A Modified U-Net Convolutional Network Featuring a Nearest-neighbor Re-sampling-based Elastic-Transformation for Brain Tissue Characterization and Segmentation

S. M. Kamrul Hasan[1] and **Cristian A. Linte**[1,2]

[1]Chester F. Carlson Center for Imaging Science, Visualization and Image-guided Navigation Lab Rochester Institute of Technology, Rochester, NY USA

[2]Biomedical Engineering Biomedical Modeling, Visualization and Image-guided Navigation Lab Rochester Institute of Technology, Rochester, NY USA

Abstract

The detection and segmentation of brain tumors from Magnetic Resonance Imaging (MRI) is a very challenging task, despite the availability of modern medical image processing tools. Neuro-radiologists still diagnose deadly brain cancers such as even glioblastoma using manual segmentation. This approach is not only tedious, but also highly variable, featuring limited accuracy and precision, and hence raising the need for more robust, automated techniques. Deep learning methods such as the U-Net deep convolutional neural networks have been widely used in biomedical image segmentation. Although this model was demonstrated to yield desirable results on the BRATS 2015 dataset by using a pixel-wise segmentation map of the input image as an auto-encoder, which assures best segmentation accuracy, the output only showed limited accuracy and robustness for a number of cases. The goal of this work was to improve the U-net model by replacing the de-convolution component with an up-sampled by the Nearest-neighbor algorithm and also employing an elastic transformation to augment the training dataset to render the model more robust, especially for the segmentation of low-grade tumors. The proposed Nearest-Neighbor Re-sampling Based Elastic-Transformed (NNRET) U-net Deep CNN framework has been trained on 285 glioma patients BRATS 2017 MR dataset available through the MICCAI 2017 grand challenge. The framework has been tested on 146 patients using Dice similarity coefficient (DSC) & Intersection over Union (IoU) performance metrics and outweighed the classic U-net model.

Keywords

Brain tissue segmentation; nearest-neighbor interpolation; deep convolutional networks; modified U-net

I. INTRODUCTION

According to American Cancer Society, more than 80,000 people are newly diagnosed with cancer every year in the United States. Of all cancer patients, approximately 32% are diagnosed with malignant brain tumors that have a 5-year survival rate of 5.3% [1]. Brain tumors - one of the most quotidian neurological brain disorders with a devastating-effect on many lives represent an uncontrolled mass of tissues found in different parts of the brain.

Gliomas constitute the most common and deadliest type of brain tumors, with a subset ranging from slow growing low- to high-graded malignant tumors known as glioblastomas. As these tumors are generally localized in the posterior cranial fossa of human brain, their detection is challenging by means of biopsy, raising the need for noninvasive, image processing-based methods for detection and characterization. Image segmentation entails the partitioning of an image into its constituent regions of interest.

The human brain consists of five types of soft tissues - white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), edema, and tumor tissue all of which appear different thanks to their differences in magnetic environment when imaged using magnetic resonance imaging (MRI), therefore enabling their differentiation (Fig. 1). As an example, high-graded gliomas (HGG) and necrotic tissues are delineated easily in a MRI image, while low-graded gliomas (LGG) and certain tumor tissue are much more difficult to identify and segment. Hence, it is important to devise segmentation techniques that are sufficiently accurate and robust to ensure accurate diagnosis and the most appropriate course of therapy.

Despite ongoing research in brain image segmentation, the vast variability and heterogeneity of the brain MRI data raises the need for more efficient segmentation techniques. To segment various types of brain tumors, some efforts have focused on the use of supervised algorithm that relied on a randomized tree for segmenting BRATS FLAIR MRI images [2]. This method yielded DICE similarity on the order of 88% for both LGG and HGG cases, but it was highly dependent on tuning the super-pixel size that could affect the final detection of the tumor. In recent years, the use of deep neural networks is becoming very popular for semantic segmentation, with a first application to medical imaging disseminated in [3]. The authors in [4] used the modified version of U-net where they employed segmentation layers in the localization pathway and combined them to form the final network output. However, their work did not show the performance of the new algorithm for both high-grade and low-grade tumor images.

The fully connected layer problem was minimized by introducing a pixel-wise best deep convolutional neural network formulation as proposed in [5]. This U-net architecture works like an auto-encoder where the input is the same as output. Auto-encoders work by compressing the input into a latent-space and then reconstruct the output. Using this U-net architecture, the authors in [8] first implemented a fully automated brain tumor segmentation method and applied it to the BRATS 2015 dataset, using a small sub-set of the data to rain their network.

The classic U-net model [6] relies on the transposed convolution or deconvolution, in a similar, yet opposite fashion to the convolutional layers. I.e., instead of mapping from 4×4 input pixels to 1 output, they map from 1 input pixel to 4×4 output pixels. Despite its learnable parameters [7], this methods performance is much slower, as the filters need additional weights to train. Additionally, it can easily lead to "uneven overlap", characterized by checkerboard-like pattern resulting in artifacts on a variety of scales affecting unusual colors (Fig. 2). While it is difficult to entirely remove these artifacts, they can be minimized.

II. METHODOLOGY

A. Overview

To address the limitations associated with the classical U-Net architecture, we first selected a kernel size to avoid overlap and to further remove the artifacts. We resized the image using nearest-neighbor interpolation. In essence, we used an up-sampling operation (i.e., nearest neighbor interpolation) followed by a convolution operation with a carefully selected stride and kernel size. The main distinguishing feature of this proposed architecture over its precursor classical U-net framework consists of the up-sampling step. The U-net architecture with an up-sampling interpolation step for resizing the images prior to performing the transposed convolution provided additional robustness to the segmentation, especially for low-graded glioma tumors. The network was trained on the BRATS 2017 image database for 100 epochs using data augmentation via elastic transformation. The achieved segmentation accuracy was comparable and slightly better than that achieved using the classical U-Net architecture and its robustness was improved, reducing segmentation artifacts on a large number of datasets. Moreover, our proposed, modified architecture was also more computationally efficient than the traditional transposed convolutional U-net approach described in [6].

B. Imaging Data

All experiments reported in this paper were conducted on the BRATS 2017 image database [6], [8]. This imaging repository consists of MRI scans from 285 glioma patients. Each patient datasets contains a T1-weighted, T2-weighted, FLAIR and a post-Gd T1c-weighted image. Of all 285 patient image datasets, 210 were acquired from high-grade (anaplastic astrocytomas and glioblastoma multiform tumors) and the remaining 75 images featured low-grade (histologically diagnosed astrocytomas or oligoastrocytomas) glioma patients. All images had been manually annotated by expert neuroradiologists according to four different tumor labels and served as ground truth segmentations against which the output of our algorithm was testes: Label 1 - background; Label 2 - necrotic, non-enhancing tumor; Label 3 - edema; and Label 4 - enhancing tumor.

In addition, for testing purposes, we also used images from 146 patients featuring brain tumors of unknown grade available from the same MICCAI 2017 Challenge on Multi-modal Brain Tumor Segmentation [8], [9]; an example of these is shown in Figure 3.

C. Proposed Model

Figure 4 shows a schematic diagram of our proposed NNRET U-net deep convolution neural network. We replaced the fully connected layers with a down-sampling layer along with max-pooling layers. Every two layers form a block and the encoding path consists of 5 convolutional blocks. The network is to a large extent similar to the approach presented in [6]. The convolution layers in all the blocks detect the local features from the previous layers and map their appearance to a feature map. The kernel size and the stride were both chosen such that the filter size can be divided by the stride. For a kernel size of 3×3 and a stride of 1 in both directions coupled with a ReLU activation function, the stride will move the filters

one pixel at a time. This down-sampling path decreases the feature size while increasing the number of feature maps from 1 to 1024.

To decrease computational cost, we reduced the input volume size from 240×240 into 15×15, further down-sampled via a $2 \times 2 \times$ then max-pooling operation (i.e., inserting pooling layer in-between successive convolutional layers), resulting in low-resolution feature maps. To decrease the spatial size of the image, we used the max pooling of size 2×2 with a stride of 2 that will move the filter 2 pixels at a time, resulting in 75% fewer activations. To avoid the cropping operation during the concatenation of feature maps, we used zero padding on every convolution layer.

For the up-sampling, we used the following modification to reconstruct the high resolution feature maps. The decoder consists of five blocks, but despite using the transposed convolution as in [8], we used nearest neighbor up-sampling layer with the scale factor of 2 at the beginning of each block, followed by two convolution layers and ReLU function that increased the spatial dimension in each block by a factor of 2.

The nearest neighbor up-sampling works like convolution that performs a mathematical operation on every pixel and its neighbors by using interpolation to increase the spatial dimension of an image. The process being the nearest neighbor up-sampling to increase image resolution is shown in Figure 4. After locating the center pixel of a cell on the input raster that corresponds to the output raster dataset, the location of the nearest center of the cell on the input raster will be determined, and that value will then be assigned to the corresponding cell on the output raster. As an example, here we explain how a 4×4 pixel image would be up-sampled by NN interpolation method. The cell centers of the output raster are equally separated and a location value needs to be determined from the input raster for each output cell. The nearest neighbor algorithm selects select those cells centers from the input raster that are closest to that of output raster. The black areas of the image can be filled either with the copies of the center pixel or the weighted combinations of the surrounding pixels.

The second step is of the algorithm is similar to the one described in [10], [11]. For the sake of completeness, we reported the receptive fields of each network layer inside the U-net architecture in Table 1.

D. Network Training

A critical component of any successful deep learning model is having sufficient, good quality data to train the classifier and overcome the risk of the model over-fitting the data. To train our model, we used the BRATS 2017 database of brain images, which we further augmented using an elastic transformation that includes both affine and non-affine transformations. Rather than a fixed displacement fields, we generated random displacement fields (Eqn. 3):

$$x(x, y) = y(x, y) = \ rand(-1, +1) \quad (1)$$

where, x and y are convolved with an intermediate value of (in pixels) and the fields are multiplied by a scaling factor α that controls intensity. Thus, we obtain the elastically transformed image in which the global shape of the tumor is undistorted, unlike in the affine-transformed image.

In addition, we also used several other transformations to augment the data specifically image scaling, translation, rotation, shear, and addition of salt and pepper noise as data augmentation, with the overall objective to improve the robustness of the model.

III. RESULTS

In this work, we used the Dice similarity coefficient which is an overlap index that quantified the agreement between two segmented image regions a ground truth segmentation and the tested segmentation method. The DSC coefficient assesses the similarity between the ground truth and the predicted output and reports it in the form of a coefficient between 0 and 1 (Eq'n. 1). The higher the DSC coefficient, the higher the similarity and the more accurate the segmentation output.

$$DSC = \frac{2.|Tr_1 \cap Pr_1|}{|Tr_1| + |Pr_1|} \quad (2)$$

In a sense, the DSC similarity coefficient is similar to the Intersection over Union (IoU) metric addition, which is also a measure of region overlap and agreement between two different segmentation results (Eq'n. 2):

$$IoU = \frac{Tr_1 \cap Pr_1}{Tr_1 \cup Pr_1} \quad (3)$$

where "//" refers to the sum of all the segmented areas, Tr_1 is the ground truth segmented tumor region and Pr_1 is the tumor region segmented by our proposed algorithm.

The proposed modified architecture yielded higher DSC than the IoU similarity metric. This model was implemented in Tensorflow with tensorlayers and was trained for a total of 100 epochs. The hyper-parameters used were as follows: a batch size setting of 10 MRI volumes with a learning rate of 0.0001 for a total of 100 epochs. Each epoch consisted of 6 sub-steps (k-fold cross validation). After completing 6 sub-steps, the algorithm computes the average DCS score across all training cases.

Each epoch training required 1.9 to 2.5 hrs on a server equipped with NVidia Titan GPUs. We validated out segmented results on three tumor sub-regions for each patient dataset. The classification of a complete tumor was deemed successful if all four labels were detected. A correctly classified core tumor consisted of three labels, except edema, and similarly, a correctly classified enhancing tumor has only label 4. For each region, segmentation result

was evaluated. The results after the first three epochs, along with the 15th epoch are shown in Table 2, which summarizes the output of the algorithm in terms of DSC coefficient, Intersection over Union (IoU) and computing time for the first few training epochs.

We tested the NNRET U-Net on 30 MRI volumes with available ground truth annotations and evaluated the performance of our method according to the DSC & IoU metrics. The results are summarized in Table 3, which shows comparable and slightly better results than the classical U-Net implementation as quantified by the DSC for both high- and low-graded glioma tumors.

To visualize the segmentation results by our proposed model, we randomly selected two MRI sequences from the testing dataset with unknown ground truth. Figure 5 shows two image datasets a HGG and LGG. Each row shows the image dataset, ground truth segmentation, and the result of our proposed segmentation algorithm in axial, coronal and sagittal views. The yellow, cyan and light green colors correspond to edema, enhancing and non-enhancing tumor cores, respectively. Note that the model was trained using only axial slices. Note that while the HGG case features enhancing regions, the LGG case does not. Moreover, the model accurately depicts and differentiates between endemic, non-enhancing and enhancing tumor regions with minimal error.

IV. SUMMARY AND CONCLUSION

In this work we described a first, preliminary implementation of a modified U-Net framework for brain tumor tissue segmentation and characterization and evaluated its performance using the MICCAI 2017 Multimodal Brain Tumor Segmentation dataset. To improve robustness beyond that of the classical U-Net framework, we substituted the deconvolution layer with an up-sampling layer that uses nearest-neighbor followed by two convolution layers and augmentation via elastic transformation. We trained the network on the well-established BRATS 2017 database featuring both high- and low-graded glioma tumors. Our proposed algorithm achieved a mean DSC score of 0.8976 and a mean IoU score of 0.8869 assessed against the ground truth annotations, hence outperforming the traditional U-Net segmentation results.

Future work will focus on using kernel-based optimization to reduce training time, as well as a more thorough validation of the algorithm. According to these preliminary results, upon further refinement, this method has the potential to evolve into a robust patient-specific bran tissue characterization tool. The output of this technique can serve as a first order tissue characterization that the clinician can then refine for improved results and more accurate diagnosis.

REFERENCES

[1]. Kohler BA, Ward E, McCarthy BJ, Schymura MJ, Ries LA, Eheman C, Jemal A, Anderson RN, Ajani UA, and Edwards BK, "Annual report to the nation on the status of cancer, 1975–2007, featuring tumors of the brain and other nervous system," Journal of the national cancer institute, vol. 103, no. 9, pp. 714–736, 2011. [PubMed: 21454908]

[2]. Soltaninejad M, Yang G, Lambrou T, Allinson N, Jones TL, Barrick TR, Howe FA, and Ye X, "Automated brain tumour detection and segmentation using superpixel-based extremely

randomized trees in flair mri," International journal of computer assisted radiology and surgery, vol. 12, no. 2, pp. 183–203, 2017. [PubMed: 27651330]

[3]. Ciresan D, Giusti A, Gambardella LM, and Schmidhuber J, "Deep neural networks segment neuronal membranes in electron microscopy images," in Advances in neural information processing systems, 2012, pp. 2843–2851.

[4]. Isensee F, Kickingereder P, Wick W, Bendszus M, and Maier-Hein KH, "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," in International MICCAI Brainlesion Workshop. Springer, 2017, pp. 287–297.

[5]. Ronneberger O, Fischer P, and Brox T, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

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

[7]. Radford A, Metz L, and Chintala S, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[8]. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, and Davatzikos C, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," Scientific data, vol. 4, p. 170117, 2017.

[9]. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, and Davatzikos C, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection," The Cancer Imaging Archive, 2017.

[10]. Dong C, Loy CC, and Tang X, "Accelerating the super-resolution convolutional neural network," in European Conference on Computer Vision. Springer, 2016, pp. 391–407.

[11]. Dong C, Loy CC, He K, and Tang X, "Image super-resolution using deep convolutional networks," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 2, pp. 295–307, 2016. [PubMed: 26761735]
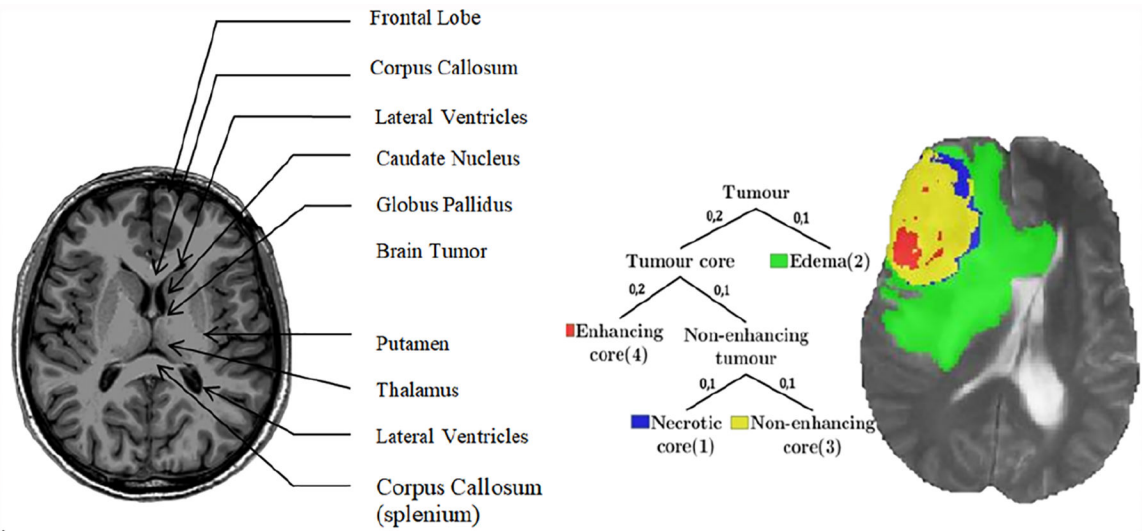
Fig. 1.
Example of the structure of the human brain depicted in a MRI image: axial MRI slice of a normal brain (left) and a segmented tumor with labels as the ground truth (right). Image courtesy of BRATS database.

a) The checkerboard pattern in two dimensions

b) Uneven overlap pattern formation

Deconvolution in last two layers. *Artifacts prior to any training.*

Deconvolution only in last layer. *Artifacts prior to any training.*

All layers use resize-convolution. *No artifacts before or after training.*

c) Artifacts due to Transposed Convolution

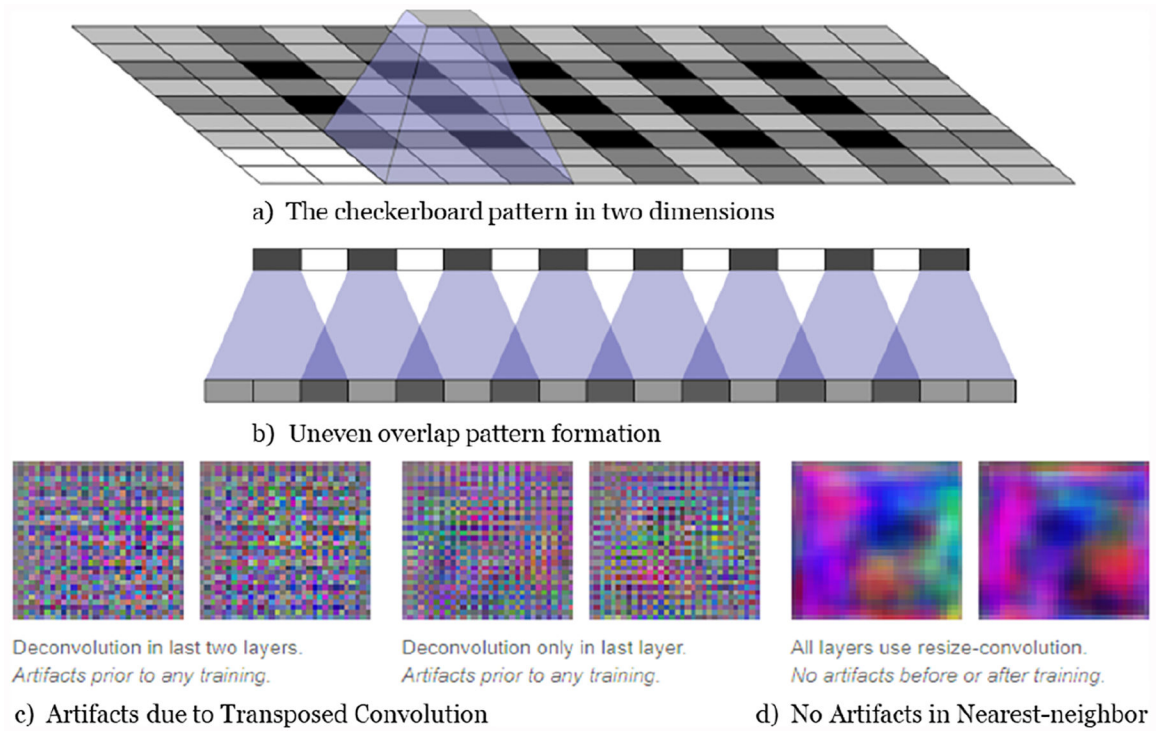d) No Artifacts in Nearest-neighbor

Fig. 2.
Schematic diagram illustrating an artifact caused by the transposed convolution operation: a) Checkerboard problem caused by applying a transposed convolution on images of improper resolution (a) resulting in uneven overlap (b), and artifacts (c) that can be minimized and essentially eliminated by applying a nearest-neighbor interpolation up-sampling operation (d).
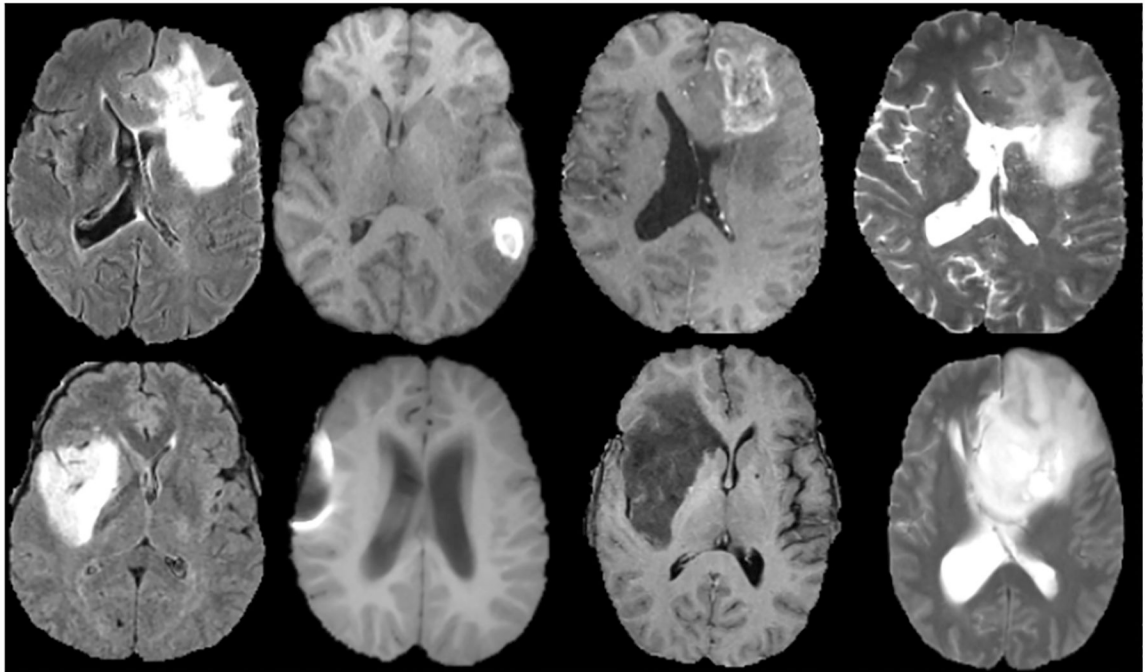
Fig. 3.
Examples of brain images from the BRATS 2017 database: high grade glioma images -
FLAIR, T1-weighted, T1c-weighted and T2-weighted (top row) and corresponding images
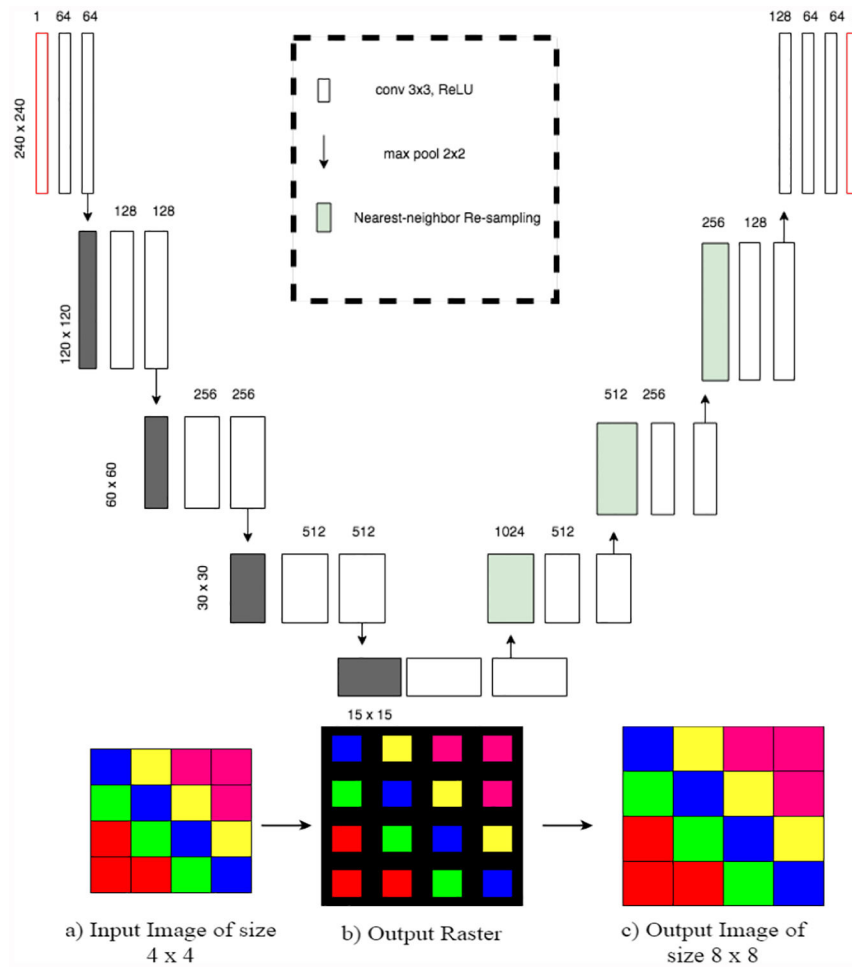for low-grade glioma, respectively (bottom row).

Fig. 4.
Schematic diagram of proposed NNRET U-net architecture (the concatenation process is not shown in the image) and nearest neighbor up-sampling for image enlargement.
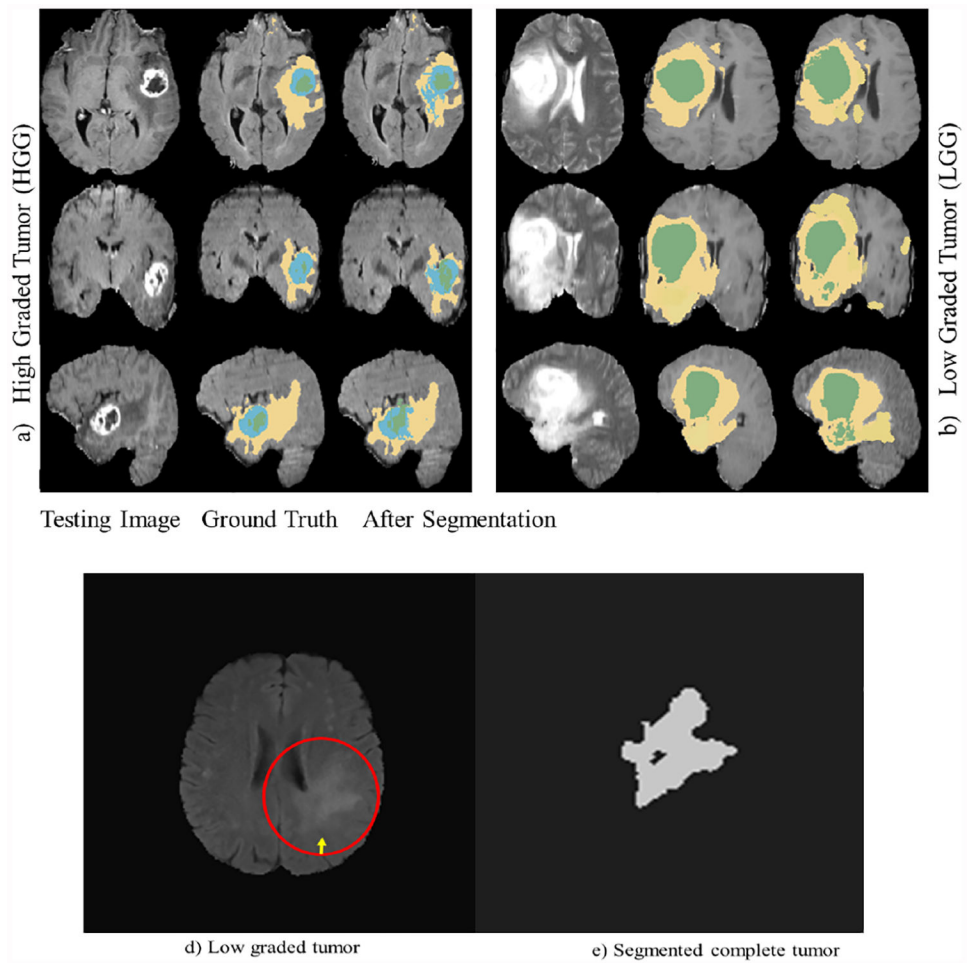
Fig. 5.
Segmentation result by NNRET U-net model. a) high-graded glioma (HGG); b) low-graded glioma (LGG) images. Left column shows the raw imaging data, middle column shows the ground truth segmentation, and the right column shows the result of the proposed segmentation method. Each row corresponds to the axial, coronal and sagittal slice (top to bottom), c) low graded tumor (pointed using red circle), d) segment complete tumor region for LGG precisely.

TABLE I

Down-sampling and up-sampling layers inside U-net architecture. The deconvolution layers in the U-net classic architecture were replaced by the nearest neighbor up-sampling layers. These up-sampling layers are similar to the deconvolution layers, but the up-sampled image are smoother and have a higher resolution than their precursors.

Layers	Type	Kernel		No. of Feature map
Layer 1	input			4
Layer 2	Conv1	3×3		64
Layer 3	conv1	3×3		64
Maxpool Layerl	pool1	2×2		
Layer 4	conv2	3×3		128
Layer 5	conv2	3×3		128
Maxpool Layer2	pool2	2×2		
Layer 6	conv3	3×3		256
Layer 7	conv3	3×3		256
Maxpool Layer3	pool3	2×2		
Layer 8	conv4	3×3		512
Layer 9	conv4	3×3		512
Maxpool Layer4	pool4	2×2		
Layer 10	conv5	3×3		1024
Layer 11	conv5	3×3		1024
NN Upsample	Up4	scale factor=2		512
Concatenate				
Layer 12	conv4	3×3		512
Layer 13	Conv4	3×3		512
NN Upsample	up3	scale factor=2		256
Concatenate				
Layer 12	conv3	3×3		256
Layer 13	conv3	3×3		256
NN Upsample	up2	scale factor=2		128
Concatenate				
Layer 12	conv2	3×3		128
Layer 13	conv2	3×3		128
NN Upsample	Up1	scale factor=2		64
Concatenate				
Layer 12	conv1	3×3		64
Layer 13	conv1	3×3		64
Sigmoid	conv1	1×1		1

TABLE II

Training accuracy after the first three epochs: every epoch consists of 6 sub-steps and upon the completion of each epoch, the mean DSC, IoU and computing time are computed. The training accuracy increased to 91% at convergence after the completion of 15 epochs.

Epoch	Dice	IoU		Time (s)
1	0.0433	0.5588		10.600
1	0.1918	0.6223		10.442
1	0.1388	0.7349		10.385
1	0.0961	0.5256		10.463
1	0.1606	0.8142		10.513
1	0.8084	0.7338		10.385
1/100	0.1937	0.6931		6817.6
2	0.7219	0.5447		10.516
2	0.4983	0.8910		10.516
2	0.3646	0.7633		10.462
2	0.5008	0.7498		10.499
2	0.9915	0.7739		10.709
2	0.4068	0.0000		10.519
2/100	0.5629	0.7162		6860.7
3	0.5019	0.6125		10.471
3	0.3489	0.8541		10.543
3	0.8318	0.9049		10.535
3	0.5502	1.0000		10.574
3	0.6907	0.7348		10.721
3	0.3784	0.7411		10.511
3/100	0.5532	0.7342		6880.5
…	…	…		…
100/100	0.9104	0.9075		6834.3

TABLE III

Assessment of proposed segmentation technique against the classical U-Net implementation on complete tumor region using DSC and IoU for both high-and low-graded glioma tumors.

Algorithm	Tumor grade	DSC		IoU
NNRET	HGG	0.8976		0.8869
NNRET	LGG	0.8459		0.8263
	Combined	0.8717		0.8566
Classic U-net	HGG	0.88		-
Classic U-net	LGG	0.84		-
	Combined	0.86		-