Modified U-Net for Automatic Brain Tumor Regions Segmentation

Keerati Kaewrak, John Soraghan and Gaetano Di Caterina
Centre for Signal and Image Processing (CeSIP),
Department of Electronic and Electrical Engineering
University of Strathclyde
Glasgow, UK
keerati.kaewrak@strath.ac.uk, j.soraghan@strath.ac.uk,
gaetano.di-caterina@strath.ac.uk

Derek Grose
Beatson West of Scotland Cancer Centre
Glasgow, UK
Derek.Grose@ggc.scot.nhs.uk

Abstract— **Novel deep learning based network architectures are investigated for advanced brain tumor image classification and segmentation. Variations in brain tumor characteristics together with limited labelled datasets represent significant challenges in automatic brain tumor segmentation. In this paper, we present a novel architecture based on the U-Net that incorporates both global and local feature extraction paths to improve the segmentation accuracy. The results included in the paper show superior performance of the novel segmentation for five tumor regions on the large BRATs 2018 dataset over other approaches.**

Keywords— segmentation, tumor, u-net

I. INTRODUCTION

Convolutional Neural Networks (CNNs) have become an efficient tool for image classification and segmentation including brain tumor segmentation [1]. Conventional CNNs comprise a convolutional layer, pooling layers and a fully connected layer. These layers are stacked together to form a full CNNs architecture. Because tumor appearances vary from patient to patient, brain tumor segmentation represents a particularly challenging segmentation problem. Early diagnosis of brain tumor is one of the factors that gives higher survival rate to patients [2, 3], hence an application of deep learning on tumor regions segmentation could benefit the tumor treatment planning by assisting the oncologists with less time-consuming tumor regions delineation acquisition.

In this paper, we present a novel deep neural network architecture that is based on the U-Net for brain tumor segmentation. Since tumor structures vary across individuals and the underestimation of tumor region can lead to inefficient treatment planning, the novel architecture uses the idea of TwoPathCNN by convolving the kernel size of 9×9 and 3×3, respectively, over the MRI input images to provide global and local feature extraction into our proposed model. These global and local feature maps are used in the modified U-Net architecture to give the model two different aspects for the segmentation classes prediction. Global feature could provide larger context information of region around the image pixel. The proposed network contains a total number of 16 convolutional layers. We implemented the proposed network to segment five tumor regions without data augmentation and obtained promising relative results for all tumor regions.

The remainder of this paper is organized as follows. Related deep learning based brain tumor segmentation work is presented in Section II. The clinical dataset and proposed network architecture are presented in Section III. Comparative experimental results are presented in section IV. Section V concludes the paper.

II. RELATED WORKS

Previous application of machine learning to brain tumor segmentation such as the semi-automatic tumor segmentation in [4] requires users to initialize the region of interest (ROI) by roughly drawing around the suspected tumor on an MRI slide. Subsequently an algorithm that combined the techniques of region and edge-based active contour and level set method together to segment the tumor region was applied. In "Tumor-cut" method [5], users drew the maximum diameter of the tumor on the input MRI slides before applying a cellular automata (CA) based seeded tumor segmentation algorithm. However, to eliminate human interaction, fully automatic segmentation is currently our main focus.

Pereira et al. [6] proposed fully-automatic deep CNNs architecture using a small kernel size (3×3) in convolutional layers to extract features with the increasing depth. Havaei et al. [7] increased training data by feeding the input image patches instead if whole image through two-pathway CNN architecture (TwoPathCNN) which consists of local and global paths using different sizes of convolution kernels to capture local and global feature maps then concatenated them together and used the concatenation feature maps to perform the prediction of class labels. In 2015 Ronneberger et al. [8] introduced the U-Net which is an encoder-decoder network architecture aimed to tackle the limitation of deep convolutional network in biomedical image segmentation that required deeper network and larger size of training set in order to work well. Semantic segmentation which assigns a certain class label to each pixel and gives localization information is a desired output for biomedical image segmentation problem rather than just a single class image classification.

As illustrated in Fig.1 the original U-Net comprises a down-sampling and up-sampling path. Both paths are connected with concatenation of corresponding feature maps. This concatenation of the feature maps provides the localization context to higher resolution feature maps, and give more precise segmentation results using only a relatively few training images. The down-sampling path of U-Net follows a typical convolutional neural network architecture which extracts essential features from input data. The down-sampling stream consists of 5 blocks of two repeated 3×3 convolutions, each followed by a rectified linear unit (ReLU)
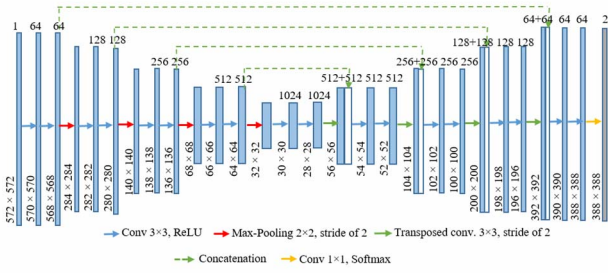
Fig.1. The original U-Net architecture.

$f(x) = \max(0, x)$ as an activation function, and a 2×2 max pooling operation with stride 2. The number of feature channels are doubled at each down-sampling step while the feature maps resolution is halved in both dimensions. This down-sampling path extracts feature map from 1 channel image to 1024 feature channels. The up-sampling path comprises 4 blocks of a 3×3 transposed convolution with stride of 2 that expands the feature maps back to the original size in the final block of up-sampling stream. The transposed convolution [9] allows the model to learn the weights for expanding the resolution of the feature maps to original size. The transposed convolution process is defined as

$$y^i = x_k^i * f \qquad (1)$$

where x represents the 2-d feature vector for channel i, f is a 3×3 matrix of 1 and $*$ is 2D convolution operator. An example of a 3×3 transposed convolution with a stride of 2 that double the image resolution in both dimensions and halves the number of feature channels is illustrated in Fig.2. A concatenation with the correspondingly cropped feature map from the down-sampling path, and two repeated 3×3 convolutions, each followed by ReLU is used to reduce the feature maps from 1024 to 64 channels before the finally layer of a 1×1 convolution [10] maps each 64-component feature vector to the desired number of classes. As the characteristics of tumor structures are different for each patient combined
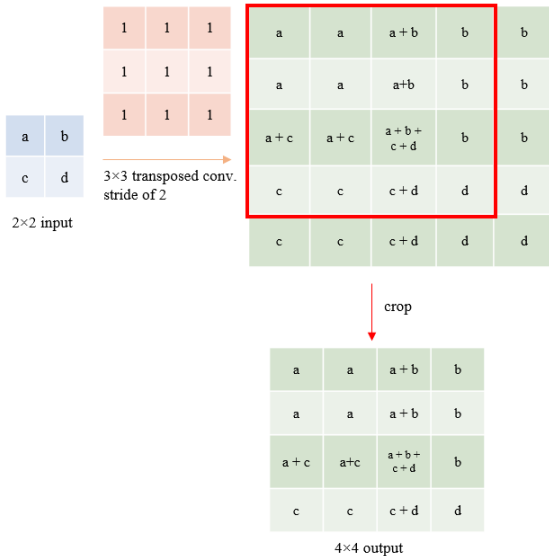

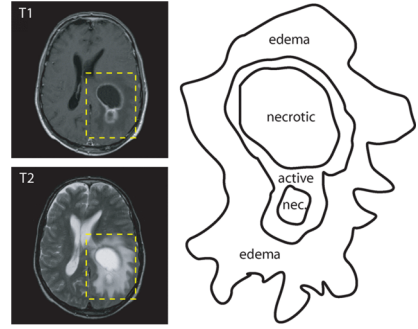Fig. 2. An example transposed convolution process using for up-sampling.

Fig. 3. Left – examples of T1 and T2 MRI image of the brain tumor, right – an illustration of tumor region and its label [13].

with a limited amount of labelled data present major challenges in the development of automatic segmentation algorithms based on deep learning. U-Net like models were used for brain tumor segmentation in [11, 12].

A brain tumor is an uncontrollable growth of abnormal cells of the brain tissues. The World Health Organization (WHO) classified brain tumor into over 100 types and also graded the tumor according to its behaviour. Low-grade tumors are slow-growing and have less chance of recurring after getting completely removed while high-grade tumors are fast-growing, more aggressive and tend to spread to other parts of the brain or spinal cord [3]. Glioma is a type of brain tumor that abnormally grows from glial cells of the brain and it is the most common type of brain tumor that found in both adults and children. MRI is currently imaging modality of choice for brain tumor assessment because of its superior soft-tissue contrast. Fig.3 gives an example of brain tumor appearance on T1 and T2 MRI images and illustrates intra-structures of tumor. Necrotic and non-enhancing region is the area which considered as dead part of the tumor appear with low intensity in T1 image with high intensity surrounding boundary while edema which is the swelling non-tumor area affected by the tumor appear brighter in T2 image. Enhancing tumor is the active part of the tumor between edema and necrotic. But in case of low-grade glioma, the distinction of these structures may not be seen clearly. Since the intra-structure appear differently on each MRI modality as shown in Fig.3, it is important to use multimodal MRI for brain tumor segmentation [13].

Dong et al. presented a model [11] that relies on data augmentation to increase training dataset however these as seen to be unable to segment enhancing tumor region in low-grade glioma case which is the most difficult region for segmentation task. Shreyas et al. [12] proposed a model that used empirically derived class weights with batch-normalization but only implemented the model on high-grade-glioma images data.

III. MATERIAL AND METHOD

A. Dataset

In this study, we trained and tested our proposed model on BRATs 2018 dataset which contains multimodal MRI scans
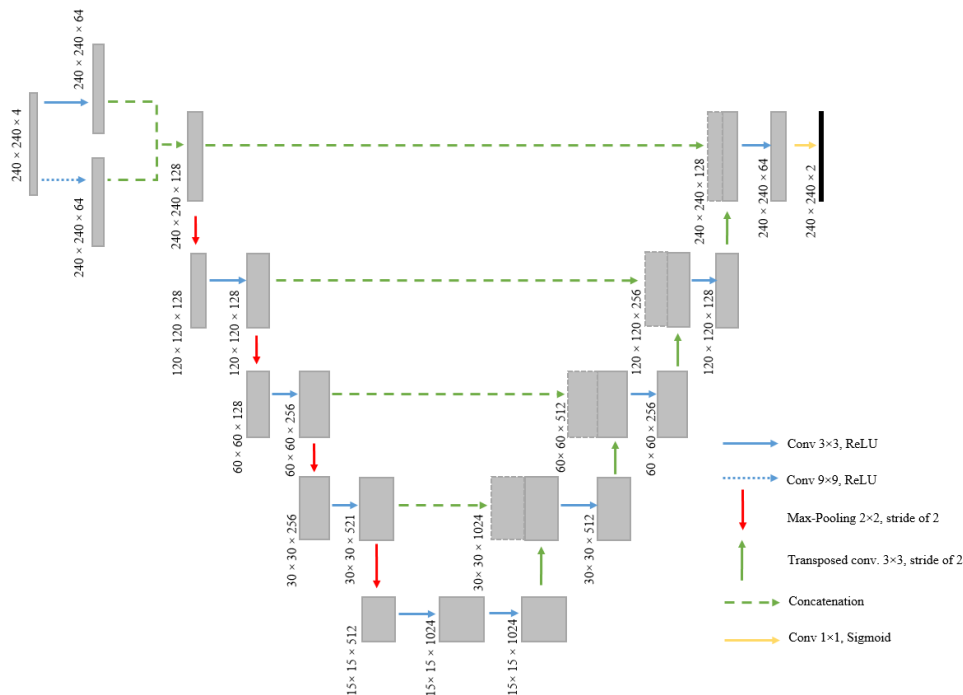
Fig. 4. Proposed network architecture.

from 210 high-grade glioma (HGG) patients and 70 low-grade glioma (LGG) patients [14]. The dataset provides four MRI scans; T1-weighted (T1), T1-weighted with gadolinium enhancing contrast (T1+Gd), T2-weighted and FLAIR image volumes for each patient. Each type of MRI scans was obtained by using different MRI sequences setting and hence tumor regions appear differently on each MRI scan [15]. These images were resampled and interpolated into $1\times1\times1$ m^3 with the size of $240\times240\times155$.

TABLE I.   Tumot regions and corresponding annotated labels from BRATs 2018 dataset.

Tumor regions	Annotated label by manually segmentation
Necrotic and non-enhancing tumor	1
Edema	2
Enhancing tumor	4
Whole tumor	1,2,4
Tumor core	1,4
Everything else	0

Table I shows that the dataset provides annotated labels of intra-tumor structures from manual segmentation by expert oncologists as follows. Label 1 for necrotic and non-enhancing tumor region which is visible in T1+Gd images. Label 2 for edema which appear with high intensity visible in T2 and FLAIR images. Label 4 for enhancing tumor or active tumor region visible in T1+Gd images. Whole tumor region contains all three intra-tumor structures (label 1,2 and 4). Tumor core contains all tumor region excepts edema (label 1 and 4) and can be seen best in T1+Gd images. The different appearances of theses intra-structure can be seen first four column of Fig.5. Finally, label 0 is for everything else [1]. We used these consensus segmentation labels as the ground truth

for the network training and evaluation. We performed data normalization for each MRI scan by subtracting the mean of each MRI scan and dividing by its standard deviation.

B. Proposed Network Architecture

The proposed networked architecture is based on U-Net [8] consisting of down-sampling and up-sampling paths. In the first block of down-sampling path, as illustrated in Fig.4 two convolutional layers with kernel size of 3×3 and 9×9 are used followed by a ReLU to capture low-level local and global feature maps from multimodal MRI input instead of two repeated 3×3 convolutional layers like original U-Net. Global and local feature maps are then concatenated. Using global and local paths in low-level feature extraction provides the network contextual information of the essential feature. A 2×2 max-pooling with stride of 2 is used to halve the image size. A 3×3 Convolutional layer followed by ReLU and 2×2 max-pooling with stride of 2 are repeated until image resolution deceases from 240×240 to 15×15 and feature maps increases from 4 to 512. Two repeated 3×3 convolutional layers followed by a ReLU are used in the fifth block of down-sampling path to double feature maps to 1024.

For the up-sampling path, a 3×3 transposed convolution with stride of 2 as illustrated in Fig.2 is used to double image size in both dimensions and to halve the feature maps. The result is then concatenated with the corresponding feature maps from the same level in the down-sampling path as shown in Fig.4 providing the higher resolution feature map the localization context and allowed more precise segmentation. A 3×3 convolutional layer followed by a ReLU is used to halve the concatenation feature maps. The process is repeated until the image resolution increased from 15×15 to original 240×240 with 64 feature maps. Finally, a 1×1 convolutional layer [10] comprising two filters of length
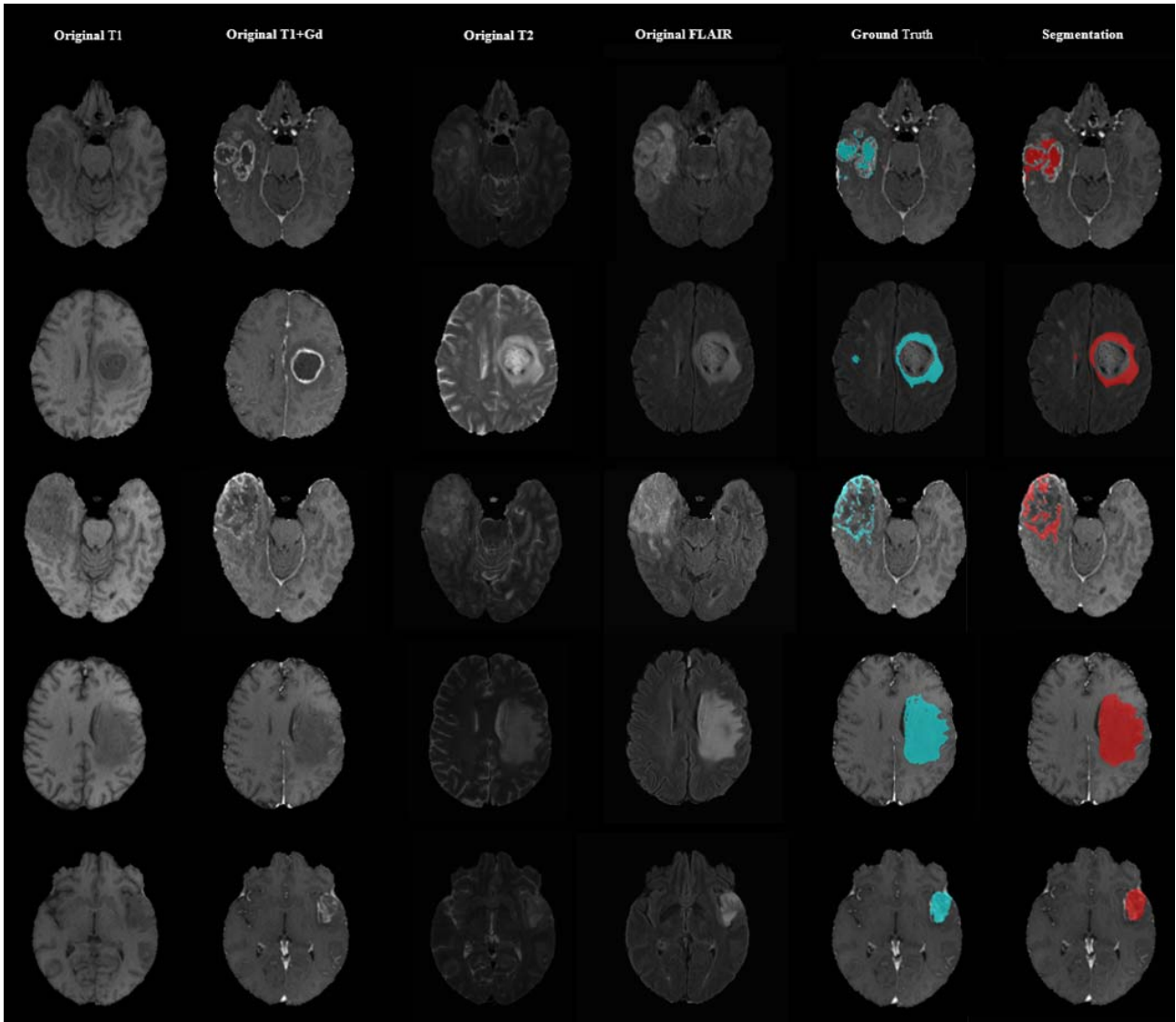
Fig. 5. Examples of MRI images and segmentation results from test set. From left to right; multimodal MRI scans, ground truth mask (blue) and segmentation mask (red) overlaid on original MRI of necrotic and non-enhancing, edema, enhancing tumor, whole tumor and tumor core regions.

1×1×64 is employed to produce a 240×240×2 output. One for the foreground and one for the background segmentation.

C. Network Training

An Adam optimizer [16] was used with a learning rate of 0.00001, values of $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an epoch size of 50. All weights were initialized using a normal distribution with mean of 0 and standard deviation of 0.01, and all biases were initialized as 0. Dice Loss (DL_2) proposed by [17] was used as a loss function and can be defined as

$$DL_2 = 1 - \frac{\sum_{n=1}^{N} p_n r_n + \varepsilon}{\sum_{n=1}^{N} p_n + r_n + \varepsilon} - \frac{\sum_{n=1}^{N} (1 - p_n)(1 - r_n) + \varepsilon}{\sum_{n=1}^{N} 2 - p_n - r_n + \varepsilon} \quad (2)$$

where r_n, p_n are the pixel value of the ground truth and the predicted probability map of nth element, respectively. The ε term is used to ensure loss function stability. We implemented the network using Tensorflow with Tensorlayer

library on a PC equipped with NVIDIA GeForce GTX 1070 GPU, an Intel(R) Core(TM) i5-8400 CPU 2.80 GHz processor, 16 GB of RAM.

D. Experiments and Evaluations

We trained and tested the proposed networks for 5 segmentation classes according to tumor regions as defined in Table I.

We evaluated the network performance by comparing predicted segmentation of the tumor regions to ground truth provided by BRATs 2018 dataset. Dice similarity coefficient (DSC) and intersection over union (IOU) which give the similarity between predicted tumor region and ground truth by comparing the overlapped regions are used and can be defined as

$$DSC = \frac{2TP}{FP + 2TP + FN} \quad (3)$$

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

where *TP*, *FP* and *FN* denote the number of true positive, false positive and false negative counts, respectively.

TABLE II.   PROPOSED NETWORK PERFORMANCE OF BRAIN TUMOR REGIONS SEGMENTATION ON BRATs 2018 DATASET.

Task	DSC		IOU	
	Training	*Testing*	*Training*	*Testing*
Whole tumor	0.94	0.81	0.89	0.69
Necrotic and Non-enhancing tumor	0.81	0.45	0.71	0.31
Edema	0.90	0.67	0.82	0.52
Enhancing tumor	0.83	0.70	0.72	0.56
Tumor core	0.93	0.62	0.88	0.47

TABLE III.   DICE SIMILARITY COEFFICIENT OF PROPOSED NETWORK COMPARING TO PUBLISHED RESULTS OF TRAINGING PHASE OF BRAIN TUMOR REGIONS SEGMENTATION.

Method	Data	DSC		
		Whole	*Core*	*Enhance*
Proposed network	Brats 2018	**0.94**	**0.93**	**0.83**
U-Net [11]	Brats 2015	0.86	0.86	0.65
TwoPathCNN [7]	Brats 2013	0.85	0.78	0.73

IV. RESULTS AND DISCUSSION

The image data of high-grade glioma (HGG) and low-grade glioma (LGG) patients were trained together for each segmentation task and five-fold cross validation were conducted to choose the best epoch performance. The segmentation masks of foreground and background of predicted tumor regions were obtained after training and testing phase end as illustrated in the last column of Fig.5.

Mean DSC and IOU of training and testing set are reported in Table II. From Table II it is seen that a DSC score > 0.80 is obtained for all tumor regions segmentation in the training phase. However, the network performance dropped in the testing phase for necrotic and non-enhancing tumor regions. Since we did not use data augmentation to increase training dataset and the necrotic and non-enhancing region has unique characteristic in each tumor patient, these could be the main reason of poor performance in this specific tumor region. A comparison of training phase segmentation results of the proposed network to other published network performance based on U-Net and TwoPathCNN with data augmentation is given in Table III. The proposed model gave highest DSC score in all segmentation tasks.

For future work, we aim to improve the network performance especially for testing phase, and modify the network for multi-class segmentation so we are able to segment all tumor regions simultaneously.

V. CONCLUSION

In this study, we proposed a novel encoder-decoder network architecture based on U-Net architecture that had proved to work well in biomedical image segmentation. We used a kernel size of 9×9 and 3×3 as global and local feature extraction paths in our proposed novel deep network for brain tumor regions segmentation. We trained and tested the proposed network on BRATs 2018 dataset. Without data augmentation and fewer convolutional layers, the proposed

model gave promising segmentation results for all tumor regions.

REFERENCES

[1]   B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging,* vol. 34, no. 10, pp. 1993-2024, 2015.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2]   I. Goetz and A.-L. Grosu, "Advanced Imaging Modalities and Treatment of Gliomas: Radiation Therapy," in *Brain Tumor Imaging*, E. Hattingen and U. Pilatus, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 135-142.K. Elissa, "Title of paper if known," unpublished.

[3]   D. N. Louis *et al.*, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathologica,* vol. 131, no. 6, pp. 803-820, 2016/06/01 2016.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[4]   X. Guo, L. Schwartz, and B. Zhao, "Semi-automatic segmentation of multimodal brain tumor using active contours," *Multimodal Brain Tumor Segmentation,* vol. 27, 2013.

[5]   A. Hamamci, N. Kucuk, K. Karaman, K. Engin, and G. Unal, "Tumor-Cut: Segmentation of Brain Tumors on Contrast Enhanced MR Images for Radiosurgery Applications," *IEEE Transactions on Medical Imaging,* vol. 31, no. 3, pp. 790-804, 2012.

[6]   S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *IEEE Transactions on Medical Imaging,* vol. 35, no. 5, pp. 1240-1251, 2016.

[7]   M. Havaei *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Medical Image Analysis,* vol. 35, pp. 18-31.

[8]   O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015: Springer International Publishing, pp. 234-241.

[9]   M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2528-2535.

[10]   M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400,* 2013.

[11]   H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Annual Conference on Medical Image Understanding and Analysis*, 2017: Springer, pp. 506-517.

[12]   V. Shreyas and V. Pankajakshan, "A deep learning architecture for brain tumor segmentation in MRI images," *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP),* pp. 1-6, 2017.

[13]   J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille, "Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification," *IEEE Transactions on Medical Imaging,* vol. 27, no. 5, pp. 629-640, 2008.

[14]   S. Bakas *et al.*, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific Data,* vol. 4, p. 170117, 2017.

[15]   O. Keunen *et al.*, "Multimodal imaging of gliomas in the context of evolving cellular and molecular therapies," *Advanced Drug Delivery Reviews,* vol. 76, pp. 98-115, 2014.

[16]   D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv e-prints*, 2014

[17]   F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *2016 Fourth International Conference on 3D Vision (3DV),* pp. 565-571, 2016.