Technology and Health Care 32 (2024) S183–S195
DOI 10.3233/THC-248016
IOS Press

Brain tumor segmentation based on the U-NET++ network with efficientnet encoder

Yunyi Chen[a,b,1], Lan Quan[c,d,1], Chao Long[e], Yuxuan Chen[a,b], Li Zu[a,f,*] and Chenxi Huang[a]
[a]*Key Open Project of Key Laboratory of Data Science and Intelligence Education, Hainan Normal University, Haikou, Hainan, China*
[b]*Department of Software Engineering, School of Informatics, Xiamen University, Xiamen, Fujian, China*
[c]*Department of Neurology and Department of Neuroscience, The First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian, China*
[d]*Xiamen Key Laboratory of Brain Center, Xiamen, Fujian, China*
[e]*School of Graduate Studies, La Consolacion University Philippines, Malolos, Philippines*
[f]*College of Mathematics and Statistics, Hainan Normal University, Haikou, Hainan, China*

Abstract.
BACKGROUND: Brain tumor is a highly destructive, aggressive, and fatal disease. The presence of brain tumors can disrupt the brain's ability to control body movements, consciousness, sensations, thoughts, speech, and memory. Brain tumors are often accompanied by symptoms like epilepsy, headaches, and sensory loss, leading to varying degrees of cognitive impairment in affected patients.
OBJECTIVE: The study goal is to develop an effective method to detect and segment brain tumor with high accurancy.
METHODS: This paper proposes a novel U-Net++ network using EfficientNet as the encoder to segment brain tumors based on MRI images. We adjust the original U-Net++ model by removing the dense skip connections between sub-networks to simplify computational complexity and improve model efficiency, while the connections of feature maps at the same resolution level are retained to bridge the semantic gap.
RESULTS: The proposed segmentation model is trained and tested on Kaggle's LGG brain tumor dataset, which obtains a satisfying performance with a Dice coefficient of 0.9180.
CONCLUSION: This paper conducts research on brain tumor segmentation, using the U-Net++ network with EfficientNet as an encoder to segment brain tumors based on MRI images. We adjust the original U-Net++ model to simplify calculations and maintains rich semantic spatial features at the same time. Multiple loss functions are compared in this study and their effectiveness are discussed. The experimental results shows the model achieves a high segmention result with Dice coefficient of 0.9180.

Keywords: Brain tumor segmentation, EfficientNet encoder, U-Net++ framework

1. Introduction

Brain tumors are formed by the growth of abnormal cells in the brain. Accurate and reliable segmentation of brain tumors is crucial for diagnosis, treatment planning, and monitoring of tumor progression. However, due to the widespread use of MRI equipment in brain examination, a large amount of brain

---

[1] Yunyi Chen and Lan Quan contributed equally to this paper.
[*] Corresponding author: Li Zu, Key Open Project of Key Laboratory of Data Science and Intelligence Education, College of Mathematics and Statistics, Hainan Normal University, Haikou, Hainan 571158, China. E-mail: 050112@hainnu.edu.cn.

MRI images will be generated in the clinic, which is impossible for doctors to manually annotate and segment. Furthermore, due to the intricate neural structures within the brain, segmenting brain tumors can be particularly challenging. Patients may experience severe symptoms such as epilepsy, aphasia, dementia, and cognitive impairment after the surgery, resulting in irreversible damage. Cleanly segmenting tumor tissue and avoiding damage to normal nerve tissue can significantly drop the probability of the above-mentioned surgical sequelae and reduce tumor recurrence rate, which is crucial for mitigating the suffering of patients and prolonging the survival time.

Automatic segmentation networks have the ability to analyze a large amount of medical data to detect and segment tumors, usually getting segmentation results in a short time and achieving higher accuracy than manual diagnosis. This paper aims to separate tumor tissue and surrounding normal nerve tissue accurately with the help of deep learning algorithms. We propose a novel U-Net++ brain tumor segmentation model using EfficientNet as the encoder. In this paper, the dense skip connections between the sub-networks of original U-Net++ network are removed, and the feature maps on the same resolution layer are connected to generate new feature maps. We then continue to recover details by upsampling these new feature maps. This method significantly reduces the amount of calculation and improves the efficiency of the model. At the same time, we combine the advantages of the cross-entropy loss function and the Dice loss function, making the two work together in the model training process to develop a stable model. According to the results, the proposed U-Net++ segmentation model obtains a Dice coefficient of 0.918.

2. Literature review

Some traditional machine learning algorithms such as support vector machine (SVM), conditional random field (CRF), and random forest (RF) are widely used in brain tumor segmentation tasks in the past few decades. Juan-Albarracín et al. [1] evaluated different unsupervised segmentation methods including K-means, Fuzzy K-means, GMM, GHMRF, which prove unsupervised segmentation algorithms are effective and competitive in brain tumor segmentation. Wu et al. [2] proposed a color-based K-means clustering segmentation method, which combines color transformation, K-means clustering and histogram clustering. This model provided good segmentation performance that helps doctors accurately detect the size and area of the tumor.

Later, with the innovation of hardware and other technologies, computing power is promoted. The development of deep learning is booming and shows vast advantages in the medical field. Mittal et al. [3] applied a model combining Stationary Wavelet Transform (SWT) and Growing Convolutional Neural Network (GCNN). SWT was used for feature extraction, and it has better results than Fourier transform on discontinuous data. After feature extraction, a random forest classifier was applied for segmentation and the GCNN network was used to train the model, which improved the brain tumor segmentation accuracy. Nema et al. [4] proposed a novel RescueNet network. The architecture consists of an encoding path (from 256×256 to 32×32) and a symmetric decoding path (from 64×64 to 256×256) that can segment the tumor. This model only requires a small amount of training data to produce good segmentation results. The authors designed the DownBlock and UpBlock structure, and utilized residual blocks to avoid gradient disappearance during backpropagation. The DICE coefficient of this RescueNet model on the BRATS 2015 and BRATS 2017 datasets reached 94.01% and 94.63%, respectively.

Researchers also found the encoder-decoder structure greatly improves the accuracy of image segmentation tasks. Ghosh et al. [5] used an improved U-Net framework to segment tumors, applying dense convolutional blocks during the downsampling process to enhance feature reusability. Meanwhile, VGG-16 architecture was used for the decoder. To improve model stability and performance, they added batchnormalization (BN) layers in the dense blocks. After performing a 5-fold cross-validation method on the TCGA-LGG dataset, this model got a Dice coefficient of 0.9099 and a Jaccard coefficient of 0.8189. The U-Net architecture has an encoder and decoder structure that preserves rich semantic features and provides excellent performance in image segmentation tasks.

Most current research efforts seek methods to obtain automated brain tumor segmentation results that provide objective, reproducible, and scalable methods for the quantitative assessment of brain tumors [6]. While the encoder-decoder architecture has been widely applied for brain tumor segmentation, choosing the optimal architecture remains a significant challenge. For example, many variants of U-Net exist, such as Residual U-Net [7], Dense U-Net [8], etc. Choosing an optimal U-Net architecture for more satisfactory performance and accuracy remains an important task.

3. Materials and methods

This paper proposes a U-Net++ model with EfficientNet as the encoder for tumor segmentation in MRI images. The decoder module is adjusted by removing the skip connections of the sub-networks of original U-Net++. We connect the upsampling result of the deep feature map with the corresponding feature maps of the same layer to maintain rich semantic features and spatial information. In addition, the cross-entropy and Dice loss function are combined, working together for model optimization. Our EfficientNet encoder uses ImageNet initialization weight to improve training speed and avoid model overfitting caused by training the network from scratch.

We apply image augmentation techniques, and the training data after augmentation is fed into the EfficientNet encoder for feature extraction, and then into the U-Net++ decoder module for training. After that, the test data is input into the fully trained model to obtain segmentation results. Then, we visualize the segmentation results and evaluate them with the Dice coefficient. The process can be seen in Fig. 1.

3.1. Data collection

The dataset selected in this paper is the LGG Segmentation Dataset available on Kaggle, which is a reliable and widely used dataset in the brain tumor segmentation field. The dataset includes images from The Cancer Imaging Archive (TCIA) and data from 110 patients in the lower-grade glioma collection of The Cancer Genome Atlas (TCGA). Overall, the dataset contains 3929 sets of brain MRI images and corresponding manual segmentation masks. Specifically, it includes 2556 sets of normal morphological data without tumors and 1373 sets of abnormal MRI images with tumors. Figure 2 provides an illustrative example of the images present in the database, where (a) represents the brain MRI image and (b) depicts the corresponding manual segmentation mask.

3.2. Image augmentations

Deep learning models rely heavily on a large amount of training data to achieve high accuracy. Insufficient data often cause overfitting problems and influences the performance of the model. However, in most computer vision tasks, the training data is very limited. Data augmentation techniques are helpful in this situation. Image augmentation makes a series of random changes like rotation, shifts, and flips to

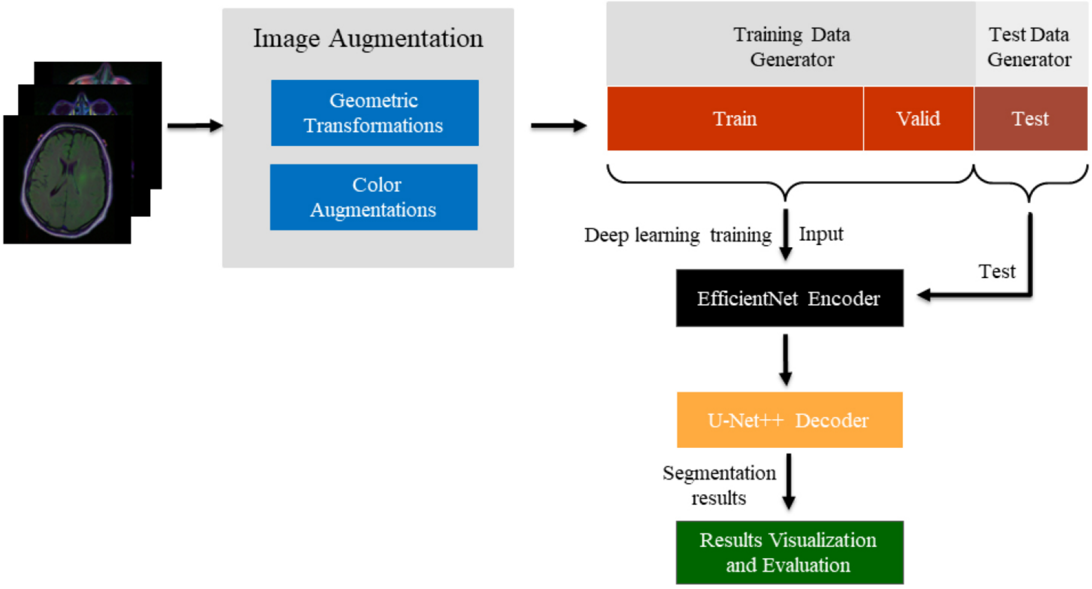

Fig. 1. The flowchart of proposed segmentation model.

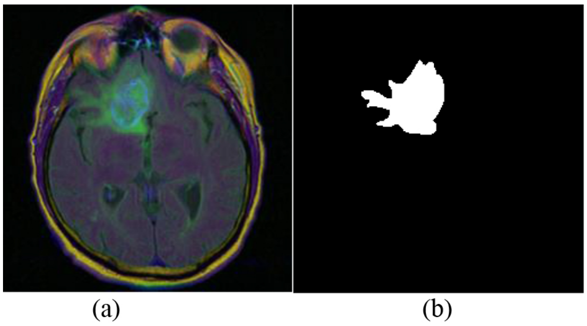

Fig. 2. Details of dataset. (a) Brain MRI image. (b) Manual Segmentation mask.

develop new training examples from the existing ones, efficiently reducing the model's dependence on certain attributes.

In this research, we apply the Albumentations library for image augmentation. Albumentations [9] is a powerful data augmentation tool that contains lots of data augmentation methods such as image resizing, scale, rotation, transpose, flipping, etc. Meanwhile, the Albumentations package can simultaneously change the original images and corresponding mask images and provide high-speed processing speed, which is suitable and popular in segmentation tasks. In this paper, we set up the following augmentation methods using the Albumentations package:

(1) Set random HorizontalFlip by the ratio of 0.5;
(2) Set contrast limited adaptive histogram equalization (CLAHE) by a ratio of 0.5;
(3) Randomly select a transform in RandomGamma transform and RandomBrightness transform for the MRI images by the ratio of 0.3;
(4) Randomly select a transform in ElasticTransform and GridDistortion for the MRI images by the ratio of 0.3;

3.3. EfficientNet encoder

EfficientNet [10] is a novel convolutional neural network architecture proposed by the Google team. It provides a family of models (EfficientNetB0 to EfficientNetB7) which can extract features efficiently and is considered the most state-of-the-art computer vision network nowadays. A convolutional neural network has three crucial dimensions: depth, width, and resolution. The depth of the network corresponds to the number of layers of the network, width corresponds to the number of neurons in the layer, and resolution is related to the height and width of the image. Usually, these three dimensions can be scaled to improve the learning effect of the model. Although choosing one among the three dimensions to scale can work out to improve the performance, when the dimension is scaled to a certain extent, the model's accuracy will fall into a bottleneck and tend to be saturated. If we deepen the network depth, the model can learn more complex features but encounters the bottleneck of gradient disappearance. In this case, the network's width and resolution should be changed to match the depth; Similarly, if the resolution is scaled, the network's depth and width should also be increased to have a larger receptive field and capture more features. Through the cooperation of the three dimensions, the model can be scaled uniformly to achieve the best model effect and slow down the saturation speed of single-dimensional scaling. Compared with other models that only scale in one dimension, EfficientNet balances the three dimensions of depth, width, and resolution to avoid model saturation caused by scaling a single dimension.

To obtain the best combination of the three variables of width, depth, and resolution more efficiently, EfficientNet uses the following compound scaling method:

$$depth : d = \alpha^{\theta} \tag{1}$$

$$width : w = \beta^{\theta} \tag{2}$$

$$resolution : r = \gamma^{\theta} \tag{3}$$

$$s.t. \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 (\alpha \geqslant 1, \beta \geqslant 1, \gamma \geqslant 1)$$

Where the value of α, β, and γ are determined by a small grid search on the original small model. θ determines the size of the available computing resources of the system, usually a user-defined parameter.

The EfficientNet family of models provides a set of scalable and efficient deep neural network architectures that can be customized to fit a wide range of computational and accuracy requirements. EfficientNet-B0 can be seen as the baseline network. EfficientNetB1-B7 are progressively larger and more complex. The best parameters of EfficientNet-B0 can be achieved by doing a small grid search over α, β, γ. Then, the above method fixes α, β, γ as constants, and the EfficientNetB1$\sim$EfficientNetB7 networks can be obtained by scaling the baseline model of EfficientNetB0 by changing θ. The core block of the EfficientNet models is the mobile inverted bottleneck (MBConv), where the squeeze and excitation (SE) module is added to give weights to each channel, achieving compression and excitation operations. EfficientNetB1$\sim$EfficientNetB7 corresponds to different numbers of MBConv blocks respectively. For example, EfficientNetB4 consists of 7 blocks, reducing the image size from the input size of 256 × 256 to 7 × 7. SE module adds parameters to each channel of the convolutional block so that the network can adaptively adjust the weights of each feature map. It first compresses the feature of each channel into a single value through an average pooling operation to obtain the global features. Then, the SE module performs the excitation operation according to the activation ratio R, uses 1x1 convolution kernels to learn the relationship between each channel, and obtains each channel's weight through the Sigmoid activation function. Then, these weighted feature maps can be fed into the next layer of the network. The SE module makes the model concentrate on the most informative channel while suppressing those unimportant features.

At the same time, the EfficientNet model uses the Swish activation function. It is a product of linear and Sigmoid functions. The value of Swish activation function can be slightly negative around zero, resulting in a smoother gradient space and making the network easier to learn the data distribution. Compared with ReLU, the Swish activation function is more suitable for deeper networks. Equations (4) and (5) state the Swish activation function method:

$$Swish(x) = x * Sigmoid(\beta x) \tag{4}$$

$$Sigmoid(y) = \frac{1}{1 + e^{-y}} \tag{5}$$

In addition, EfficientNet can represent information with fewer channels. Hence, as the computational complexity of the decoder grows, EfficientNet can build a more computationally efficient system. When using a complex decoder like U-Net++, which needs to perform a lot of operations on the extracted features, EfficientNet has a more significant advantage over other encoders [11]. Furthermore, to avoid the difficulty of finding suitable regularization parameters when training from scratch, we initialize the EfficientNet network with ImageNet weights to train the model faster and more accurately. ImageNet pre-trained weights are able to capture a wide range of visual features and are often used as a starting point for training deep neural networks on computer vision tasks. These pre-trained weights can speed up training and improve the performance of deep learning models.

3.4. U-Net++ network

Traditional convolutional neural networks obtain deep features of the original images by reducing the resolution of images. However, such low-resolution feature maps usually lose useful spatial information that affects the segmentation accuracy. Models with an encoder-decoder structure, such as the U-Net network [12], are helpful in this condition. U-Net consists of a contracting path and an expanding path, which can restore spatial information and realize accurate segmentation. The network directly combines low-resolution but semantically-rich deep feature maps with high-resolution, fine-grained shallow feature maps by skip connections to recover details in target regions. However, directly combining the shallow feature map with the upsampling deep feature map obtained in the expansion path may have a large semantic gap, affecting the model's effectiveness [13].

Compared with the U-Net network, U-Net++ [14] is a more powerful medical image segmentation model, which retains the symmetrical U-shaped down-sampling and up-sampling structure. U-Net++ consists of sub-U-Net networks of different depths. It redesigns the skip connections of the U-Net network to compose a more powerful network. In this way, U-Net++ forms multiple upsampling paths from different depths, integrates U-Net with various receptive fields and contains richer semantic information, which can obtain a better performance than traditional U-Net network. Therefore, this paper adopts the U-Net++ structure for brain tumor segmentation.

U-Net++ redesigns the skip connections path, making it densely connect the sub-networks. The dense skip connection fuses the output of the previous convolutional layer with the corresponding upsampling output of the lower dense block. Through this, the previous feature map is accumulated to the current node. This process can be expressed as follows:

$$x^{i,j} = \begin{cases} H(x^{i-1,j}), & j = 0 \\ H([[x^{i,j}]_{k=0}^{j-1}, u(x^{i+1,j-1})]), & j > 0 \end{cases} \tag{6}$$

Where $H()$ is the convolution operation, $u()$ represents the upsampling layer, and $[\,]$ is the connection layer.

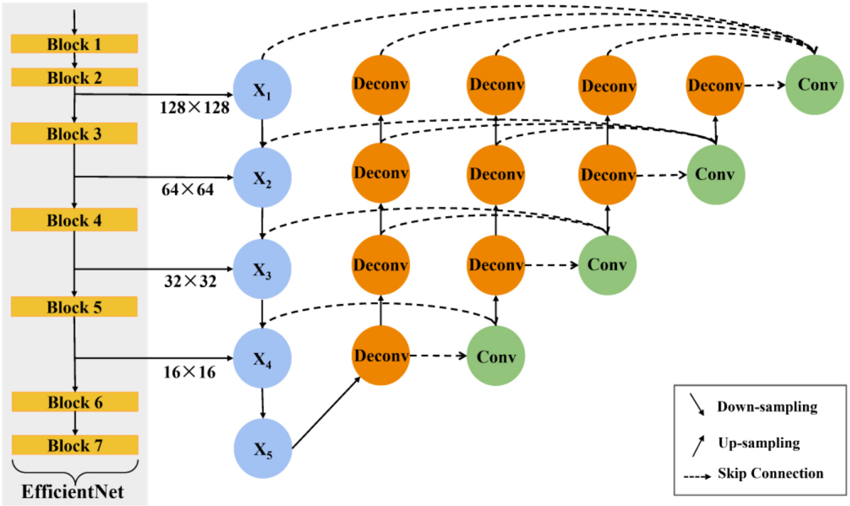

Fig. 3. The structure of the proposed U-Net++ segmentation model with EfficientNet encoder.

We extracted four feature maps of different sizes from the EfficientNet network as the input of the U-Net++. At the same time, to reduce the amount of computation, we avoid the dense connections between sub-networks of the original U-Net++. Instead, we connect the upsampling results of the same layer to form a new feature block, and then continue to upsampling the new feature block to recover details. Figure 3 details the proposed U-Net++ model with the EfficientNet Encoder.

Compared with the original U-Net++ network, this approach is cost-friendly. The original U-Net++ network introduces a nested architecture that includes multiple sub-networks. They are densely connected, and each layer of the encoding path is connected to all layers of the corresponding decoding path. While this connectivity pattern enables the network to capture high-level semantic features and achieve high performance, it also leads to high computational costs. As shown in Fig. 3, our proposed network addresses this issue by removing the dense connections between sub-networks and instead connecting only the upsampling results of the same layer. By retaining and connecting the feature maps at the same resolution level, the network can capture deep semantic features and bridge the semantic gap. This method significantly reduces the number of parameters and balances the computational cost.

3.5. *Loss function*

Our segmentation model combines the cross-entropy loss function and the Dice loss function to take advantage of them. The model can be optimized under the control of both loss functions simultaneously. The calculation process of the cross-entropy loss function can be seen in Eq. (7).

$$L == -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \tag{7}$$

In Eq. (7), N is the number of classes, y represents labels, and $p(y)$ is the probability of belonging to the y label.

It predicts the category of each pixel and averages all the pixels. Therefore, when the pixel categories on the picture are unbalanced, mainstream pixels can easily dominate the training process, reducing the capability of extracting the weak pixel categories. The dice coefficient is a metric function used to measure the similarity of sets. It is usually used to calculate the pixel similarity between two samples, suitable for semantic segmentation tasks. The Dice loss function formula is shown below:

$$\textit{Dice Loss} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} = 1 - \frac{2TP}{2TP + FN + FP} \tag{8}$$

Where $|X \cap Y|$ is the intersection between X and Y, and $|X|$ and $|Y|$ represents the number of samples of X and Y. For segmentation tasks, X is the ground truth, and Y is the predicted value. And TP is the true positive, FP is the false positive, TN is the true negative, FN is the false negative.

In this paper, we combine the binary cross-entropy loss function and the dice loss function. The sum of the two loss functions is used as a measure of model loss. The Dice loss function calculates the proportion of overlap between the predicted segmentation results and the ground truth, which reduces the impact of imbalanced data on model optimization and alleviates the problem caused by the imbalance between foreground and background [15]. However, it may make the training process unstable. When the target area is small, the prediction error of some pixels will cause a huge fluctuation in the loss value. Both the cross-entropy loss function and the Dice loss function have their shortcomings, so we combine the cross-entropy with the Dice loss function to help improve the stability of the model.

4. Results and discussion

In our experiment, the Adam optimizer was selected, which is a stochastic gradient-based optimization. The Adam optimizer is computationally efficient, requires little memory, and only requires first-order gradients [16]. Besides, ImageNet pre-trained weights are used to initialize the EfficientNet model, which provides a reasonable initial weight of the encoder and avoids overfitting problems. We divide the dataset into training set, validation set, and test set at a ratio of 8:1:1. There are 100 epochs in the training process, and the learning rate is dynamically adjusted during the process with an initial learning rate of 0.001. We utilize the cosine annealing [17] to adjust the learning rate, which can prevent the training from falling into a local optimal value.

We adopt the Dice coefficient to evaluate the segmentation results in this experiment, which can calculate the similarity of two samples and is effective in segmentation tasks. Equation (9) shows the computing process of Dice coefficient. To ensure the reliability of the results, we experiment 10 times and take the average value as the final result.

$$\textit{Dice Coefficient} = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FN + FP} \tag{9}$$

In Eq. (9), X is the ground truth, and Y is the predicted value. $|X \cap Y|$ is the intersection between X and Y, $|X|$ and $|Y|$ represents the number of samples of X and Y. TP is the true positive, FP is the false positive, TN is the true negative, FN is the false negative.

In our experiment, the EfficientNetB4 encoder is applied. The encoder is initialized by ImageNet weight to avoid model overfitting. The encoder output layers with shapes of 16 × 16, 32 × 32, 64 × 64, and 128 × 128 are inputted to the proposed U-Net++ network. The combination of binary cross-entropy loss and dice loss is used to measure model loss. It shows that our network achieves a dice coefficient of 0.9180, proving the proposed segmentation model has good performance on the brain tumor segmentation task. Figure 4 visualizes the predicted segmentation result. To facilitate a more comprehensive comparison of the results, Fig. 5 depicts the ground truth brain tumor mask. By comparing Figs 4 and 5, it is evident that our segmentation model is capable of accurately identifying the location and shape of brain tumors.

Table 1
Results of different EfficientNet encoder

Model	Dice coefficient
EfficientNetB0	0.9177
EfficientNetB4	0.9180

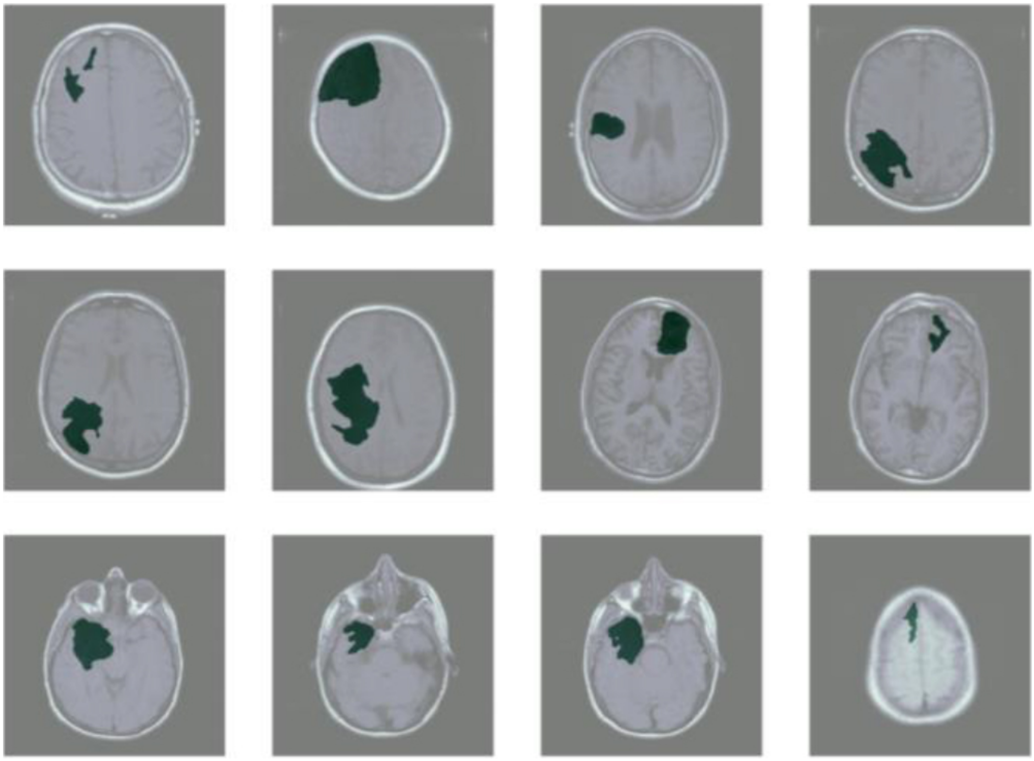

Fig. 4. Segmentation model results.

To compare the effect of different EfficientNet encoders, we apply EfficientNetB0 and EfficientNetB4 from the EfficientNet family as the encoder, respectively. Both EfficientNetB0 and EfficientNetB4 are initialized with ImageNet pre-trained weights. The Dice coefficient results are demonstrated in Table 1.

Table 1 shows the segmentation model with EfficientNetB0 as the encoder obtains a Dice coefficient of 0.9177. The EfficientNetB4 encoder has more channels and model parameters, and therefore has the ability to learn more feature information, performing slightly better than EfficientNetB0.

Additionally, to compare the impact of the loss function on the proposed segmentation model, we conduct experiments using four different loss functions. All the loss functions are experimented with the EfficientNetB4 encoder to control variables. We compare the Dice loss function, cross-entropy loss function, Focal loss function [18], and loss function combining cross-entropy and Dice loss. The focal loss function is an extension of the cross-entropy loss function. Equation (10) shows the calculation process of it. Compared with the Dice loss function, the Focal loss function focuses on balancing hard and easy samples. It allows the model to focus on hard, misclassified examples by assigning weights, avoiding easy-to-classify data dominating the overall loss value.

$$\textit{Focal Loss} = -\alpha_t(1-p_t)^{\gamma} * \log(p_t) \tag{10}$$

Table 2
Results of different loss functions

Loss function	Dice coefficient
Dice Loss	0.9163
Cross-entropy Loss	0.9130
Focal Loss	0.9088
Cross-entropy+Dice Loss	0.9180

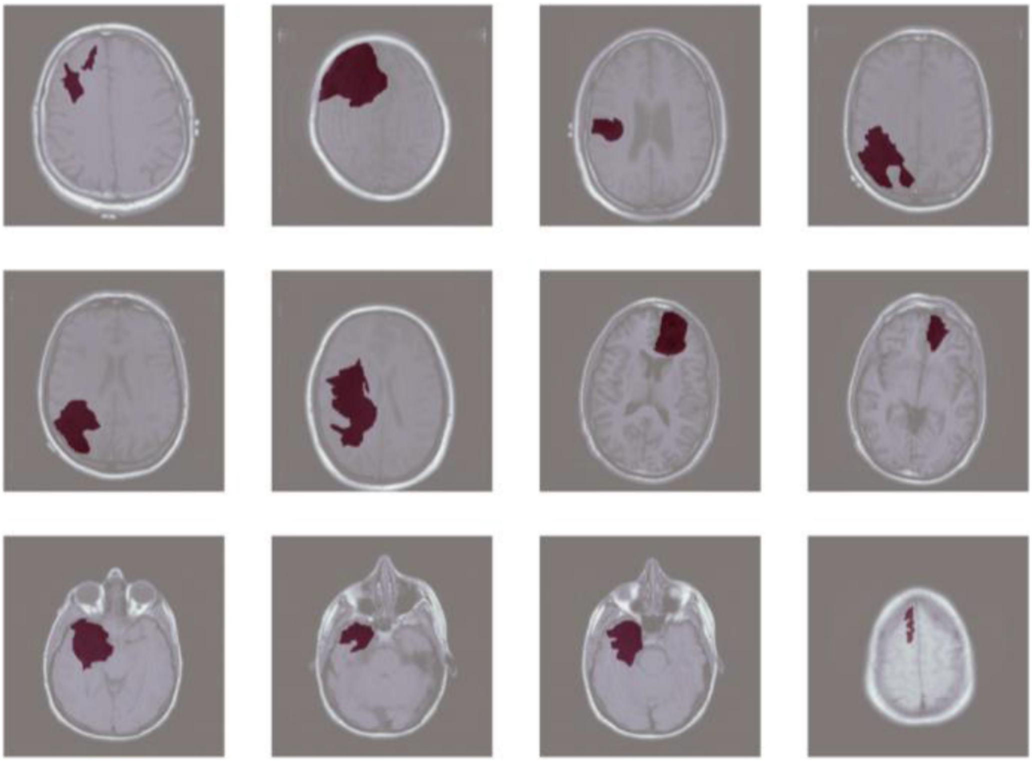

Fig. 5. Ground truth brain tumor mask.

Where α_t is a weighted term, γ is an adjustable positive parameter, and p_t is the predicted results.

Table 2 details specific experimental results of different loss functions. Results in Table 2 show that the loss function combining the cross-entropy and Dice loss has the best performance. The model's optimization effect using the cross-entropy loss function or the Dice loss function alone is not as good as the combined effect of the two on the model. The focal loss function is expected to have good performance since it extends the cross-entropy loss function and avoids the domination of easy-to-classify samples. However, results show that the Focal loss function performs the worst on this dataset among the four loss functions. The reason may be attributed to the large number of learnable samples we have provided through data augmentation technology. This approach increased the diversity of the dataset and provided the model with more examples to learn from. As a result, the advantage of the Focal loss function in addressing class imbalance was diminished.

In addition, to verify the effectiveness of the network, this paper compares the performance of this U-Net++ model with other segmentation models proposed by other researchers. Tamer et al. [19] test the results of brain tumor segmentation on the LGG dataset using the original U-Net model. Finally,

Table 3
Comparison of other segmentation models

Authors	Models	Dice coefficient
Tamer A[19]	U-Net	89.02%
Sourodip Ghosh[5]	U-Net+VGG-16	90.99%
Liping Yi[20]	SU-Net	78.5%
Rai H M[21]	Vanilla U-Net	82.06%
\	**Proposed model**	**91.80%**

they obtained a Dice coefficient of 89.02%, indicating that the U-Net++ model proposed in this paper performs better than the original U-Net model. Ghosh et al. [5] apply an improved U-Net framework to segment tumors, where VGG-16 architecture is used for the decoder. Finally, this improved model achieves a Dice coefficient of 90.99%. Yi et al. [20], and Rai et al. [21] proposed other models for segmentation. The detailed comparison results can be clearly seen in Table 3, which proves that the U-Net++ segmentation model with the EfficientNet encoder proposed in this paper reaches satisfying performance, surpassing other models in segmentation accuracy.

5. Conclusions

Precise segmentation of brain tumors provides important support for neuropathological analysis, diagnostic report generation, surgical plan design, and treatment plan formulation. The complex anatomy of brain tumors makes manually segmentation time-consuming and easily affected by subjective factors. Therefore, establishing efficient algorithms to segment brain tumors automatically have great clinical significance.

In recent years, the development of medical imaging technology and deep learning algorithms has promoted the progress of related work. This paper conducts research on brain tumor segmentation, using the U-Net++ network with EfficientNet as an encoder to segment brain tumors based on MRI images. The data is augmented using the Albumentations package. EfficientNet is a powerful feature extraction network, especially beneficial when the decoder needs to perform lots of complex calculations. We adjust the original U-Net++ model by deleting the dense skip connections between sub-networks and connecting the upsampling results with the feature maps of the same layer, which simplifies calculations and maintains rich semantic features and spatial information at the same time. Besides, we combine the cross-entropy and Dice loss functions to make them work together for model optimization. In the experiments, four different loss functions are tested to compare the impact of the loss function. Finally, the proposed model is compared with other models such as U-Net and SU-Net to verify the model's validity. The experimental results show that a high-accuracy segmentation model for brain tumors based on MRI images is realized. Our model achieves a satisfactory segmentation performance, obtaining a Dice coefficient of 0.9180.

The proposed U-Net++ network focuses on brain tumor segmentation with 2D MRI images. However, 2D slices may not capture the full spatial information of the brain, leading to potential inaccuracies in the segmentation process. Compared with 2D slices, 3D MRI images provide comprehensive information by capturing the brain's volume with various flow directions. Therefore, future work is expected to extend our segmentation model to 3D space, which can potentially improve the accuracy and robustness of the model. Moreover, we will continue to search methods to expand the dataset and try to adjust our model by adding dropout layers or BatchNormalization layers to prevent overfitting and improve the stability and speed during the training process.

Acknowledgments

This study was supported by the (1) Key Open Project of Key Laboratory of Data Science and Intelligence Education (Hainan Normal University), Ministry of Education (DSIE202303); (2) Open Fund of Key Laboratory of Embedded System and Service Computing (Tongji University) under Grant ESSCKF 2022–03; (3) the National Natural Science Foundation of PR China (No. 12261029); and (4) the Higher Education Project of Hainan Provincial Department of Education (No. HnjgY2022ZD-4).

Conflict of interest

None to report.

References

[1] Juan-Albarracín J, Fuster-Garcia E, Manjon JV, et al. Automated glioblastoma segmentation based on a multiparametric structured unsupervised classification. PloS One. 2015; 10(5): e0125143.

[2] Wu MN, Lin CC, Chang CC. Brain tumor detection using color-based k-means clustering segmentation. Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007). IEEE. Vol. 2. 2007. pp. 245-250.

[3] Mittal M, Goyal LM, Kaur S, et al. Deep learning based enhanced tumor segmentation approach for MR brain images. Applied Soft Computing. 2019; 78: 346-354.

[4] Nema S, Dudhane A, Murala S, et al. RescueNet: An unpaired GAN for brain tumor segmentation. Biomedical Signal Processing and Control. 2020; 55: 101641.

[5] Ghosh S, Chaki A, Santosh KC. Improved U-Net architecture with VGG-16 for brain tumor segmentation. Physical and Engineering Sciences in Medicine. 2021; 44(3): 703-712.

[6] Magadza T, Viriri S. Deep learning for brain tumor segmentation: A survey of state-of-the-art. Journal of Imaging. 2021; 7(2): 19.

[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. pp. 770-778.

[8] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. pp. 4700-4708.

[9] Buslaev A, Iglovikov VI, Khvedchenya E, et al. Albumentations: fast and flexible image augmentations. Information. 2020; 11(2): 125.

[10] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning. PMLR. 2019. pp. 6105-6114.

[11] Silva JL, Menezes MN, Rodrigues T, et al. Encoder-decoder architectures for clinically relevant coronary artery segmentation. arXiv preprint arXiv2106.11447. 2021.

[12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham. 2015. pp. 234-241.

[13] Le Duy Huynh NB. A U-NET++ with pre-trained EfficientNet backbone for segmentation of diseases and artifacts in endoscopy images and videos. CEUR Workshop Proceedings. 2020; 2595: 13-17.

[14] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation. Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham. 2018; 3-11.

[15] Zhao R, Qian B, Zhang X, et al. Rethinking dice loss for medical image segmentation. 2020 IEEE International Conference on Data Mining (ICDM). IEEE. 2020. pp. 851-860.

[16] Zhang D, Chen Y, Chen Y, et al. Heart disease prediction based on the embedded feature selection method and deep neural network. Journal of Healthcare Engineering. 2021. 2021.

[17] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv1608.03983. 2016.

[18] Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision. 2017. pp. 2980-2988.

[19] Tamer A, Youssef A, Ibrahim M, et al. Deep learning techniques for the fully automated detection and segmentation of brain MRI. 2022 5th International Conference on Computing and Informatics (ICCI). IEEE. 2022. pp. 310-315.

[20] Yi L, Zhang J, Zhang R, et al. SU-net: an efficient encoder-decoder model of federated learning for brain tumor segmentation. International Conference on Artificial Neural Networks. Springer, Cham. 2020. pp. 761-773.
[21] Rai HM, Chatterjee K, Dashkevich S. Automatic and accurate abnormality detection from brain MR images using a novel hybrid UnetResNext-50 deep CNN model. Biomedical Signal Processing and Control. 2021; 66: 102477.