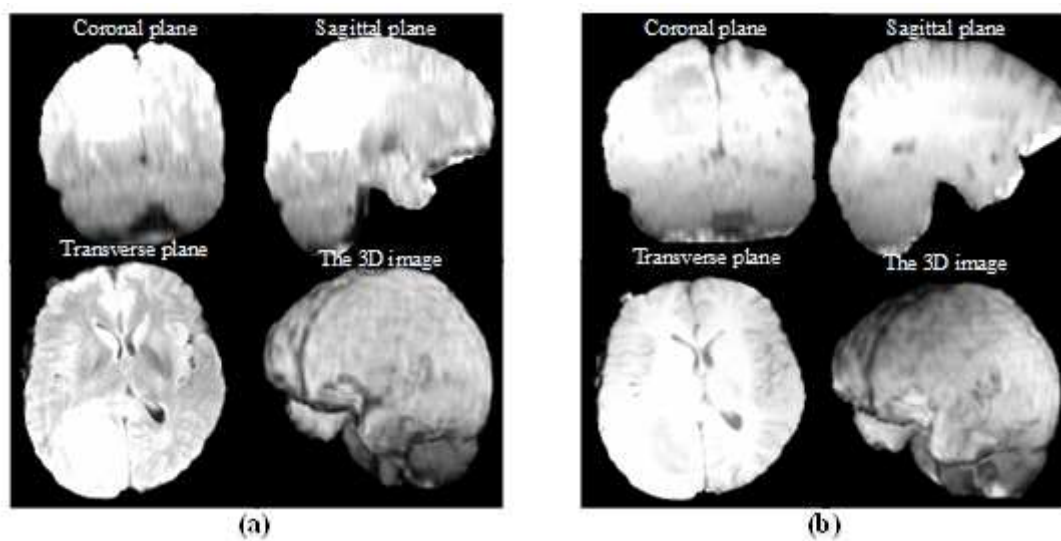
Figure 6

The MRI images of brain tumors from different angles.
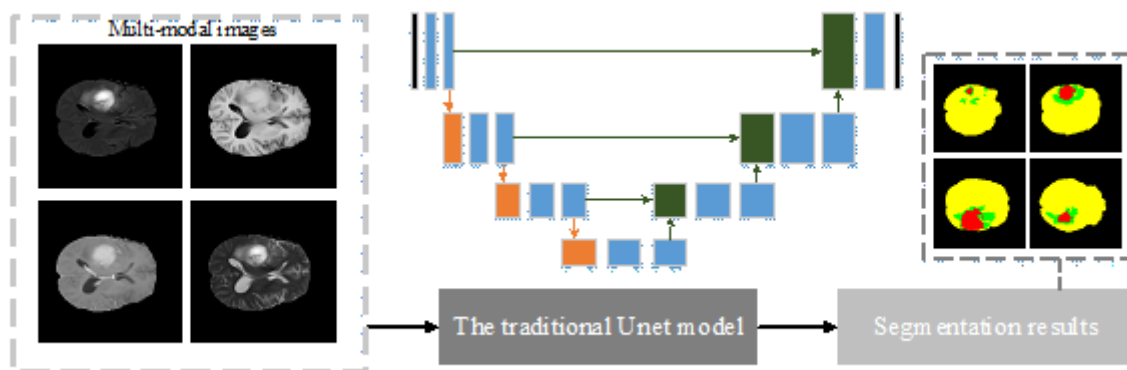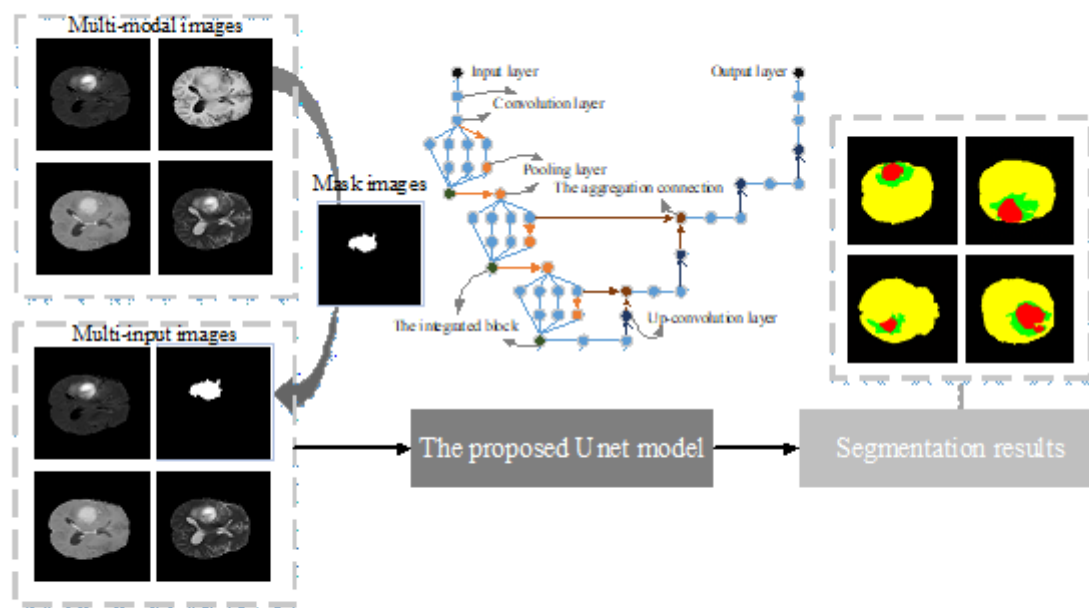
Figure 7

The process of the proposed multi-input Unet model. (a) Traditional Unet model; (b) The proposed model.
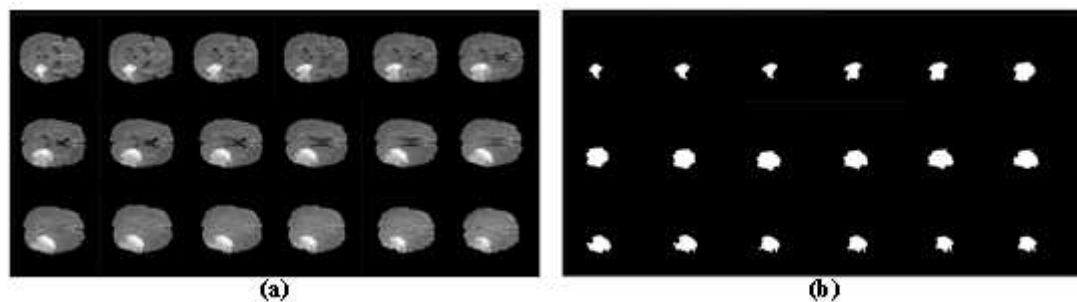
Figure 8

The mask images by region growing algorithm. (a) 2D slices; (b) The corresponding mask images.
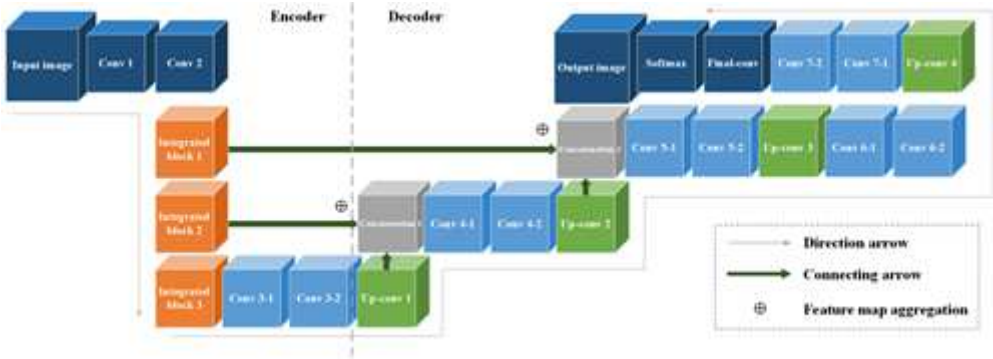
Figure 9

The encoder and decoder of the proposed model.
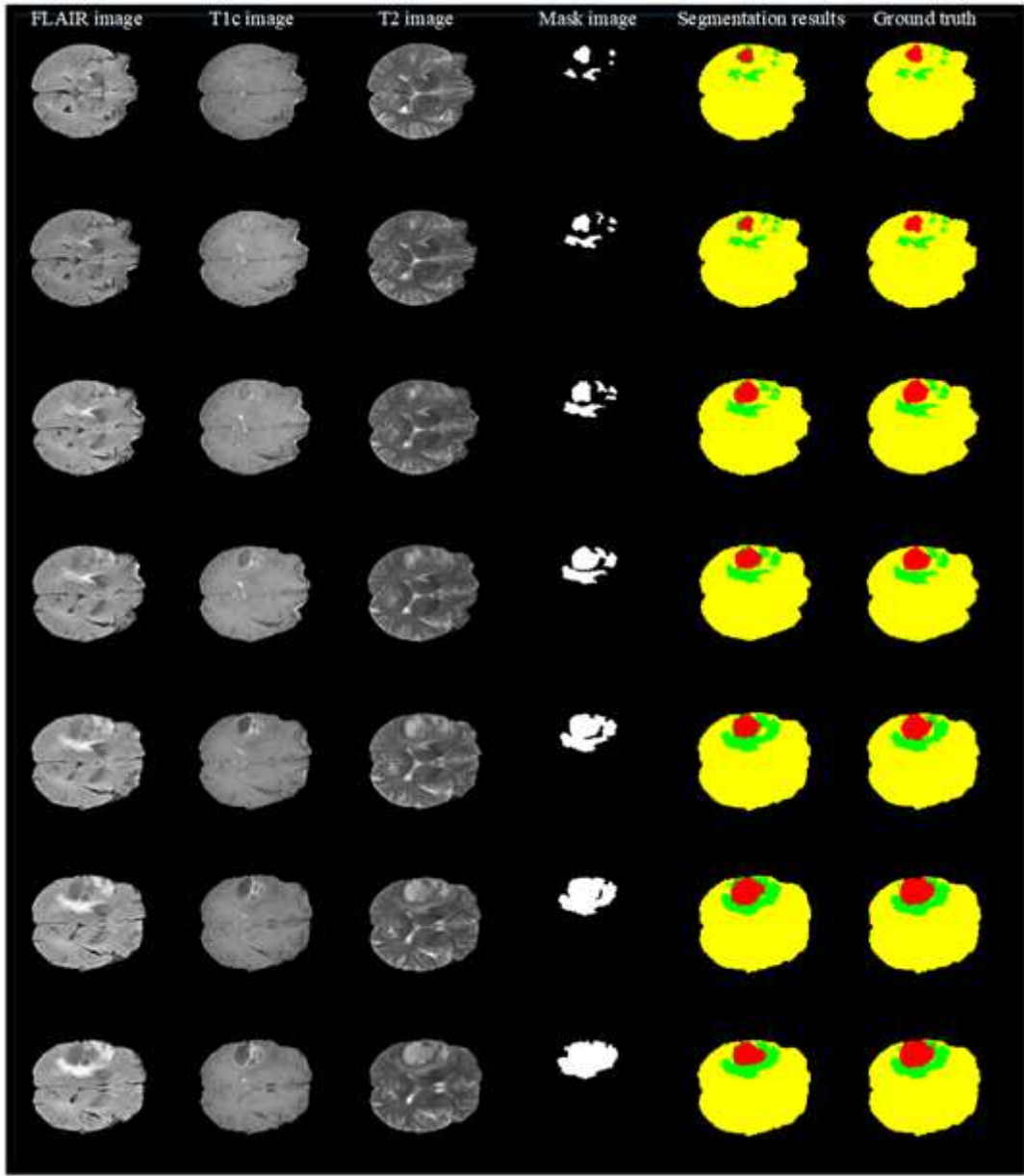
Figure 10

Segmentation results of the proposed model.
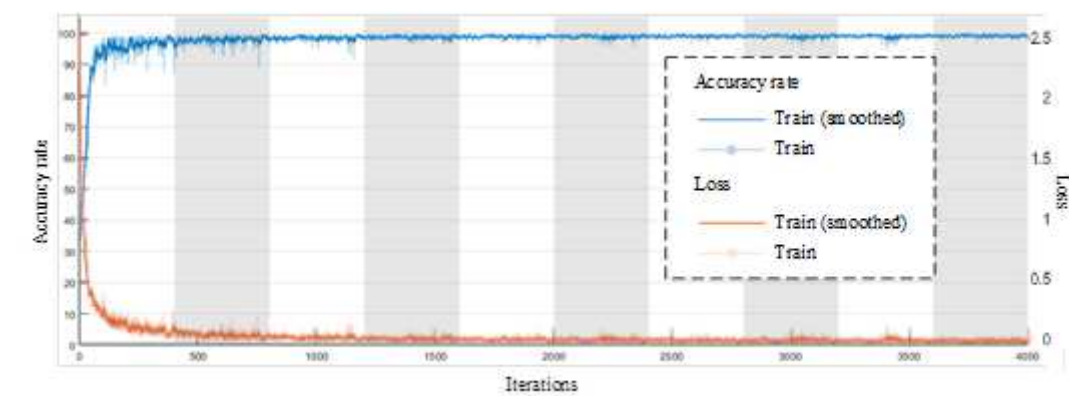
Figure 11

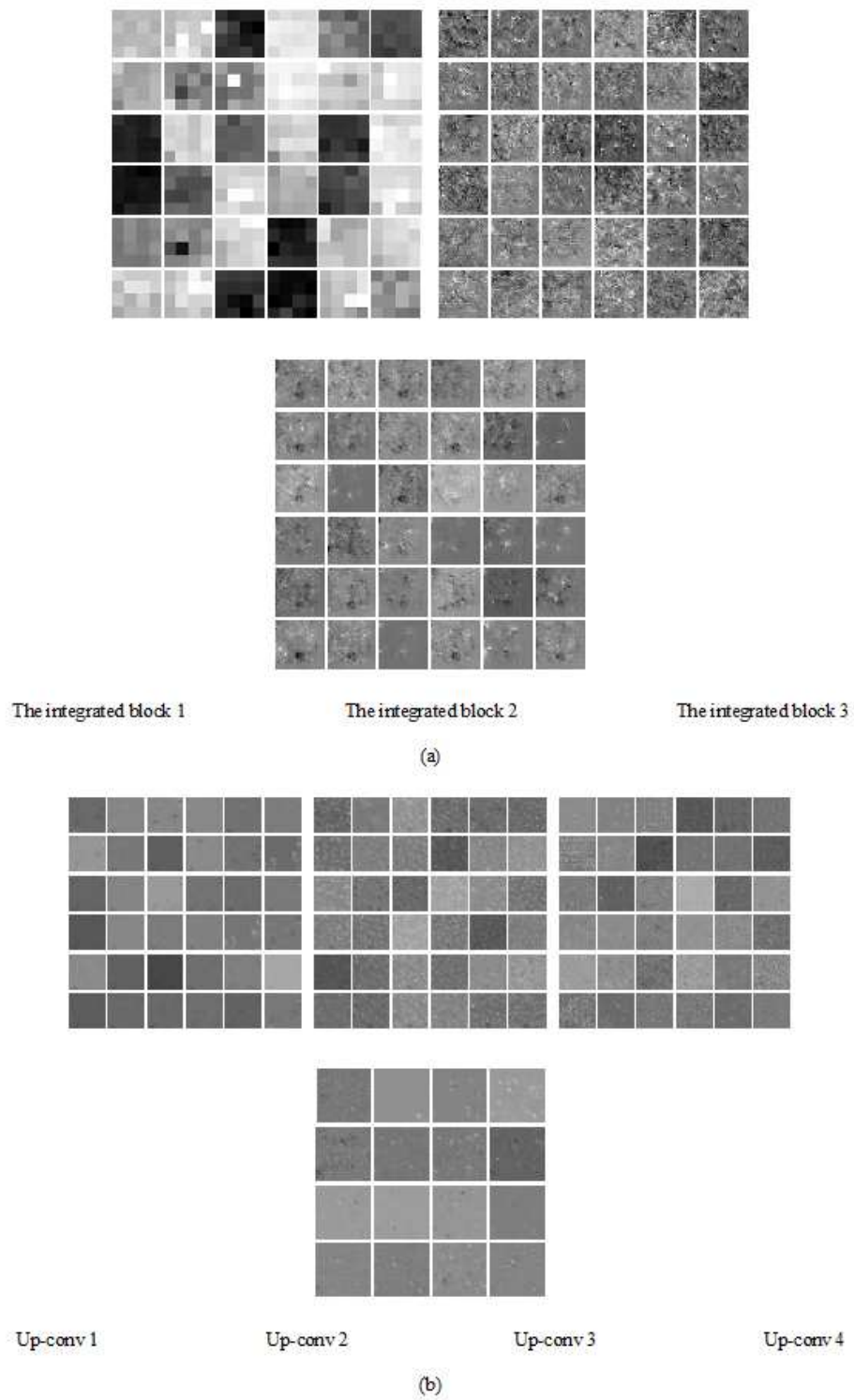The accuracy and loss rate of the proposed model.

The integrated block 1   The integrated block 2   The integrated block 3

(a)

Up-conv 1   Up-conv 2   Up-conv 3   Up-conv 4

(b)

Figure 12

The feature maps of the proposed model. (a) Part of the feature maps in the integrated blocks 1-3; (b) Part of the feature maps in the up-sampling (Up-conv 1-4).
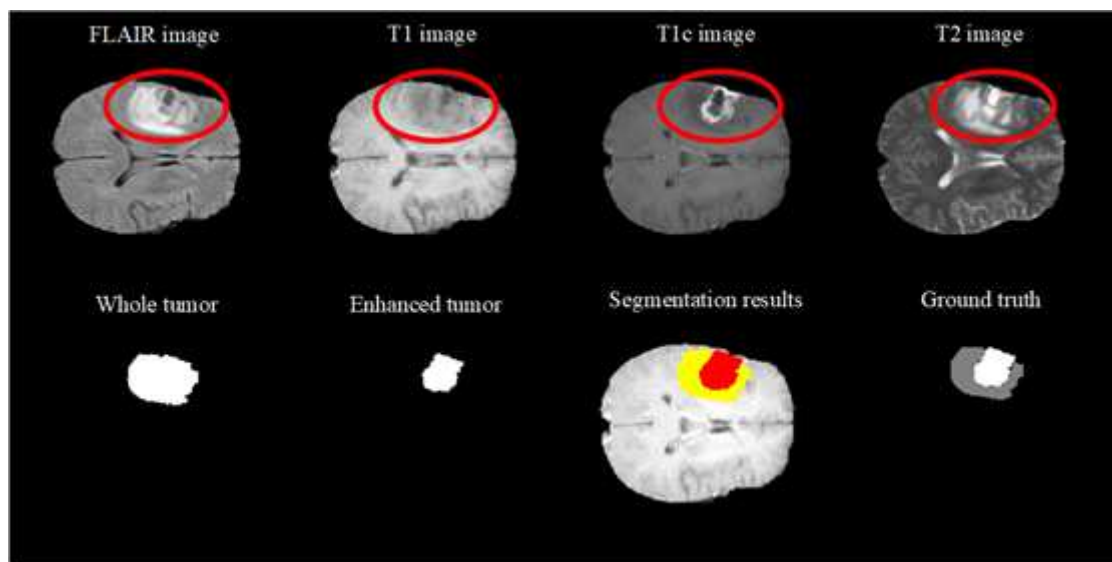
Figure 13

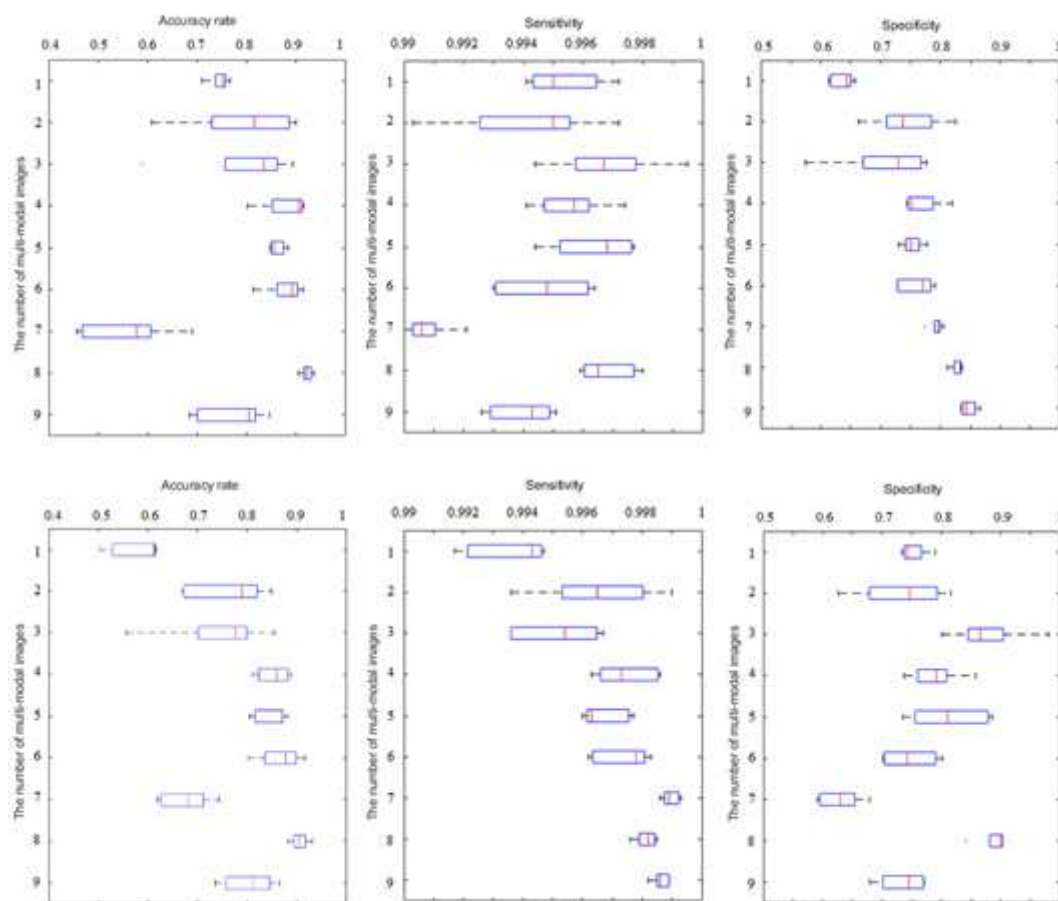Segmentation results of the multi-modal images.

Figure 14

The performance of the proposed multi-input Unet model in different multi-modal images. (The first and second rows represent the segmentation results of whole tumor and core lesion, respectively. The first to third columns represent three performance indicators: accuracy, sensitivity, and specificity, respectively. Here, numbers 1-4 represent the single-modal images :T2 images, T1 images, FLAIR images, and T1c images. Numbers 5-9 represent the multi-modal images, which are T2+T1c images, FLAIR+T1c images, T1+T1c images, FLAIR+T1c+T2 images, and FLAIR+T1c+T2+T1 images.)
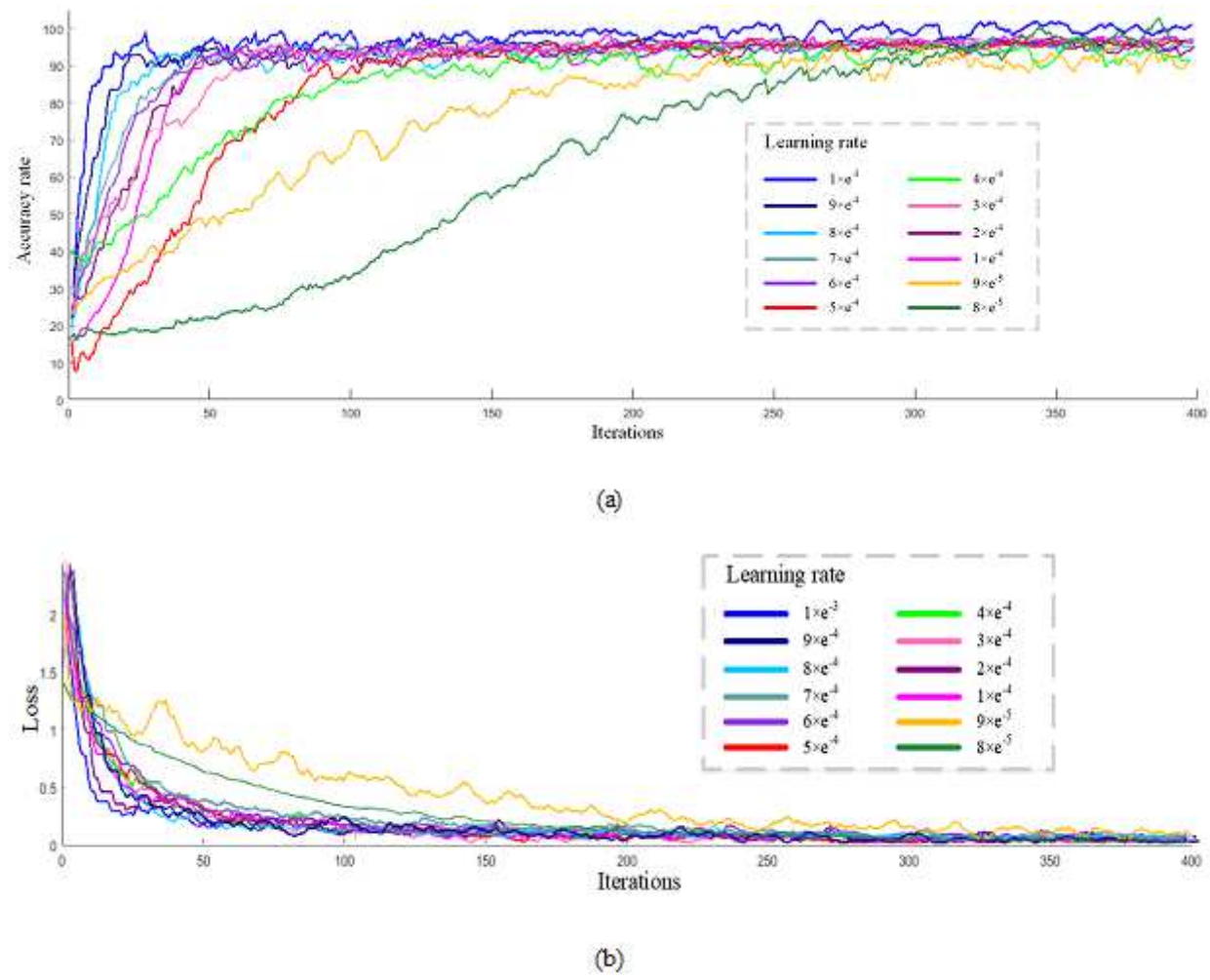

(a)

(b)

Figure 15

The accuracy and loss rate curves of the proposed model under different learning rates.

Figure 16

The performance of the proposed algorithm in different multi-input images.(where the numbers 1-9 represent T2 , T1, FLAIR, T1c, T2+T1c, FLAIR+T1c, T1+T1c, FLAIR+T1c+T2, and FLAIR+T1c+T2+T1 with the mask images, respectively. The numbers 1'-9' indicate the corresponding multi-input images without the mask images.)