





Segmentation of the Multimodal Brain Tumor Images Used Res-U-Net

Jindong Sun¹ , Yanjun Peng^{1,2} , Dapeng Li¹, and Yanfei Guo¹

¹ College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China

pengyanjuncn@163.com

² Shandong Province Key Laboratory of Wisdom Mining Information Technology, Qingdao, China

Abstract. Gliomas are the most common brain tumors, which have a high mortality. Magnetic resonance imaging (MRI) is useful to assess gliomas, in which segmentation of multimodal brain tissues in 3D medical images is of great significance for brain diagnosis. Due to manual job for segmentation is time-consuming, an automated and accurate segmentation method is required. How to segment multimodal brain accurately is still a challenging task. To address this problem, we employ residual neural blocks and a U-Net architecture to build a novel network. We have evaluated the performances of different primary residual neural blocks in building U-Net. Our proposed method was evaluated on the validation set of BraTS 2020, in which our model makes an effective segmentation for the complete, core and enhancing tumor regions in Dice Similarity Coefficient (DSC) metric (0.89, 0.78, 0.72). And in testing set, our model got the DSC results of 0.87, 0.82, 0.80. Residual convolutional block is especially useful to improve performance in building model. Our proposed method is inherently general and is a powerful tool to studies of medical images of brain tumors.

Keywords: Brain tumor segmentation · Deep learning · Magnetic resonance images

1 Introduction

Gliomas are the most frequent primary brain tumors, which have the highest mortality rate [1, 3, 4, 18]. They can be categorized to low-grade gliomas (LGG) and high-grade gliomas (HGG). HGG is more aggressive form of the disease, which has a median survival rate of two years or less. The slower growing low-grade variants, such as low-grade astrocytomas and oligodendrogliomas, usually makes life expectancy of several years [15]. MRI is a basic modality commonly used in brain structure analysis, which provides images with high contrast for soft tissues and high spatial resolution and can be useful to evaluate unknown health risk [2, 6, 15].

In recent years, lots of automatic approaches have been proposed for accurate segmentation in brain tumors, and these works can be roughly categorized into machine learning methods, deep learning methods and both-combined methods. Machine learning method is based on probabilistic models, which can learn from brain tumor patterns that do not follow a specific model, such as Conditional Random Field (CRF), Random Forrest (RF) and Support Vector Machine (SVM). Deep learning method learns the feature representation in a data-driven way [7], such as convolutional neural network (CNN), parallelized long short-term memory network (LSTM) and fully convolutional network (FCN). In addition, some authors combined probabilistic model (CRF, RF or SVM) and deep learning method to develop a novel method [5, 10, 12].

The fully convolutional neural networks (FCN), a new variant of CNN, gained the great interest in the segmentation competition of PASCAL VOC 2012. The deep convolutional learning model with substantially enlarged depth advanced the state-of-art performance on segmentation tasks that it alleviated the optimization degradation issue by approximating the objective function with residual functions instead of simply stacking layers, and residual block are skip connections between layers of the network. FCN based approaches are the pioneering work of deep learning in medical image segmentation, although the segmented result is not good enough.

In the end-to-end methods, with the combination of encoding layers or decoding layers, they achieved the success of image segmentation in pixel level. Compared to primary convolutional neural network, the end-to-end method can avoid a lot of duplicate calculations. U-Net architecture, based on fully convolution, had been successfully applied to medical image segmentation [9, 17, 19, 21]. This model is a popular and efficient network for segmentation in brain tumors. Naser and Deen [16] proposed a new approach to achieve segmentation in gliomas. They combined U-Net model for convolutional segmentation and pre-trained VGG16 model for transferring learning and a fully connected classifier for tumor grading. For clinical usage, the challenge is how to pursue the best accuracy for segmentation within limited computational budgets. Li et al. [13] proposed a multi-modality aggregation network (MMAN), which was able to extract multi-scale features of brain tissues and harness complementary information from multi-modality MRI images for fast and accurate segmentation. They applied dilated convolutional layers with different kernel size to obtain large-scale features without increasing too many parameters and computational costs. Ding et al. [8] developed a novel multi-path adaptive fusion network. In this model, they applied the idea of skip-connection in ResNets to the dense block so as to effectively reserve and propagate more low-level visual features. Liu et al. [14] investigated the performance of U-Net model in brain tumor, stroke, white matter hyperintensities (WMHs), eye, cardiac, liver, musculoskeletal, skin cancer, and neuronal pathology. They reported the different extended U-shaped networks and analyzed their pros and cons.

In this work, inspired on the groundbreaking proposal on U-Net, we focus on building the U-Net architecture by using residual convolutional blocks. We eval-

uated performances of different residual blocks. In addition, it is a key element to keep gradients independent and distributed identically. We aim to get better segmentation score in BraTS 2020 challenge.

2 Method

2.1 Pre-processing

In this work, we applied cropping and random-slicing methods. As for cropping, due to the GPU memory limitation, we cropped the zero-pixel region which in MRI images before training. The zero-pixel area of image boundary does not help to improve the segmentation accuracy. The original size of MRI images is array size of $155 \times 240 \times 240$. In model, we employed max-pooling function four times that every dimension size must be divided by 16 (2^4). Therefore, considering factors above, we set the size of 3D MRI images as $144 \times 192 \times 192$. As for multimodal 3D images, it is $4 \times 144 \times 192 \times 192$.

For each MRI images, we cropped to nine slices randomly. This step can effectively prevent overfitting during training stage. We randomly take 9 consecutive 3D sequences with length of 16 in the first dimension of the MRI images into training. After randomly cropping, the array size of MRI images is $9 \times 16 \times 192 \times 192$. As for multimodal 3D images, it is $4 \times 9 \times 16 \times 192 \times 192$. In addition, we do the same operation for each epoch during training stage. So the sequences that input to the neural network are generally different for per image and per epoch. This randomization makes the neural network model powerful generalization, especially in limited training data sets. We ensure that all pixels of the brain are trained in training step.

In addition, we employed z-score normalization in medical images [11]. It is accomplished by linearly transforming the original intensities between mean and standard deviation into the corresponding learned landmarks, which defined as:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

where μ is the mean of the MRI sequence in pixel level and σ is the standard deviation of the MRI sequence in pixel level.

2.2 Architecture

We build the architecture of deep learning referring to Fig. 1. It is an end-to-end method of deep learning, which is also a pixel-to-pixel method. Each layer in this model is five-dimensional array size of $bs \times c \times h \times w \times d$, where bs is batch size dimension, c is the channel or multimodal (Flair, T1, T1c and T2) sequences and h , w , d are spatial dimensions. Each convolutional layer and de-convolutional layer contains batch normalization and activation function.

In building this architecture, we refined three primary residual blocks and employed these blocks into encoding stage, in which there are res-block-1, res-block-2 and res-block-3. The residual block is a kind of skip-connect architecture,

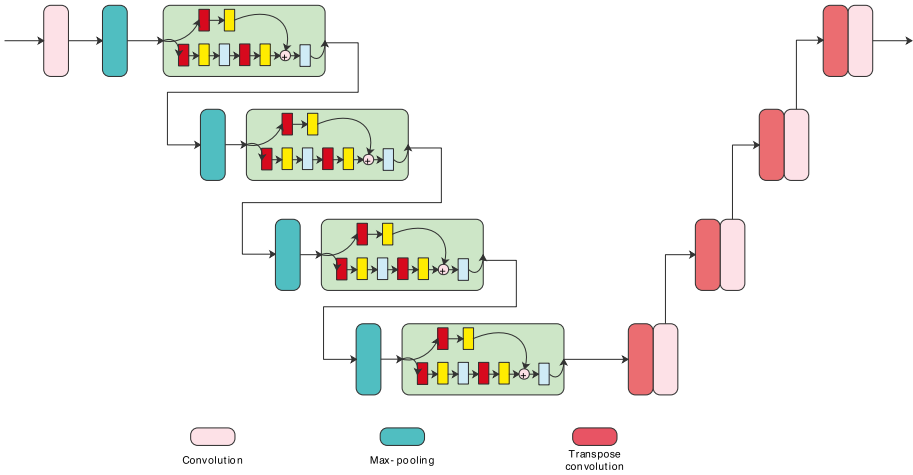


Fig. 1. Our proposed model using res-block-1 in encoding stage. Res-block-1 is shown in Fig. 2

avoiding gradient vanished with increasing depth of network, in which the gradient is effectively transferred to the shallow layer during training network. We apply the randomized leaky rectified liner unit (RReLU) as activated function for neural network.

We designed the res-block-1 block by using a dual-path convolution, an addition operation and a RReLU function. Referring to Fig. 2, we employed two convolutional layers with batch normalization in main path, where RReLU was adopted after the first convolutional layer. And in the skip-path, we employed a convolutional layer and a batch normalization layer. These two path are added by weighted, after that the output data feature activated by a RReLU function.

Referring to Fig. 3, we designed the res-block-2 block by using the same dual-path architecture like res-block-1. In res-block-2, we putted RReLU function to the first position, so that the output feature which computed at the dual-path added each other, and then it putted fused feature to next neural unit.

Referring to Fig. 4, the res-block-3 is the main single convolutional block with a primary skip-connect weights, in which the last convolutional layer is connected after weighted addition operation.

In this model, we apply RReLU as activated function for neural network [20], which defined as:

$$RReLU(x) = \begin{cases} x & x > 0 \\ ax & otherwise \end{cases} \tag{2}$$

where a is randomly sampled from uniform distribution $U(L, R)$. L is lower bound of the uniform distribution and R is upper bound of the uniform distribution. We set L of $1/8$ and set R of $1/3$.

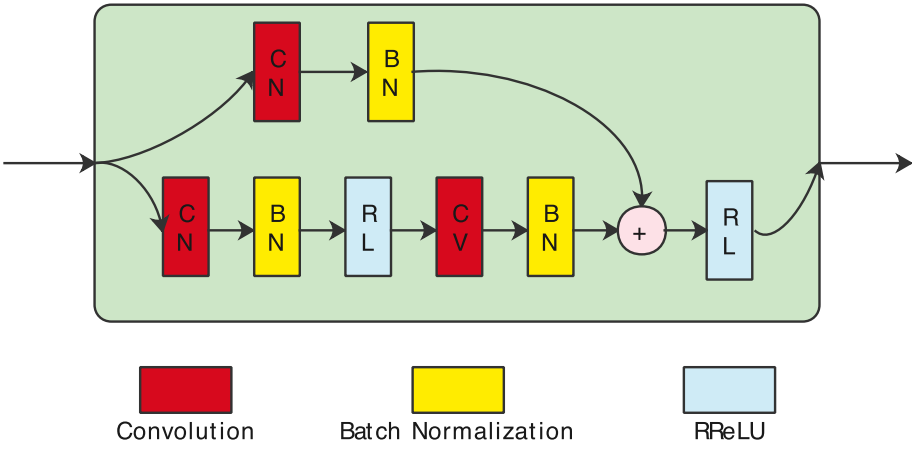


Fig. 2. Res-block-1.

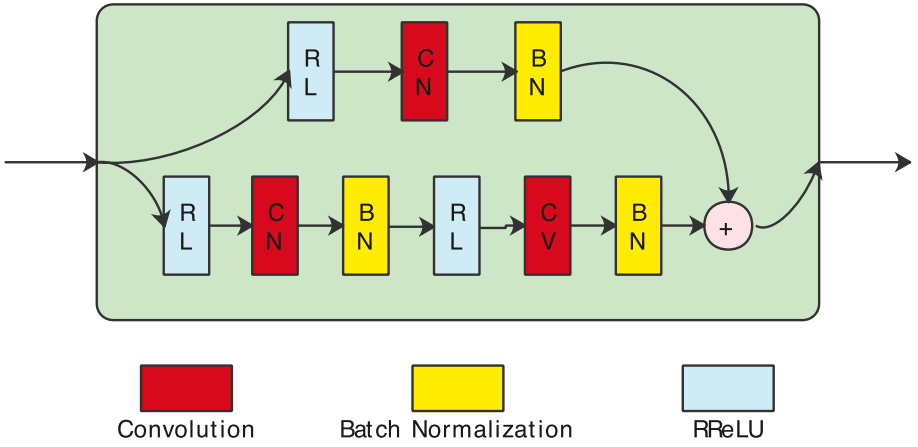


Fig. 3. Res-block-2.

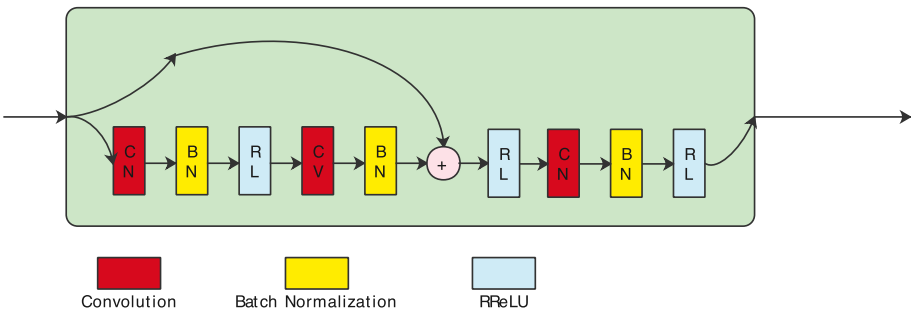


Fig. 4. Res-block-3.

As for loss function, it is used to calculate the loss of training which used in back propagation. We used the Categorical Cross-entropy. The loss function can be described as:

$$\begin{aligned} \text{loss}(x, \text{class}) &= -\log\left(\frac{\exp(x(\text{class}))}{\sum_j \exp(x(j))}\right) \\ &= -x(\text{class}) + \log\left(\sum_j \exp(x(j))\right) \end{aligned} \quad (3)$$

which combines *LogSoftmax* and *NLLLoss* in one single class. As for *NLLLoss* ($-x(\text{class})$) function, the negative log likelihood loss, it is useful to train a classification problem with *class* classes, and obtaining log-probabilities in a neural network is easily achieved by adding a *LogSoftmax* layer in the last layer. The Categorical Cross-entropy function is useful to solve the classification problem with multi-classes.

In encoding stage, we employ lots of convolutional layers to extract features from MRI images. And we set parameters of convolutional function with kernel size of 3, stride of 1, padding of 1. Channels, in encoding stage, are 32, 64, 128, 256 and 512 respectively. We use max-pooling function to down-sampling so that model get deep features and learn segmentation ability from its.

In decoding stage, we employ transposed convolutional layer to up-sampling, which makes the output 3D images with the same size of the input 3D images. The transposed convolution is effective and very easy to implement.

3 Experiments

Our method was evaluated on BraTS 2020 dataset.

3.1 Dataset

The BraTS 2020 dataset contains four modes for every patient: Flair, T1, T1c and T2. We trained our model in BraTS 2020 training set, which contains 369 MRI scans including high-grade and low-grade brain tumor. In addition, the validation set contains 125 scans of glioblastoma and testing set contains 166 scans of glioblastoma. BraTS challenge has always been focusing on the evaluation of state-of-art methods for the segmentation for brain tumors in multimodal magnetic resonance imaging scans. Metrics for this challenge are computed through the online evaluation platform that the ground truth labels are not available for public. Every region of gliomas needs to be segmented pixel-to-pixel sequences for 4 meaningful regions: the enhancing tumor (ET), the tumor core (TC), the whole tumor (WT) and normal tissues.

3.2 Setup

Some of the hyper-parameters of the architectures were shown in Table 1. We approached brain tumor segmentation as a multi-class classification problem,

segmented normal tissue, necrosis, edema, non-enhancing, and enhancing tumor from MR images respectively. However, the given MR images are not suitable for pouring into neural network directly that the redundant data will cost large GPU memory. So we cropped images of effective parts in three dimensions. Similarly, the same process was used on label set. Additionally, in brain tumor segmentation, the number of samples of necrosis and enhancing tumor is small in training set. To deal with that, we normalized all pixel-level image using z-score (zero-mean) normalization, which made the input data follow a normal distribution and speeded up training. The learning rate was linearly decreased each epoch during the training stage. Our model was developed using PyTorch. We train the model using four GPUs of Nvidia RTX 2080 TI with 40 h.

3.3 Evaluation

The evaluation metrics of brain tumor segmentations consist of three types of measures: Dice similarity coefficient (DSC), Sensitivity and Specificity. The DSC measures the spatial overlap between the automatic segmentation and the label. It is defined as:

$$DSC = \frac{2TP}{FP + 2TP + FN} \quad (4)$$

where FP, FN and TP are false positive, false negative detections and true position, respectively. Sensitivity, also called the true positive rate or probability of detection, measures the proportion of positives that are correctly identified as such:

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

A larger value of Sensitivity denotes a higher proximity of abnormal tissue between label and prediction of segmentation. Finally, specificity, also called the true negative rate, measures the proportion of negatives. It is defined as:

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

where TN is true negative detections. A larger value of Specificity denotes a higher proximity of normal tissue between label and prediction of segmentation.

3.4 Result

We evaluate our proposed model on validation with three different residual block, and compared with other state-of-art methods. Lastly, we report the result of segmentation on BraTS 2020 testing dataset. The performance of our model is presented on Fig. 5.

Referring to Table 2, the Res-Block-1 get better performance of segmentation than others. In Dice metric of WT region, the Res-Block-2 gain a little advantage. In addition, all the res-block get good score in segmentation of WT region.

Table 1. Hyper-parameters of our proposed model. The weights and bias in initialization are each convolutional layers' setting. We set the randomized leaky ReLU with default parameter setting.

Stage	Hyper-parameters	Value
Initialization	Weights	1.0
	Bias	0.0
Training	Max epoch	100
	Batch size	4
	Learning rate	0.002
	Learning rate decay	0.95
	GPU RTX 2080Ti	4
Randomized leaky ReLU	Lower	1/8
	Upper	1/3
	Inplace	True

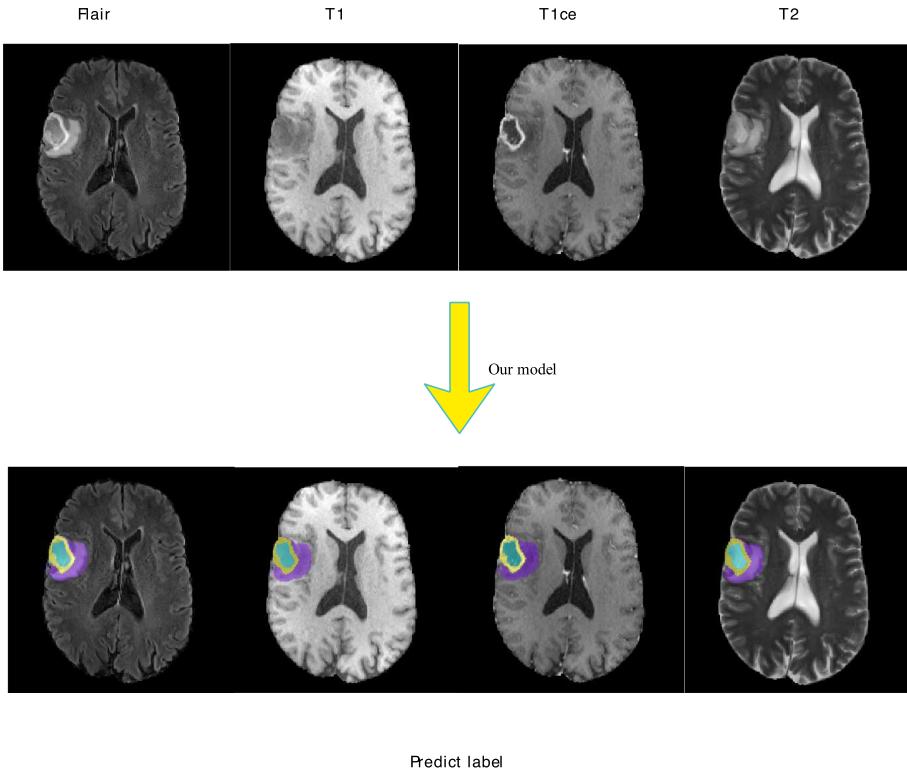


Fig. 5. Our predict MRI images using res-block-1 block.

Table 2. Segmentation result of Res-Block-1, Res-Block-2 and Res-Block-3 on BraTS 2020 Validation dataset

Method	Dice			Sensitivity			Specificity		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
Res-Block-1	0.723	0.892	0.788	0.732	0.908	0.776	0.999	0.998	0.999
Res-Block-2	0.690	0.897	0.748	0.717	0.920	0.750	0.999	0.998	0.999
Res-Block-3	0.680	0.891	0.740	0.686	0.885	0.693	0.999	0.999	0.999

Table 3. Segmentation result of our proposed model, DeepLab and U-Net model on BraTS 2020 Validation dataset

Method	Dice			Sensitivity			Specificity		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
Our method	0.723	0.892	0.788	0.732	0.908	0.776	0.999	0.998	0.999
DeepLab	0.707	0.884	0.774	0.742	0.876	0.754	0.999	0.999	0.999
U-Net	0.688	0.871	0.703	0.720	0.835	0.687	0.999	0.999	0.999

However, it is diverse to design res-block in U-shaped like model. Further experimental investigations are needed to estimate the performance of these decoding method in segmentation of medical image.

We compared our model with DeepLab and U-Net model by applying the same pre-processing methods for quantitative study. Our study is focus on performance of deep learning neural model. The results are reported on Table 3. The most difficult tasks in this brain tumor segmentation is marking the tumor core region for LGG and the enhancing tissues for HGG. To compare with two classical end-to-end model and referring to Table 3, our proposed model outperformed these models in Dice metrics.

Table 4. Segmentation result of our proposed model on BraTS 2020 testing dataset

Method	Dice			Sensitivity			Specificity		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.803	0.872	0.823	0.808	0.896	0.819	0.999	0.999	0.999
StdDev	0.202	0.132	0.250	0.220	0.138	0.246	0.0003	0.001	0.0006
Median	0.854	0.913	0.910	0.886	0.936	0.920	0.999	0.999	0.999
25quantile	0.774	0.850	0.832	0.792	0.879	0.796	0.999	0.998	0.999
75quantile	0.916	0.944	0.951	0.935	0.965	0.956	0.999	0.999	0.999

The BraTS challenge testing result is reported on Table 4. In segmentation of the core and the enhancing tumor, our proposed method have better performance on testing set.

4 Conclusion

In this work, we propose a U-shaped architecture using residual block. We evaluated performance of different residual block in this U-shaped architecture. Residual block is an effective block to build deep neural network in feature extraction stage. In brain tumor segmentation, there are lots of deep-learning models including 2D and 3D model that the architecture becomes more and more complex as the development of computer hardware and the result of segmentation becomes more and more precise. Our research approach is a powerful tool to studies of 3D medical images of brain tumors and our proposed model is an effective deep-learning model, especially in 3D brain tumor segmentation.

Acknowledgement. This work was supported in part by National Natural Science Foundation of China under Grant No. 61976126, Shandong Natural Science Foundation under Grant No. ZR2019MF003, No. ZR2017MF054.

References

1. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *Cancer Imaging Arch.* **286** (2017)
2. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection, July 2017. <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
3. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **4**, 170117 (2017)
4. Bakas, S., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint [arXiv:1811.02629](https://arxiv.org/abs/1811.02629) (2018)
5. Bauer, S., Nolte, L.-P., Reyes, M.: Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 354–361. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23626-6_44
6. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of MRI-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* **58**(13), R97 (2013)
7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
8. Ding, Y., Gong, L., Zhang, M., Li, C., Qin, Z.: A multi-path adaptive fusion network for multimodal brain tumor segmentation. *Neurocomputing* **412**, 19–30 (2020)
9. Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y.: Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In: Valdés Hernández, M., González-Castro, V. (eds.) MIUA 2017. CCIS, vol. 723, pp. 506–517. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60964-5_44

10. Havaei, M., Jodoin, P.M., Larochelle, H.: Efficient interactive brain tumor segmentation as within-brain KNN classification. In: 2014 22nd International Conference on Pattern Recognition, pp. 556–561. IEEE (2014)
11. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. *Pattern Recogn.* **38**(12), 2270–2285 (2005)
12. Lee, C.-H., Wang, S., Murtha, A., Brown, M.R.G., Greiner, R.: Segmenting brain tumors using pseudo-conditional random fields. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008. LNCS, vol. 5241, pp. 359–366. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85988-8_43
13. Li, J., Yu, Z.L., Gu, Z., Liu, H., Li, Y.: MMAN: multi-modality aggregation network for brain segmentation from MR images. *Neurocomputing* **358**, 10–19 (2019)
14. Liu, L., Cheng, J., Quan, Q., Wu, F.X., Wang, Y.P., Wang, J.: A survey on U-shaped networks in medical image segmentations. *Neurocomputing* **409**, 244–258 (2020)
15. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2014)
16. Naser, M.A., Deen, M.J.: Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. *Comput. Biol. Med.* **121**, 103758 (2020)
17. Noori, M., Bahri, A., Mohammadi, K.: Attention-guided version of 2D UNet for automatic brain tumor segmentation. In: 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), pp. 269–275. IEEE (2019)
18. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **35**(5), 1240–1251 (2016)
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
20. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. arXiv preprint [arXiv:1505.00853](https://arxiv.org/abs/1505.00853) (2015)
21. Xu, F., Ma, H., Sun, J., Wu, R., Liu, X., Kong, Y.: LSTM multi-modal UNet for brain tumor segmentation. In: 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC), pp. 236–240. IEEE (2019)