

Deep Convolutional Neural Networks for the Segmentation of Gliomas in Multi-sequence MRI

Sérgio Pereira^{1,2}(✉), Adriano Pinto¹, Victor Alves², and Carlos A. Silva¹

¹ CMEMS-UMinho Research Unit, Guimarães, Portugal
id5692@alunos.uminho.pt, csilva@dei.uminho.pt

² Centro Algoritmi, Universidade do Minho, Braga, Portugal

Abstract. In their most aggressive form, the mortality rate of gliomas is high. Accurate segmentation is important for surgery and treatment planning, as well as for follow-up evaluation. In this paper, we propose to segment brain tumors using a Deep Convolutional Neural Network. Neural Networks are known to suffer from overfitting. To address it, we use Dropout, Leaky Rectifier Linear Units and small convolutional kernels. To segment the High Grade Gliomas and Low Grade Gliomas we trained two different architectures, one for each grade. Using the proposed method it was possible to obtain promising results in the 2015 Multimodal Brain Tumor Segmentation (BraTS) data set, as well as the second position in the on-site challenge.

Keywords: Magnetic Resonance Imaging · Brain tumor · Glioma · Segmentation · Deep learning · Deep Convolutional Neural Network

1 Introduction

Gliomas are brain tumors originated from the glial cells, and can be divided into Low Grade Gliomas (LGG) and High Grade Gliomas (HGG). Although the former are less aggressive, the mortality rate of the later is high [4, 19]. In fact, the most aggressive gliomas are called Glioblastoma Multiforme, with most patients not surviving more than fourteen months, on average, even when under treatment [29]. The accurate segmentation of the tumor and its sub-regions is important for treatment and surgery planning, but also for follow-up evaluation [4, 19].

Over the years, several approaches were proposed for brain tumor segmentation [4, 19]. Some probabilistic methods explicitly model the underlying data [9, 12, 20, 23]. In these approaches, besides the model for the tissue intensities, it is possible to include priors on the neighborhood through Markov Random Field models [20], estimate a tumor atlas at segmentation time [9, 12, 20] and take advantage of biomechanical tumor growth models [9, 12]. Agn et al. [1] used a generative method based on Gaussian Mixture Models and probabilistic atlases, extended with a prior on the tumor shape learned by convolutional Restricted Boltzmann Machines.

Other approaches learn a model directly from the data in a supervised way [3, 14, 17, 22, 27, 30]. In their core, all of these supervised methods have classifiers that learn how to classify each individual voxel into a tissue type, which may result in isolated voxels, or small clusters, misclassified inside another tissue; however, it is possible to regularize the segmentation by taking the neighborhood into account using Conditional Random Fields [3, 14, 17, 18]. Among the classifiers, Random Forests obtained some of the most promising results [17, 27, 30]. Bakas et al. [2] employed a hybrid generative-discriminative approach. The method is semi-automatic, requiring the user to select some seed points in the image. These points will be used in a modified version of Glistr [9] to obtain a first segmentation; then, it is refined with the gradient boosting algorithm. Lastly, a probabilistic refinement based in intensity statistics is used to obtain the final segmentation.

All the previous supervised methods require the computation of hand-crafted features, which may be difficult to design, or require specialized knowledge on the problem. On the other hand, Deep Learning methods automatically extract features [13]. In Convolutional Neural Networks (CNNs), a set of filters is optimized and convolved with the input image to compute certain characteristics; so, CNNs can deal with the raw data directly. Those filters represent weights of the neural network. Since the filters are convolved over the features, the weights are shared across neural units in the resulting feature maps. In this way, the number of weights in these networks is lower than in neural networks constituted by only fully-connected (FC) layers, making them less prone to overfitting [13]. Overfitting can be a severe problem in neural networks; so, Dropout appears as a regularization method that removes nodes of the network according to some probability in each training step, thus enforcing all nodes to learn good features [25]. Some methods employing CNN for brain tumor segmentation were already proposed [8, 10, 15, 28]. Havaei et al. [10] used a complex architecture of parallel branches and two cascaded CNNs; training of the network was accomplished in two stages: first with balanced classes and, then, a refinement of the last layer was accomplished using a number of samples of each class closer to the observed in brain tumors. Lyksborg et al. [15] trained a CNN in each of the three orthogonal planes of the Magnetic Resonance Imaging (MRI) images, using them as an ensemble of networks for segmentation. Dvořák and Menze [8] used CNNs for structured predictions.

Inspired by Simonyan and Zisserman [24], we developed CNN architectures using small 3×3 kernels. In this way, we can have more convolutional layers, with the opportunity to apply more non-linear transformations of the data. Additionally, we use data augmentation to increase the amount of training data and Leaky Rectifier Linear Units (LReLU) as non-linear activation function. This approach and architecture obtained the second position in the 2015 BraTS challenge.

2 Materials and Methods

The processing pipeline has three main stages: pre-processing, classification through CNNs and post-processing; Fig. 1 presents an overview of the proposed method and interactions between the Training and Testing stages.

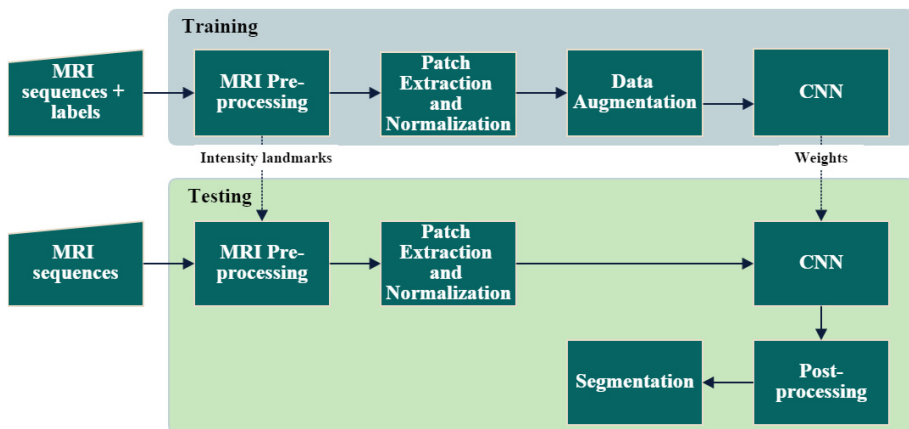


Fig. 1. Overview of the processing pipeline. During training, we artificially augment the data, but at test time we use just the original version of patches.

2.1 Data

BraTS 2015 [11, 19] includes two data sets: Training and Challenge. The Training data set comprises 220 acquisitions from patients with HGG and 54 from patients with LGG. Four MRI sequences are available for each patient: T1-, T1- post contrast (T1c), T2- and FLAIR-weighted. In this data set, the manual segmentations are publicly available. In the Challenge data set, both the manual segmentations and tumor grade are unknown. This set contains 53 subjects with the same MRI sequences as the Training set. All images were already rigidly aligned with the T1c and skull stripped; the resolution was guaranteed to be coherent among all MRI sequences and patients by interpolation of the sequences with thickest slices to $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ voxels.

2.2 Method

Given the differences between HGG and LGG, a model was trained for each grade. Thus, when segmenting a data set where the tumor grade is unknown, we require the user to visually inspect the images and identify the grade beforehand. After this procedure, the remaining pipeline is automatic, without requiring further intervention of the user, for example, to select parameters, seed points or regions of interest.

Pre-processing. The bias field in each MRI sequence was corrected using the N4ITK method [26]. This procedure was similar for all sequences, using 20, 20, 20 and 10 iterations, a shrink factor of 2 and a B-spline fitting distance of 200. After that, the intensities of each individual MRI sequence were normalized [21]. The method for this normalization procedure learns a standardized histogram

with a set of intensity landmarks from the Training set, then, the intensities between two landmarks are linearly transformed to fit in the same landmarks of the standardized histogram; we selected 12 matching landmarks both in LGG and HGG. Finally, the patches are extracted in the axial slices and are normalized to have zero mean and unit variance in each sequence; the mean and variance are calculated for in each sequence using all training patches.

In brain tumor images the classes are highly imbalanced. There are much more samples of normal tissue than tumor tissue; additionally, among the tumor classes there are also classes more common than others, for example, edema represents a bigger volume than necrosis, which may even not exist in some patients. To cope with this, around 40 % of our training samples are extracted from normal tissue, while the remaining 60 % corresponds to brain tumor samples with approximately balanced numbers of samples across classes. However, since some classes are rare, the number of training samples of some tissues must be reduced to keep the classes balanced; so, during training each patch is rotated on the fly (in a parallel process) by 90, 180 and 270 to artificially augment the training data; at test time the patches are not rotated and we classify just the central voxel.

Convolutional Neural Network. In convolutional layers of CNNs the features are extracted by convolving a set of weights, organized as kernels, with the input. These weights are optimized during training to enhance different features of the images. The computation of the i^{th} feature map in layer l (F_i^l) is defined as

$$F_i^l = f \left(b_i^l + \sum_j W_{i,j}^l * X_j^{l-1} \right) \quad (1)$$

where f denotes the activation function, b represents the bias, j indexes the input channel, W denotes the kernels and X^{l-1} the output of the previous layer.

The architectures of the CNNs were developed following [24] and are described in Table 1; several variations were experimented, but these were found to obtain better results in the validation set. By using small kernels, we can stack more layers and have a deeper architecture, while maintaining the same effective receptive field of bigger kernels. For example, two layers with 3×3 filters have the same receptive field of one layer with 5×5 kernels, but we have fewer weights to train and we can apply two non-linear transformations to the data. We trained a deeper architecture for HGG than for LGG; adding more layers to the LGG architecture did not improve results, possibly because of the nature of LGG, such as its lower contrast in the core, when compared to HGG. The input consists in 33×33 axial patches in each of the 4 MRI sequences. Max-pooling consists in downsampling the features maps by only keeping the maximum inside a neighborhood of units in the feature maps; in this way, the computational load of the next layers decrease and small irrelevant details can be discarded. However, segmentation must also detect fine details in the image, thus, in our architectures, max-pooling is performed with some overlapping of the receptive fields, to keep important details

Table 1. Architecture of the CNN for HGG (left) and LGG (right). All non-linearities were LReLU, with the exception of the last FC layer, where softmax was used.

Input: 33 X 33 X 4				Input: 33 X 33 X 4			
Layer	Receptive Field	Stride	# of feat. maps/FC units	Layer	Receptive Field	Stride	# of feat. maps/FC units
Conv.	3X3	1X1	64	Conv.	3X3	1X1	64
Conv.	3X3	1X1	64	Conv.	3X3	1X1	64
Max-Pool.	3X3	2X2	-	Max-Pool.	3X3	2X2	-
Conv.	3X3	1X1	128	Conv.	3X3	1X1	128
Conv.	3X3	1X1	128	Conv.	3X3	1X1	128
Conv.	3X3	1X1	128	Max-Pool.	3X3	2X2	-
Max-Pool.	3X3	2X2	-	FC	-	-	256
FC	-	-	256	FC	-	-	256
FC	-	-	256	FC	-	-	5
FC	-	-	5				

for segmentation. In all the FC layers we use Dropout with $p = 0.5$ as regularization, in order to reduce overfitting. Besides preventing nodes to co-adapt to each other, Dropout works as an extreme case of bagging and ensemble of networks, since in each mini-batch there are different nodes exposed to a small and different portion of the training data [25]. LReLU was the activation function in almost all layers, expressed as

$$f(x) = \max(0, x) + \alpha \min(0, x) \quad (2)$$

where α denotes the leakyness parameter defined as $\alpha = \frac{1}{3}$. Contrasting with ReLU, which imposes a constant 0 in the negative part of the function, LReLU has a small negative slope in that part of the function. This is useful for training, since imposing a constant forces the back-propagated gradient to become 0 in the negative values [16]. The loss function was defined as the Categorical Cross-entropy

$$H = - \sum_{j \in \text{voxels}} \sum_{k \in \text{classes}} c_{j,k} \log(\hat{c}_{j,k}) \quad (3)$$

where \hat{c} denotes the probabilistic predictions (after the softmax activation function) and c denotes the target. Training is accomplished by optimizing the loss function through Stochastic Gradient Descent using Nesterov's Momentum with momentum coefficient of 0.9. The learning rate ϵ was initialized with $\epsilon = 0.003$ and linearly decreased after each epoch during the first 25 epochs until $\epsilon = 0.00003$. All convolutional layers operate over padded inputs to maintain its sizes in the output.

The CNNs were implemented using Theano [5] and Lasagne [7].

Post-processing. A morphological filter was applied to impose volumetric constraints. Consequently, the clusters are identified and we remove those with less than 10,000 voxels in HGG and 3,000 voxels in LGG.

2.3 Evaluation

Although we segment each image into five classes (normal tissue, necrosis, edema, non-enhancing tumor and enhancing tumor), the evaluation appraises three tumor regions: Enhancing tumor, Core (necrosis + non-enhancing tumor + enhancing tumor) and the Complete tumor (all tumor classes). To evaluate the segmentations, the following metrics were computed: Dice Similarity Coefficient (DSC), Positive Predictive Value (PPV), Sensitivity and Robust Hausdorff Distance. The DSC [6] measures the overlap between the manual and the automatic segmentation. It is defined as,

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

where TP, FP and FN denote the numbers of true positive, false positive and false negative detections, respectively. PPV represents the proportion of detected positive results that are really positive and is defined as,

$$PPV = \frac{TP}{TP + FP}. \quad (5)$$

Sensitivity measures the proportion of positive detections that are correctly identified as such and is useful to evaluate the number of true positive and false negative detections, being defined as

$$Sensitivity = \frac{TP}{TP + FN}. \quad (6)$$

The metrics provided by the organizers for the Challenge set were DSC and robust Hausdorff Distance. The Hausdorff Distance measures the distance between the surface of computed (∂P) and manual (∂T) segmentation, as

$$Haus(\partial P, \partial T) = \max\left\{ \sup_{p \in \partial P} \inf_{t \in \partial T} d(p, t), \sup_{t \in \partial T} \inf_{p \in \partial P} d(t, p) \right\} \quad (7)$$

In the robust version of this measure, instead of calculating the maximum distance between the surface of the computed and manual segmentation, it is taken into account the 95 % quantile.

3 Results and Discussion

Some segmentation examples obtained in the Training data set are illustrated in Fig. 2, where we can observe the necrosis, edema, non-enhanced and enhanced tumor classes; quantitative results in the same set are presented in Table 2 and Fig. 3. These results were obtained by 2-fold cross-validation and 3-fold cross-validation in HGG and LGG, respectively. Observing Table 2, metrics in the Core and Enhanced regions of LGG are lower than in HGG, which may be due to the lower contrast of the former. In fact, the contrast in the Core region is lower in LGG [19] than in HGG. Additionally, although brain tumors are

very heterogeneous, LGG tend to be smaller than HGG, with less Core tissues, as observed from the first and third rows of Fig. 2b. Another issue with LGG is the smaller number of training patients, when compared to HGG. From the boxplots in Fig. 3, we can observe the higher dispersion in the Core region of LGG compared to HGG; in the enhanced tumor in LGG the boxplots range almost the full scale of the metrics, possibly because some of these tumors do not possess enhancing tumor. However, the results for the Complete region are similar in

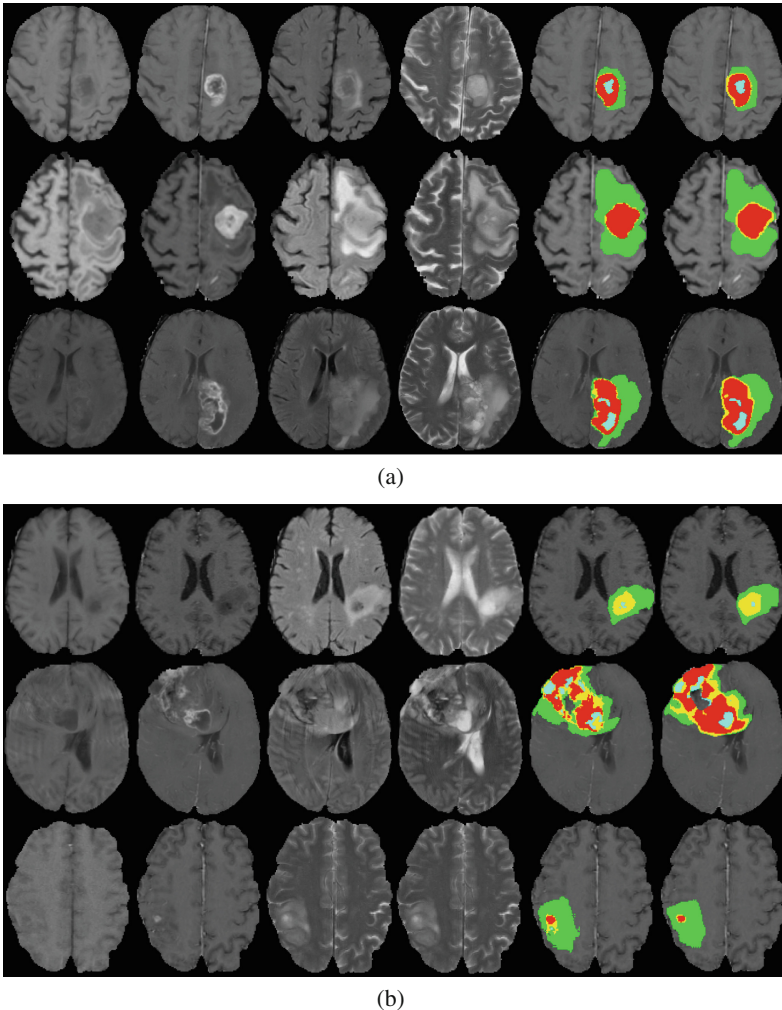


Fig. 2. Segmentation examples on the training data set from (a) HGG and (b) LGG. From left to right: T1, T1c, FLAIR, T2, manual segmentation and obtained segmentation. Colors in the segmentations represent: blue - necrosis, green - edema, yellow - non-enhanced tumor, red - enhanced tumor (Color figure online).

LGG and HGG, with similar dispersion as observed in the boxplots. There are some outliers in Fig. 3, mainly in HGG, which may be due to the high variability of brain tumors and to the bigger amount of patients with HGG. Following the results in Table 2, in Fig. 2 the boundaries of the complete tumor seem well defined, both in LGG and HGG. However, from the second and third rows in

Table 2. Results (mean) obtained with BraTS 2015 training data set.

	DSC			PPV			Sensitivity		
	Complete	Core	Enhanced	Complete	Core	Enhanced	Complete	Core	Enhanced
LGG	0.86	0.64	0.40	0.86	0.67	0.39	0.88	0.71	0.51
HGG	0.87	0.75	0.75	0.89	0.76	0.80	0.86	0.79	0.75
LGG + HGG	0.87	0.73	0.68	0.89	0.74	0.72	0.86	0.77	0.70

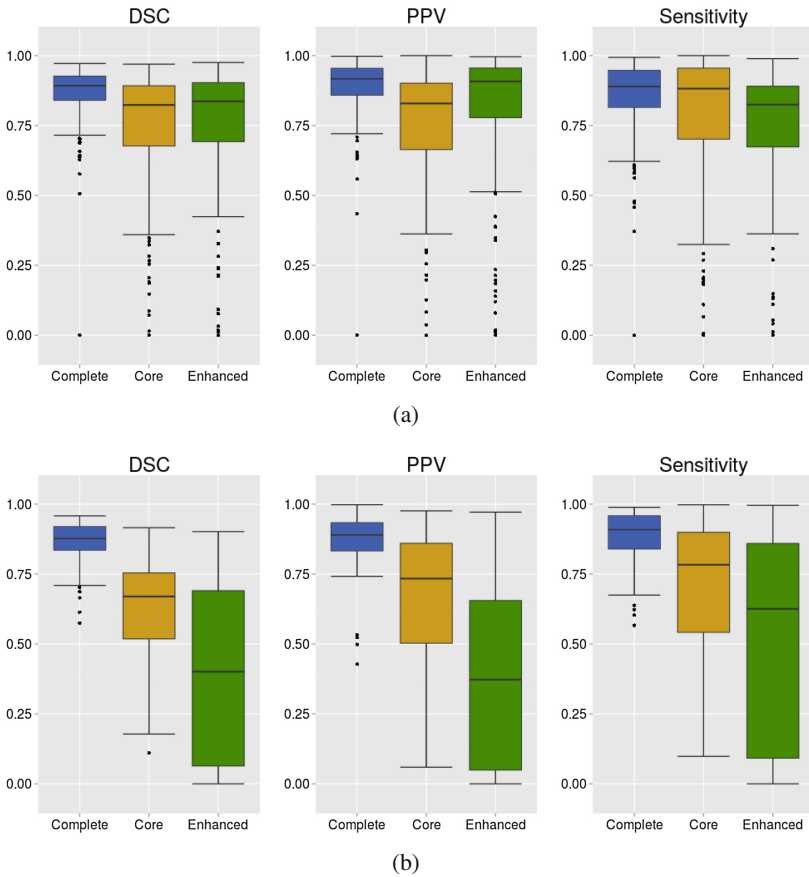


Fig. 3. Boxplot of the results in each of the evaluated brain tumor regions using the Training data set in (a) HGG and (b) LGG; black dots represent outliers

Fig. 2b it seems that we are over-segmenting the Core classes in LGG; nevertheless, the second example looks particularly difficult with a big portion of tumor Core tissues in a very heterogeneous distribution, sharp shapes and details.

Figure 4 presents segmentation examples obtained in the Challenge data set, while Table 3 and Fig. 5 present the quantitative results. In this case, all subjects

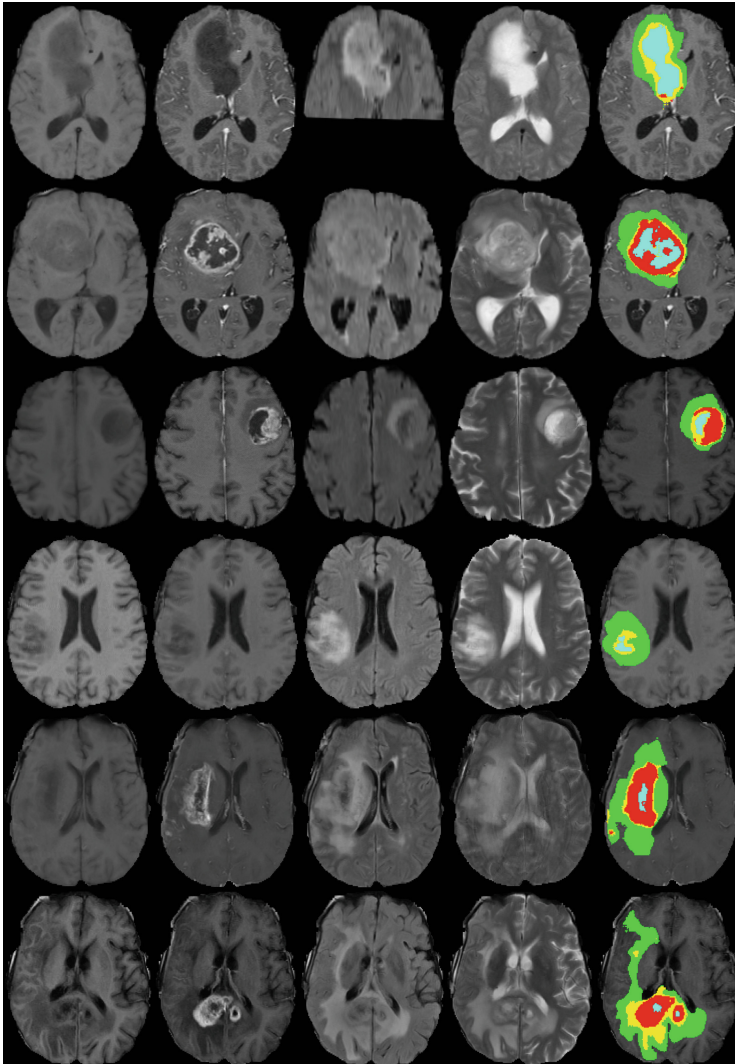
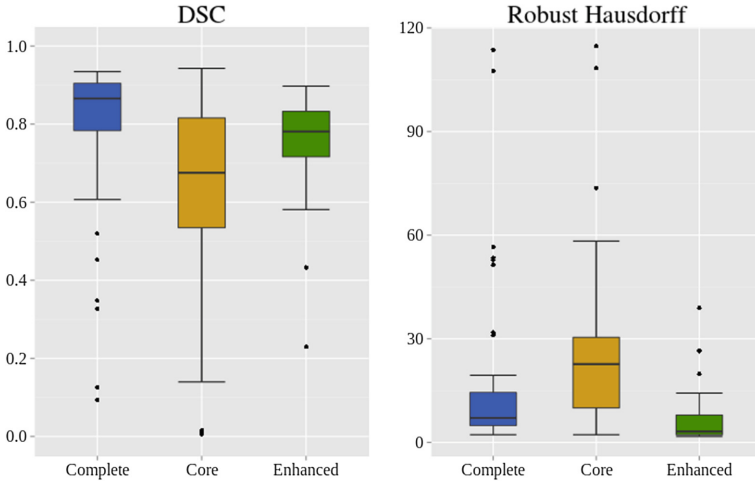


Fig. 4. Segmentation examples on the challenge data set. From left to right: T1, T1c, FLAIR, T2 and obtained segmentation. Colors in the segmentations represent: blue - necrosis, green - edema, yellow - non-enhanced tumor, red - enhanced tumor (Color figure online).

Table 3. Results (mean) using the challenge data set of BraTS 2015.

DSC			Robust Hausdorff		
Complete	Core	Enh.	Complete	Core	Enh.
0.78	0.65	0.75	15.83	26.54	6.99

**Fig. 5.** Boxplots of DSC and Robust Hausdorff Distance obtained using the challenge data set of BraTS 2015.

in each grade of the Training data set were used for training the CNN, with the exception of six validation patients in each grade. To train the CNNs we extracted around 4,000,000 training patches of HGG and 1,800,000 of LGG, and we used mini-batches of 128 training samples. However, the number of training patches was 4 times bigger due to the data augmentation. Observing Fig. 4, the segmentations seem coherent with the expected tumor tissues, for example, the enhanced tumor portions appear delineated following the enhancing parts in T1c. Also, the complete tumor appears to be well delineated, when comparing with the FLAIR and T2 sequences, where the edema is hyperintense.

The training stage of each CNN took around one week. However, the entire processing pipeline takes approximately 8 min to segment each patient, using GPU processing on a Intel Core i7 3.5 GHz CPU, 32 GB of RAM, with a Nvidia Geforce GTX 980 computer running Ubuntu 14.04 OS.

4 Conclusions and Future Work

In this paper, we presented a CNN to segment brain tumors in MRI. Excluding when the user needs to identify the tumor grade, all steps in the processing pipeline are automatic. Although simple, this architecture shows promising

results, with space for further developments, especially in the Core region and segmentation of LGG; in the Challenge data set the proposed method was ranked in the second position. As future work, we want to make a totally grade independent method, possibly through a joint LGG/HGG training or an automatic grade identification procedure before segmentation.

Acknowledgments. This work is supported by FCT with the reference project UID/EEA/04436/2013, by FEDER funds through the COMPETE 2020 Programa Operacional Competitividade e Internacionalização (POCI) with the reference project POCI-01-0145-FEDER-006941. Sérgio Pereira was supported by a scholarship from Fundação para a Ciência e Tecnologia (FCT), Portugal (scholarship number PD/BD/105803/2014). Brain tumor image data used in this article were obtained from the MICCAI 2013 Challenge on Multimodal Brain Tumor Segmentation. The challenge database contain fully anonymized images from the Cancer Imaging Archive.

References

1. Agn, M., Puonti, O., Law, I., af Rosenschöld, P.M., van Leemput, K.: Brain tumor segmentation by a generative model with a prior on tumor shape. In: Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge, pp. 1–4 (2015)
2. Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davatzikos, C.: Segmentation of gliomas in multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework. In: Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge, pp. 5–12 (2015)
3. 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, Part III. LNCS, vol. 6893, pp. 354–361. Springer, Heidelberg (2011)
4. 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)
5. Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., Bengio, Y.: Theano: a CPU and GPU math expression compiler. In: Proceedings of the Python for Scientific Computing Conference (SciPy), June 2010
6. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* **26**(3), 297–302 (1945)
7. Dieleman, S., Schler, J., Raffel, C., Olson, E., Snderby, S.K., Nouri, D., Maturana, D., Thoma, M., Battenberg, E., Kelly, J., Fauw, J.D., Heilman, M., diogo149, McFee, B., Weideman, H., takacsg84, peterderivaz, Jon, instagibbs, Rasul, D.K., CongLiu, Britefury, Degrave, J.: Lasagne: First release, August 2015. <http://dx.doi.org/10.5281/zenodo.27878>
8. Dvorák, P., Menze, B.: Structured prediction with convolutional neural networks for multimodal brain tumor segmentation. In: Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge, pp. 13–24 (2015)
9. Gooya, A., Pohl, K.M., Bilello, M., Cirillo, L., Biros, G., Melhem, E.R., Davatzikos, C.: Glistr: glioma image segmentation and registration. *IEEE Trans. Med. Imaging* **31**(10), 1941–1954 (2012)

10. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. arXiv preprint (2015). [arXiv:1505.03540](https://arxiv.org/abs/1505.03540)
11. Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P.: The virtual skeleton database: an open access repository for biomedical research and collaboration. *J. Med. Internet Res.* **15**(11), e245 (2013). <http://www.jmir.org/2013/11/e245/>
12. Kwon, D., Shinohara, R.T., Akbari, H., Davatzikos, C.: Combining generative models for multifocal glioma segmentation and registration. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part I. LNCS, vol. 8673, pp. 763–770. Springer, Heidelberg (2014)
13. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
14. 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, Part I. LNCS, vol. 5241, pp. 359–366. Springer, Heidelberg (2008)
15. Lyksborg, M., Puonti, O., Agn, M., Larsen, R.: An ensemble of 2D convolutional neural networks for tumor segmentation. In: Paulsen, R.R., Pedersen, K.S. (eds.) SCIA 2015. LNCS, vol. 9127, pp. 201–211. Springer, Heidelberg (2015)
16. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the ICML, vol. 30 (2013)
17. Meier, R., Bauer, S., Slotboom, J., Wiest, R., Reyes, M.: Appearance-and context-sensitive features for brain tumor segmentation. In: BraTS Challenge Manuscripts, pp. 20–26 (2014)
18. Meier, R., Karamitsou, V., Habegger, S., Wiest, R., Reyes, M.: Parameter learning for crf-based tissue segmentation of brain tumors. In: Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge, pp. 48–51 (2015)
19. Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, D., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekaruddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S., Riklin Raviv, T., Reza, S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015)
20. Menze, B.H., van Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., Golland, P.: A generative model for brain tumor segmentation in multi-modal images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 151–159. Springer, Heidelberg (2010)
21. Nyúl, L.G., Udupa, J.K., Zhang, X.: New variants of a method of mri scale standardization. *IEEE Trans. Med. Imaging* **19**(2), 143–150 (2000)
22. Pinto, A., Pereira, S., Correia, H., Oliveira, J., Rasteiro, D.M., Silva, C.A.: Brain tumour segmentation based on extremely randomized forest with high-level features. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3037–3040. IEEE (2015)
23. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. *Med. Image Anal.* **8**(3), 275–283 (2004)

24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint. (2014). [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
25. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
26. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. *IEEE Trans. Med. Imaging* **29**(6), 1310–1320 (2010)
27. Tustison, N.J., Shrinidhi, K., Wintermark, M., Durst, C.R., Kandel, B.M., Gee, J.C., Grossman, M.C., Avants, B.B.: Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ants. *Neuroinformatics* pp. 1–17 (2014)
28. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. In: *MICCAI Brain Tumor Segmentation Challenge (BraTS)*, pp. 1–5 (2014)
29. Van Meir, E.G., Hadjipanayis, C.G., Norden, A.D., Shu, H.K., Wen, P.Y., Olson, J.J.: Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma. *CA Cancer J. Clin.* **60**(3), 166–193 (2010)
30. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O.M., Das, T., Jena, R., Price, S.J.: Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS*, vol. 7512, pp. 369–376. Springer, Heidelberg (2012)