




Deep Convolutional Neural Networks Using U-Net for Automatic Brain Tumor Segmentation in Multimodal MRI Volumes

Adel Kermi¹ , Issam Mahmoudi¹, and Mohamed Tarek Khadir²

¹ LMCS Laboratory, National Higher School of Computer Sciences (ESI),
BP.68M, 16309 Oued-Smar, El-Harrach, Algiers, Algeria
{a_kermi, di_mahmoudi}@esi.dz

² LabGed Laboratory, Department of Computer Sciences,
University Badji-Mokhtar of Annaba, BP.12, 23000 Annaba, Algeria
khadir@labged.net

Abstract. Precise 3D computerized segmentation of brain tumors remains, until nowadays, a challenging process due to the variety of the possible shapes, locations and image intensities of various tumors types. This paper presents a fully automated and efficient brain tumor segmentation method based on 2D Deep Convolutional Neural Networks (DNNs) which automatically extracts the whole tumor and intra-tumor regions, including enhancing tumor, edema and necrosis, from pre-operative multimodal 3D-MRI. The network architecture was inspired by U-net and has been modified to increase brain tumor segmentation performance. Among applied modifications, Weighted Cross Entropy (WCE) and Generalized Dice Loss (GDL) were employed as a loss function to address the class imbalance problem in the brain tumor data. The proposed segmentation system has been tested and evaluated on both, BraTS'2018 training and validation datasets, which include a total of 351 multimodal MRI volumes of different patients with HGG and LGG tumors representing different shapes, giving promising and objective results close to manual segmentation performances obtained by experienced neuro-radiologists. On the challenge validation dataset, our system achieved a mean enhancing tumor, whole tumor, and tumor core dice score of 0.783, 0.868 and 0.805 respectively. Other quantitative and qualitative evaluations are presented and discussed along the paper.

Keywords: Brain tumor segmentation · 3D-MRI · Machine learning · Deep learning · Convolutional Neural Networks · U-net · BraTS'2018 challenge

1 Introduction

Brain tumor segmentation in multimodal Magnetic Resonance Imaging (MRI) is widely used as a vital process for surgical planning and simulation, treatment planning prior to radiation therapy, therapy evaluation [1–5], and intra-operative neuro navigation and image neurosurgery [6–8]. However, segmenting brain tumor manually is

not only a challenging task, but also a time-consuming one, favoring therefore, the emergence of computerized approaches.

Despite considerable research works and encouraging results in the medical imaging domain, fast and precise 3D computerized brain tumors segmentation remains until now a challenging process and a very difficult task to achieve because brain tumors may appear in different size, shape, location and image intensity [2–5]. Many recent research adopted deep learning methods [9], specifically Convolutional Neural Networks (CNNs) [9–12]. CNN shown their effectiveness and were proved successful to automatically classify the normal and pathological brain MRI scans in the past few BraTS challenges as well as other semantic and medical segmentation problems.

This paper proposes an automated and efficient segmentation method of whole tumor and intra-tumor structures, including enhancing tumor, edema and necrosis, in multimodal 3D-MRI. It is based on 2D Deep Convolutional Neural Networks (DNNs) using a modified U-net architecture [10]. The proposed DNN model is trained to segment both High Grade Glioma (HGG) and Lower Grade Glioma (LGG) volumes.

The rest of the paper is organized as follows. First, Sect. 2 presents an overview of the proposed segmentation method. Experimental results with their evaluations are given in Sect. 3. Finally, a conclusion and future work are presented in Sect. 4.

2 The Proposed Method

The proposed segmentation system is entirely automated. The brain tumor segmentation process is based on deep learning more precisely on 2D Convolutional Neural Networks. It includes the main following steps: pre-processing of the 3D-MRI data, training using a U-net architecture, and brain tumoral structures prediction.

2.1 Data and Pre-processing

The BraTS’2018 challenge training dataset consists of 210 pre-operative multimodal MRI scans of subjects with HGG and 75 scans of subjects with LGG, and the BraTS’2018 challenge validation dataset includes 66 different multimodal 3D-MRI [13–16]. Images were acquired at 19 different centers using MR scanners from different vendors and with 3T field strength. They comprise co-registered native (T1) and contrast-enhanced T1-weighted (T1Gd) MRI, as well as T2-weighted (T2) and T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI. All 3D-MRI of BraTS’2018 dataset have a volume dimension of $240 \times 240 \times 155$. They are distributed, co-registered to the same anatomical template and interpolated to the same resolution (1 mm^3). All MRI volumes have been segmented manually, by one to four raters, and their annotations were approved by experienced neuro-radiologists. Each tumor was segmented into edema, necrosis and non-enhancing tumor and active/enhancing tumor.

First, a minimal pre-processing of MRI data is applied, as in [11]. The 1% highest and lowest intensities were removed, then each modality of MR images was normalized by subtracting the mean and dividing by the standard deviation of the intensities within the slice. To address the class imbalance problem in the data, data augmentation technique [17] were employed. This consists in adding new synthetic images by

performing operations and transformations on data and the corresponding manual tumors segmentation images obtained by human experts (i.e., ground truth data). The transformations comprise rotation, translation, and horizontal flipping and mirroring.

2.2 Network Architecture and Training

The CNN used in this study has a similar architecture as that of U-net [10]. Our network architecture can be seen in Fig. 1. It consists of a contracting path (left side) and an expanding path (right side). The contracting path consists of 3 pre-activated residual blocks, as in [18, 19], instead of plain blocks in the original U-net. Each block has two convolution units each of which comprises a Batch Normalization (BN) layer, an activation function, called Parametric Rectified Linear Unit (PReLU) [20], instead of ReLU function used in the original architecture [10], and a convolutional layer, like in [12], instead of using Maxpooling [10], with Padding = 2, Stride = 1 and a 3×3 size filter. For down sampling, a convolution layer with a 2×2 filter and a stride of 2 is applied. At each down sampling step, the number of feature channels is doubled. The contracting path is followed by a fourth residual unit that acts as a bridge to connect both paths. In the same way, the expanding path is built using 3 residual blocks. Prior to each block, there is an upsampling operation which increases the feature map size by 2, followed by a 2×2 convolution and a concatenation with the feature maps corresponding to the contracting path. In the last layer of the expanding path, a 1×1 convolution with the Softmax activation function is used to map the multi-channel feature maps to the desired number of classes.

In total, the proposed network model contains 7 residual blocks, 25 convolution layers, 15 layers of BN and 10159748 parameters to optimize.

The designed network was trained with axial slices extracted from training MRI set, including HGG and LGG cases, and the corresponding ground truth segmentations. The goal is to find the network parameters (weights and biases) that minimize a loss function. In this work, this can be achieved by using Stochastic Gradient Descent algorithm (SGD) [17], at each iteration SGD updates the parameters towards the opposite direction of the gradients. In our network model, we used a loss function that adds Weighted Cross Entropy (WCE) [17] and Generalized Dice (GDL) [21] to address the class imbalance problem present in brain tumor data. So, the two components of the loss function are defined as follows:

$$WCE = -\frac{1}{K} \sum_k \sum_i^L W_i g_{ik} \log(p_{ik}) \quad (1)$$

$$GDL = 1 - 2 \frac{\sum_i^L w_i \sum_k g_{ik} p_{ik}}{\sum_i^L w_i \sum_k (g_{ik} + p_{ik})} \quad (2)$$

where L is the total number of labels, K denotes the batch size. w_i represents the weight assigned to the i th label. As in [21], we set w_i to $\frac{1}{(\sum_k g_{ik})^2} \cdot p_{ik}$ and g_{ik} representing the value of the (i th, k th) pixel of the segmented binary image and of the binary ground truth image, respectively.

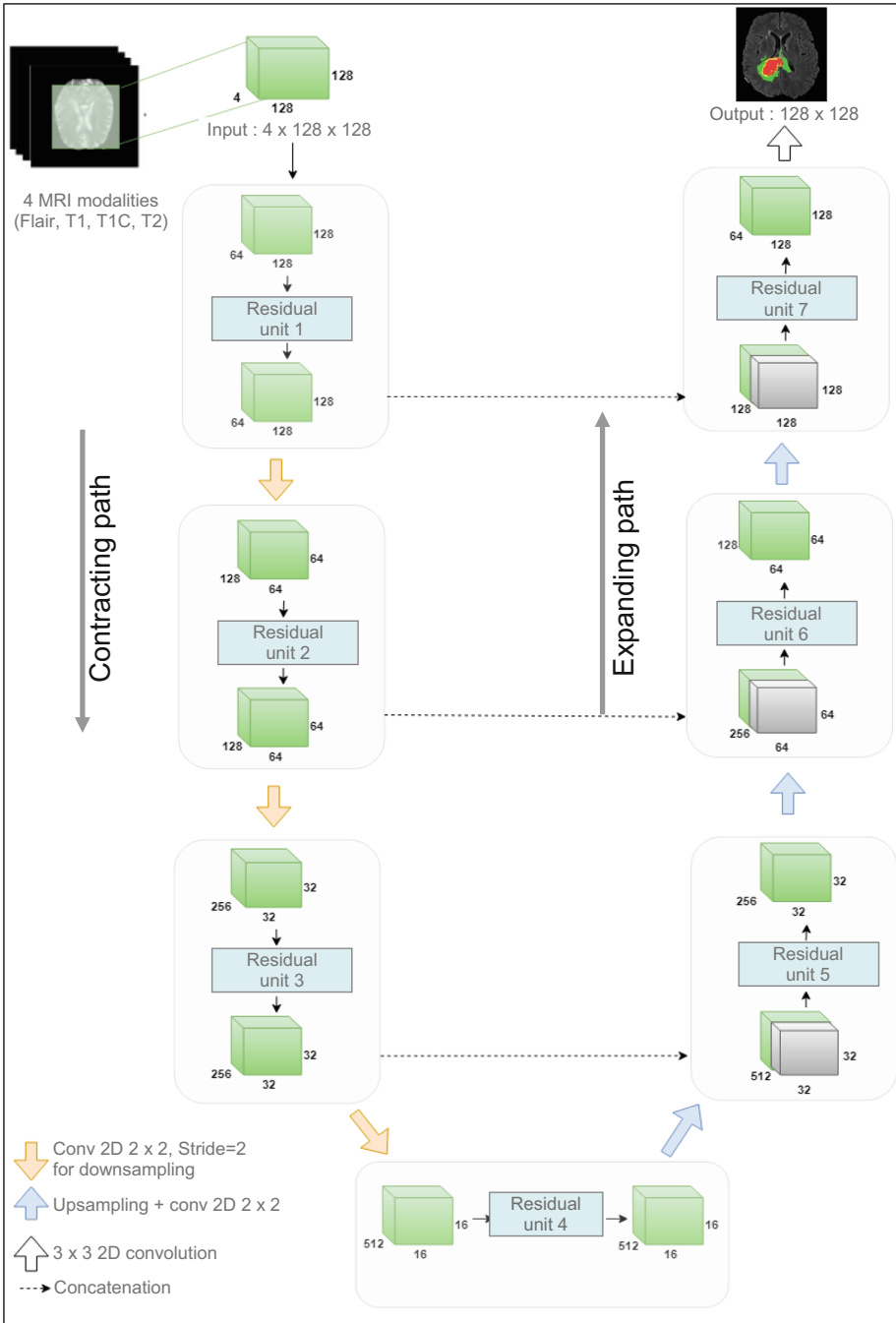


Fig. 1. Architecture of the proposed Deep Convolutional Neural Network.

2.3 Brain Tumoral Structures Prediction

After network training, prediction may be performed. This step consists to provide the network with the four MRI modalities of an unsegmented volume that it has never processed or encountered before, and it must be able to return a segmented image.

3 Experimental Results and Discussion

In this study, we have tested and evaluated our segmentation system on pre-operative multimodal MRI scans of both the training/testing and the validation datasets of the BraTS'2018 challenge [22]. The results of automatically segmented tumors, denoted by A , can be compared with manually segmented tumors by human experts, denoted by B , which are considered as ground truth for evaluation. The results presented in subsequent sections, were obtained during the BraTS'2018 challenge, for training and validation [22]. The top 63 approaches are further compared in terms of results (*Dice*, *Sensitivity* and *Specificity*) and one surface measure based on the Hausdorff distance (*HD*), in [23] on 191 cases. These measures allow to assess the segmentation results accuracy, as well as measuring the similarity between the segmentations A and B [2, 24]. The Dice metric is computed as a performance metric. It measures the similarity between two volumes A and B , corresponding to the output segmentation of the model and clinical ground truth annotations, respectively. The Sensitivity metric measures the proportion of positive voxels of the real brain tumor that are correctly segmented as such, while Specificity metric indicates how well the true negatives are predicted. Employing Sensitivity and Specificity can provide a good assessment of the segmentation result. The *HD* metric indicates the segmentation quality at the tumor's border by evaluating the greatest distance between the two segmentation surfaces A and B , and is independent of the tumor size.

3.1 Performance on 20% of BraTS'2018 Training Dataset (Testing Set)

Preliminary segmentation results for the 285 3D-MRI of the BraTS'2018 training data set have been obtained using 80% of this data set (i.e., 228 subjects) for training and the remaining 20% (i.e., 57 subjects) for validation purposes. Results obtained by our automated system for 10 sample cases are shown in Figs. 2 and 3. Figure 2 shows segmentation results from 5 multimodal MRI where HGG tumors are present and Fig. 3 shows other segmentation results from other 5 MRI with LGG tumors. In these figures, each row represents one clinical case. In the first four columns from left to right, images show one axial slice of MRI acquired in Flair, T1, T1C and T2 modality, respectively, used as input channels to our CNN model. In the fifth and the sixth columns, images show the ground truth (GT) and the prediction labels respectively, where we can distinguish intra-tumoral regions by color-code: enhancing tumor (yellow), peritumoral edema (green) and necrotic and non-enhancing tumor (red). As it can be seen, tumors in the brain MRI of the 10 cases vary in size, shape, position and intensity. By visual inspection, the proposed method usually generates segmentations (Prediction) sensibly similar to the ones obtained by the experts (GT).

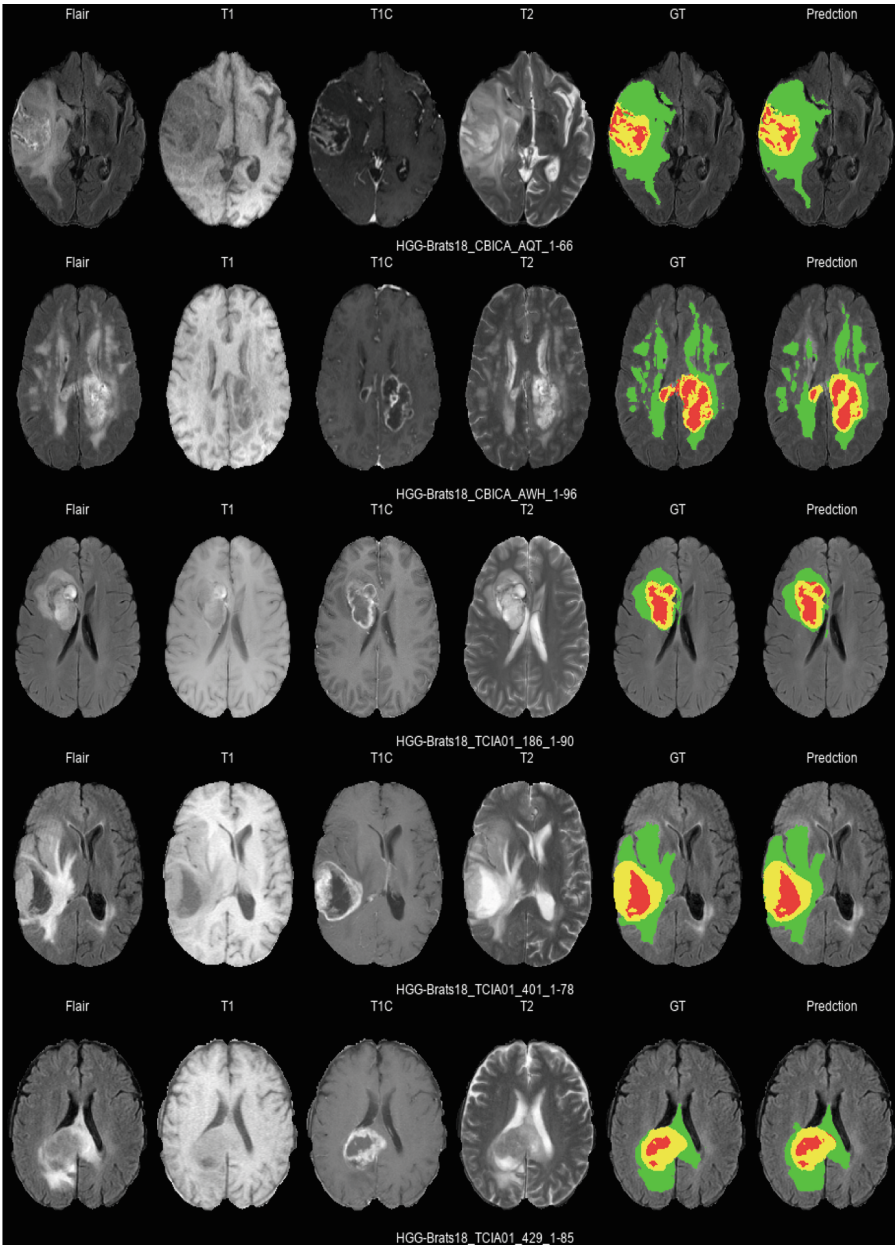


Fig. 2. Intra-tumoral structures segmentation results from 5 multimodal 3D-MRI with HGG of BraTS'2018 training dataset corresponding to 5 different subjects. (Color figure online)

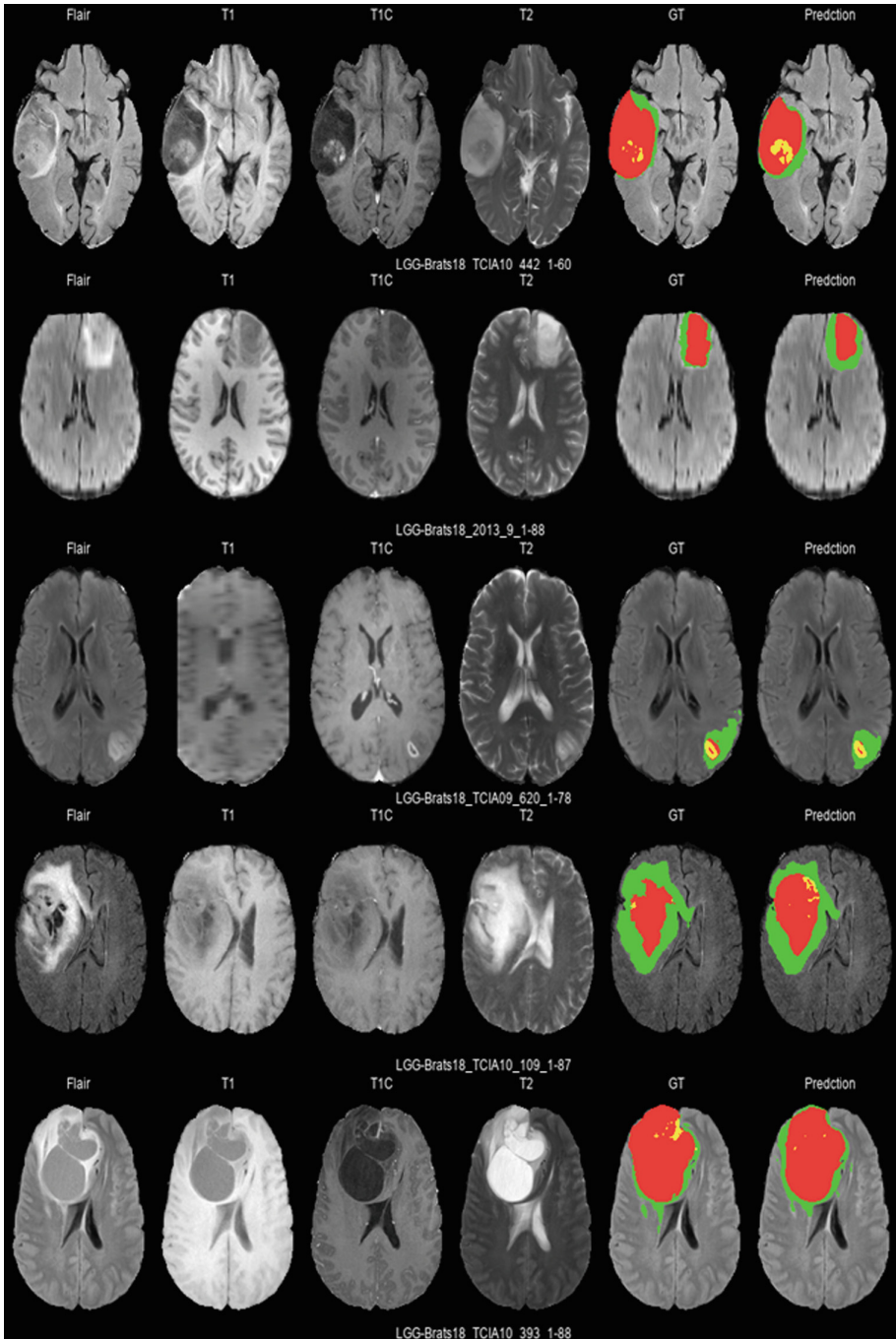


Fig. 3. Intra-tumoral structures segmentation results from 5 other multimodal 3D-MRI with LGG of BraTS'2018 training dataset corresponding to 5 different subjects. (Color figure online)

A quantitative evaluation of segmentation results of Enhancing Tumor (ET), Whole Tumor (WT) and Tumor Core (TC) using the four previously mentioned metrics is given in Tables 1 and 2. Mean, standard deviation, median and 25th and 75th percentile are given for Dice and Sensitivity metrics in Table 1 and for Specificity and Hausdorff distance in Table 2. Values presented in Table 1 show high performance on the Dice metric for WT region, but lower performance for ET and TC regions. Moreover, the proposed method showed good performances for the segmentation of the three intra-tumoral regions. However, the effectiveness, of the approach, is high-lighted in the case of HGG tumors, when compared with the LGG ones.

Table 1. Quantitative evaluation of segmentation results on 20% of BraTS’2018 training dataset (57 MRI scans) using *Dice* and *Sensitivity* metrics.

	<i>Dice</i>			<i>Sensitivity</i>		
	ET	WT	TC	ET	WT	TC
Mean	0.717	0.867	0.798	0.778	0.907	0.84
Std. Dev.	0.275	0.078	0.226	0.3	0.107	0.223
Median	0.831	0.887	0.889	0.912	0.943	0.927
25 quantile	0.726	0.839	0.808	0.777	0.896	0.847
75 quantile	0.859	0.926	0.935	0.951	0.974	0.961

Table 2. Quantitative evaluation of segmentation results on 20% of BraTS’2018 training dataset (57 MRI scans) using *Specificity* and *Hausdorff distance* metrics.

	<i>Specificity</i>			<i>Hausdorff95</i>		
	ET	WT	ET	WT	ET	WT
Mean	0.999	0.998	0.999	4.742	8.706	6.4
Std. Dev.	0.001	0.001	0.002	2.079	2.822	3.685
Median	1	0.999	1	4.123	8.062	5.099
25 quantile	0.999	0.998	0.999	3	6.442	3.871
75 quantile	1	0.999	1	6.633	10.951	8.303

3.2 Performance on BraTS’2018 Validation Dataset

For our participation to BraTS’2018 competition, 100% of the training, including the previous testing dataset (i.e., 285 subjects) is used for training. The performance on BraTS’2018 validation dataset, which is composed of 66 subjects, is diffused in the challenge leaderboard Web site¹ and presented with more statistics in Tables 3, 4, Figs. 4 and 5. In this context, we can compare the obtained segmentation results with those of other participants. The method achieved a mean ET, WT, and TC dice score of 0.783, 0.868 and 0.805 respectively. These scores are close to those obtained by the top

¹ <https://www.cbica.upenn.edu/BraTS18/lboardValidation.html>.

performing methods. Also, an average *HD* scores of 3.728, 8.127 and 9.84 for ET, WT, and TC, respectively were obtained. In addition, it was observed that our DNN model maintains similar WT scores on both, 20% of BraTS’2018 training/testing set used for validation and the final validation dataset proposed for the competition purposes. However, a slight increase in performance on the validation dataset was observed in the ET and TC compartments. It should be noted that this performance was not obtained by overfitting the validation data (i.e., our DNN model has not previously trained on MRI volumes of BraTS’2018 validation dataset).

Table 3. Quantitative evaluation of segmentation results on BraTS’2018 validation dataset (66 MRI scans) by using *Dice* and *Sensitivity* metrics.

	<i>Dice</i>			<i>Sensitivity</i>		
	ET	WT	TC	ET	WT	TC
Mean	0.783	0.868	0.805	0.826	0.895	0.807
Std. Dev.	0.216	0.101	0.199	0.241	0.149	0.222
Median	0.846	0.898	0.891	0.901	0.955	0.895
25 quantile	0.769	0.855	0.756	0.82	0.901	0.71
75 quantile	0.893	0.919	0.928	0.969	0.971	0.965

Table 4. Quantitative evaluation of segmentation results on BraTS’2018 validation dataset (66 MRI scans) by using *Specificity* and *Hausdorff distance* metrics.

	<i>Specificity</i>			<i>Hausdorff95</i>		
	ET	WT	ET	WT	ET	WT
Mean	0.997	0.991	0.997	3.728	8.127	9.84
Std. Dev.	0.004	0.007	0.003	4.471	10.426	15.385
Median	0.998	0.993	0.998	2.236	4.243	5.431
25 quantile	0.997	0.988	0.997	1.637	3	2.871
75 quantile	0.999	0.996	0.999	3.317	7.778	10.728

This performance can be explained by the fact that the number of learned cases (training dataset) used later for the segmentation of the validation dataset is larger than the one used for the BraTS’2018 segmentation on the training/testing dataset. This represents 285 and 228 cases for both trainings respectively. It is also possible that the slight improvements obtained on the validation dataset, are due to the fact that this latter contains more MRI with HGG tumors than MRI with LGG tumors. Indeed, the segmentation efficiency obtained using the proposed network, is more evident on HGG volumes when compared to LGG ones.

Boxplots showing the dispersion of Dice and Sensitivity scores are represented in Fig. 4 and boxplots of the dispersion of Specificity and HD scores are represented in Fig. 5. In these figures, boxplots show quartile ranges of the scores; whiskers and dots ‘●’ indicate outliers; and ‘x’ indicates the mean score.

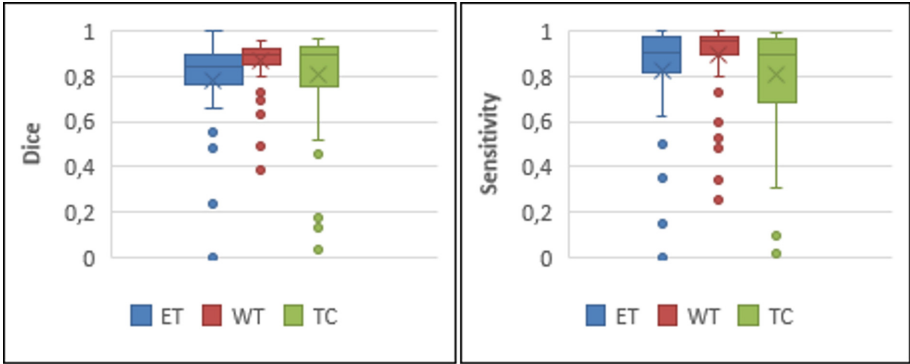


Fig. 4. Dispersion of *Dice* and *Sensitivity* scores for results segmentation of ET, WT, and TC in multimodal MRI scans of the 66 subjects of BraTS’2018 validation dataset.

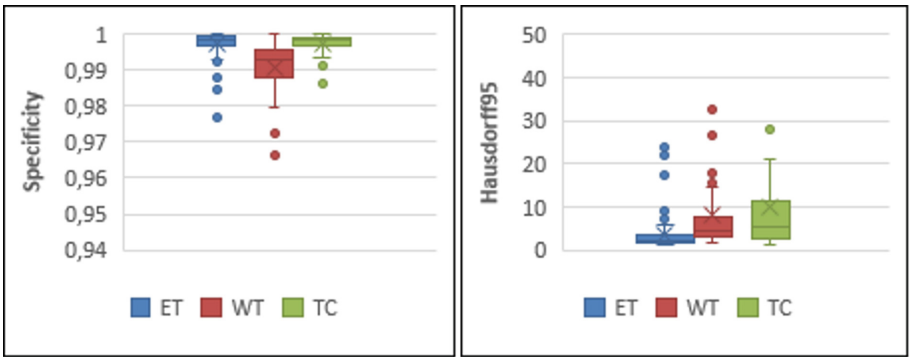


Fig. 5. Dispersion of *Specificity* and *Hausdorff distance* scores for results segmentation of ET, WT, and TC in multimodal MRI scans of the 66 subjects of BraTS’2018 validation dataset.

4 Conclusion and Future Work

In this paper, a fully automatic and accurate method for segmentation of whole brain tumor and intra-tumoral regions using a 2D deep convolutional network based on a well-known architecture in medical imaging called “U-net” is proposed. The constructed DNN model was trained to segment both HGG and LGG volumes.

The proposed method was tested and evaluated quantitatively on both BraTS’2018 training and challenge validation datasets. The total learning computation time of the 285 multimodal MRI volumes of BraTS’2018 training dataset is 185 h on a Cluster machine with Intel Xeon E5-2650 CPU@ 2.00 GHz (64 GB) and NVIDIA Quadro 4000–448 Core CUDA (2 GB) GPU. The average segmentation time of a brain tumor and its components from a given MRI volume is about 62 s on the same GPU. The different tests showed that the segmentation results were very satisfactory, and the evaluation measures confirm that our results are very similar to those manually obtained by the experts, although the proposed method can be further improved.

As future work, a more powerful GPU to further accelerate learning phase of DNN is planned. Thus, a larger number of CNN topologies as well other data augmentation methods may be tested. Also, other interesting perspective consists to use ensemble learning methods, like Stacking and Blending, to improve segmentation performance in tumor core and active tumor regions. Finally, a future work possibility may focus on refining the segmentation results by reducing the false-positive rate using post-processing techniques, such as: applying a conditional random field (CRF) and removing small connected components.

Acknowledgments. This work was in part financially supported by an Algerian research project (CNEPRU) funded by the Ministry of Higher Education and Scientific Research (Project title and number: “PERFORM”, B*04120140014). We would like to thank Bakas, S., Ph.D. and Post-doctoral researcher at SBIA of Perelman School of Medicine University of Pennsylvania – USA for providing the entire BraTS’2018 datasets employed in this study. We also gratefully acknowledge the support of CERIST, the Algerian Research Center for Scientific and Technical Information, for allowing us the use of “IBNBADIS” Cluster, without which we could not have performed these tests.

References

1. Dass, R., Priyanka, Devi, S.: Image segmentation techniques. *Int. J. Electron. Commun. Technol. (IJECT)* **3**, 66–70 (2012)
2. Kermi, A., Andjouh, K., Zidane, F.: Fully automated brain tumour segmentation system in 3D-MRI using symmetry analysis of brain and level sets. *IET Image Process.* **12**, 1964–1971 (2018)
3. Khotanlou, H., Colliot, O., Atif, J., Bloch, I.: 3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models. *Fuzzy Sets Syst.* **160**, 1457–1473 (2009)
4. Prastawaa, M., Bullitt, E., Geriga, G.: Simulation of brain tumors in MR images for evaluation of segmentation efficacy. *Med. Image Anal.* **13**, 297–311 (2009)
5. Sharma, N., Aggarwal, L.M.: Automated medical image segmentation techniques. *J. Med. Phys.* **35**, 3–14 (2010)
6. Despotovic, I., Goossens, B., Philips, W.: MRI segmentation of the human brain: challenges, methods, and applications. *Comput. Math. Methods Med.* **2015**, 23 (2015)
7. Wirtz, C.R., et al.: The benefit of neuronavigation for neurosurgery analyzed by its impact on glioblastoma surgery. *Neurol. Res.* **22**, 354–360 (2000)
8. Yanyun, L., Zhijian, S.: Automated brain tumor segmentation in magnetic resonance imaging based on sliding-window technique and symmetry analysis. *Chin. Med. J.* **127**, 462–468 (2014)
9. Işın, A., Direkoğlu, C., Şah, M.: Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Comput. Sci.* **102**, 317–324 (2016)
10. 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
11. Havaei, M., et al.: Brain tumor segmentation with Deep Neural Networks. *Med. Image Anal.* **35**, 18–31 (2017)

12. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016)
13. Bakas, S., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **4**, 170117 (2017)
14. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *The Cancer Imaging Archive* (2017)
15. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *The Cancer Imaging Archive* (2017)
16. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2015)
17. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press, Cambridge (2016)
18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
19. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, Max (eds.) *ECCV 2016*. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
20. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034. IEEE Computer Society (2015)
21. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) *DLMIA/ML-CDS -2017*. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28
22. Kermi, A., Mahmoudi, I., Khadir, M.T.: Brain tumor segmentation in multimodal 3D-MRI of BraTS'2018 datasets using Deep Convolutional Neural Networks. In: *Pre-conference Proceedings of the 7th International MICCAI BraTS'2018 Challenge*, Granada, Spain, pp. 252–263 (2018)
23. 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)
24. Castillo, L.S., Daza, L.A., Rivera, L.C., Arbelàez, P.: Volumetric multimodality neural network for brain tumor segmentation. In: *Proceedings of the 6th MICCAI BraTS Challenge*, Quebec, Canada, pp. 34–41 (2017)