

Brain Tumor Segmentation and Parsing on MRIs using Multiresolution Neural Networks

Laura Silvana Castillo*, Laura Alexandra Daza*, Luis Carlos Rivera* and Pablo Arbeláez

Department of Biomedical Engineering, Universidad de los Andes, Bogotá, Colombia.

Abstract. Brain lesion segmentation is a critical application of computer vision to the biomedical image analysis. The difficulty is derived from the great variance between instances, and the high computational cost of processing three dimensional data. We introduce a neural network for brain tumor semantic segmentation that parses their internal structures and is capable of processing volumetric data from multiple MRI modalities simultaneously. As a result, the method is able to learn from small training datasets. We develop an architecture that has four parallel pathways with residual connections. It receives patches from images with different spatial resolutions and analyzes them independently. The results are then combined using fully-connected layers to obtain a semantic segmentation of the brain tumor. We evaluated our method using the 2017 BraTS Challenge dataset, reaching average dice coefficients of 89%, 88% and 86% over the training, validation and test images, respectively.

Keywords: Semantic segmentation, brain tumors, machine learning, deep learning, MRI.

1 Introduction

Brain tumors are abnormal formations of mass that apply pressure to the surrounding tissues, causing several health problems such as unexplained nausea, seizures, personality changes or even death [1]. They have different shapes, sizes and internal structures, which makes the task of detection and classification difficult and highly dependent on the experience of the specialist, even for experts. These lesions can be classified into Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG). LGGs are benign, slowly growing tumors that can become life-threatening in the course of disease. HGGs are malignant, fast growing tumors capable of inducing the development of new tumors in different parts of the central nervous system. Without an appropriate treatment, HGGs can be lethal in just a few months [2, 3]. Even after a diagnosis had been made there is a high probability that the treatment could not be the best one for that specific case.

* Authors with equal contribution

Nowadays, doctors make use of Magnetic Resonance Imaging (MRI) to visualize the brain of a patient to look for any life-threatening abnormality. However, finding those structures in a 3D medical image is a complicated task, highly prone to error [4]. In spite of the fact that the treatment selection is based directly on the diagnosis, these days that process is made manually, which causes it to be inefficient and observer-dependent. The responsibility to find whether there is an abnormality or not in the exam lies on the neurologist’s hands, who decides, based on his own experience, if there is a lesion and what is the best way to proceed. As a consequence, the uncertainty of the patient’s outcome is significant.

For more than a decade, automatic brain lesion segmentation has been a topic of interest. Initial approaches to solve this problem were based on the detection of abnormalities using healthy-brain atlases and probabilistic models [5]. Later, results were improved using deformable registration fields along with Markov Random Fields (MRF) [6]. Subsequent approaches using machine learning techniques, such as Random Forests [7, 8], yielded better results, reaching an average dice coefficient of 60% in the 2012 BraTS Challenge.

In the last years, Convolutional Neural Networks (CNN) have shown outstanding results in detection, classification and segmentation tasks, matching humans. Part of this success is due to the rapid improvement of machine’s computational power and to the CNNs ability of abstracting features in different hierarchical representations of an image [9]. Fully convolutional networks (FCN) proved to be an effective way to perform pixel-by-pixel classification [10], obtaining a mean Intersection over Union (IoU) of 67% in the PASCAL-VOC dataset in 2012, where the task was to produce an accurate segmentation of 20 different categories in natural images. This method offers the advantage of combining coarse and shallow semantic information from images with an arbitrary input size [10]. In 2015, U-Net, an architecture based on FCN and specialized in the task of segmenting medical images, was developed. U-Net’s architecture has a contracting path to extract local information and an expanding path to locate the object within the whole image [11]. Recently V-Net, an expansion of this method to process three-dimensional data, was presented. V-Net demonstrated a remarkable behavior in the MICCAI 2012 PROMISE Challenge dataset for prostate segmentation in computerized tomography (CT), obtaining an average dice coefficient of 82% [12]. Another method to process medical images is Deepmedic, a neural network that segments brain tumors using information from different MRI modalities. It takes as inputs 3D patches (small volumetric cuts) extracted from MRIs at different modalities, and analyzes the information using two pathways. It then uses fully connected layers to obtain a segmentation of each category [13].

In this paper, we aim at providing an efficient, accurate and objective way of automatically estimating the volume and location of a brain tumor. For this purpose, we use the 2017 BraTS Challenge dataset, which utilizes multi-institutional pre-operative MRI scans and focuses on the segmentation of intrinsically heterogeneous (in appearance, shape, and histology) brain tumors, namely gliomas

[4, 14, 15, 16]. This open source dataset has MRIs from 210 patients with HGG and 75 patients with LGG tumors for training; there are four different MRI modalities per patient and annotations made by several specialists. In terms of methodology, inspired by DeepMedic’s [13] success on modeling multiscale information, we developed a neural network with four contracting pathways and residual connections that receive patches centered on the same voxel, but with different spatial resolutions. During the testing stage, the average dice coefficient and the Hausdorff distance were calculated to measure the performance of the methods.

2 Methodology

2.1 Multimodality Volumetric Neural Network

Multiple resolutions: different architectures for semantic segmentation, such as VGG [17] and FCN [10], take advantage of multiple image resolutions to simultaneously extract fine details and coarse structures from the input data. This is done using groups of convolutional layers and non-linearities, usually Rectified Linear Units (ReLU), followed by pooling operations. However, as the image resolution is reduced, so is the accuracy in the segmentation location. To overcome this drawback, we designed a network that extracts features from different input resolutions in a parallel and independent manner. This allows us to retrieve detailed appearance data along with accurate semantic information. After that, we can combine those results to obtain the final segmentation.

Figure 1 shows an overview of our approach. Our method has four identical parallel pathways, each one with six convolutional layers and two residual connections. All the paths receive patches centered at the same voxel, but extracted from different versions of the image (original and downsampled by factors of three, six and eight). The patches have input sizes of 36^3 , 20^3 , 18^3 and 15^3 for the different resolution pathways. We tested different downsample factors and input sizes, and the best result was chosen empirically on the validation set. In addition, deconvolutional layers are used to upsample the outputs when necessary. Finally, the results are concatenated and introduced in the fully connected layers to be combined and then classified. The classification layer is a convolution with kernel size of 1^3 and the final output is predicted using a softmax classifier.

Patch-wise approach: given the amount of data in a MRI, the memory requirements to process each image are substantial. Furthermore, the use of multiple modalities increases the input size even more, resulting in considerable memory consumption. On the other hand, segmenting brain tumors is a highly imbalanced problem, in which the background voxels cover over 90% of the images, while the remaining elements can belong to any of the three internal structures of the tumor. Nonetheless, both of these problems can be addressed by analyzing small patches rather than the whole image. This is because patches do not only reduce the input data size, but can also be used to balance the

number of instances per category that the model will see. With this in mind, we trained our method using patches extracted randomly from the training images. The only constraints imposed were that 50% of the patches must be centered on a foreground voxel, and no patches centered on background voxels that don't belong to the brain were extracted.

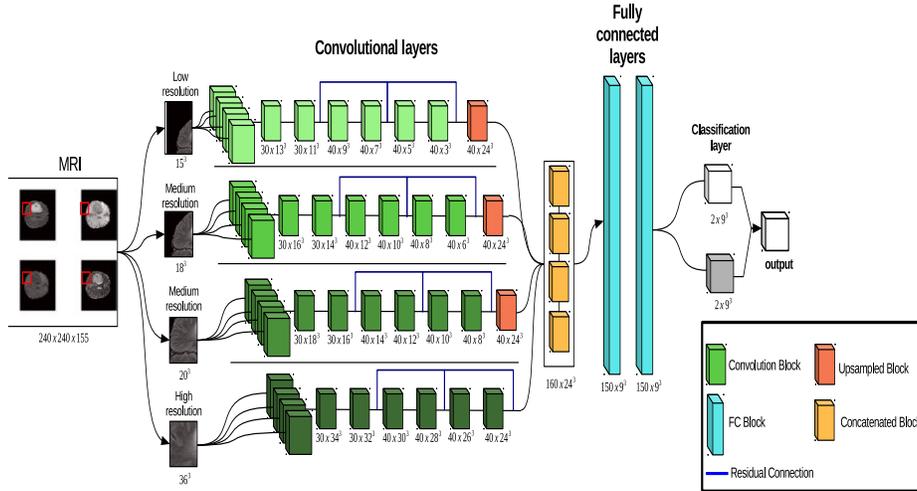


Fig. 1. Proposed Architecture. The kernels of the convolutions in the four pathways are 3^3 and no padding was made in those operations. The input of the four paths are 3D patches of each modality centered in the same voxel, but the lower resolution patches are obtained from downsampled versions of the image by factors of 3, 6 and 8, respectively.

2.2 Data

The method was trained, validated and tested using the BraTS challenge 2017 datasets. The training dataset includes 210 different MRI files from high grade glioma (HGG) cases, and 75 MRIs from low grade gliomas (LGG). The validation and test datasets include 46 and 146 different MRI files, respectively. Every image has four modalities: T1, T1 contrast-enhanced, T2 and FLAIR. The ground truth annotations were made by experts and manually-revised by board-certified neuroradiologists, and were made publicly available only for the training dataset. The annotations contain four different categories representing the background and the internal structures of the tumor as shown in Figure 2 and listed below [4, 14, 15, 16]:

0. Everything Else.

1. Necrosis and Non-Enhancing tumor.
2. Edema.
4. Enhancing tumor.

The internal structures are used to obtain the segmentations of the three glioma sub-regions evaluated in the challenge: The enhancing tumor (ET); the tumor core (TC), that includes the necrotic area, the non-enhancing and enhancing tumors; and the whole tumor (WT), represented by the edema.

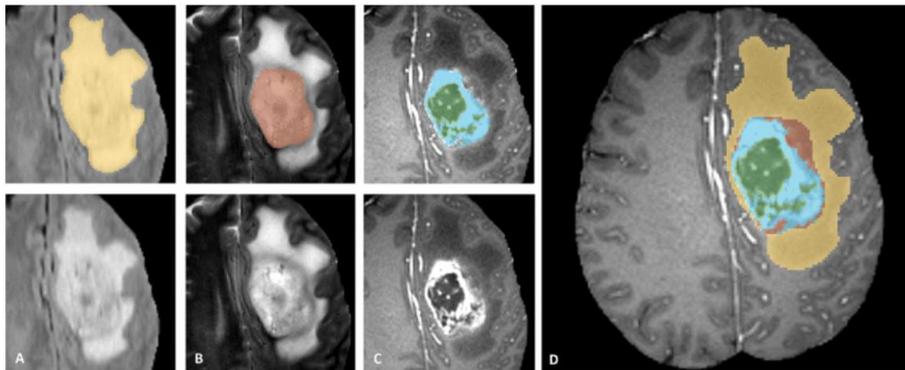


Fig. 2. Manual annotation through expert raters. Shown are image patches with the tumor structures that are annotated in the different modalities (top left) and the final labels for the whole dataset (right). Image patches show from left to right: the whole tumor visible in FLAIR (a), the tumor core visible in T2 (b), the enhancing tumor structures visible in T1c (blue), surrounding the cystic/necrotic components of the core (green) (c). Segmentations are combined to generate the final labels of the tumor structures (d): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue). (Figure taken from the BraTS IEEE TMI paper [4])

2.3 Training

The architecture was trained using the 285 cases from the training dataset, without discriminating between the two categories of glioma, seeking to obtain a robust model that could segment both lesions without difficulty. Our method’s inputs are patches of size 36^3 that are extracted randomly, making sure that 50% of them are centered at a voxel labeled as tumor, as explained in section 2.1. The data is normalized individually per MRI volume by setting the mean to 0 and the variance to 1. Data augmentation is made to avoid overfitting of the model due to the small size of the training dataset, and it is performed on the fly to prevent memory issues. The process is made by reflecting randomly chosen volumes along the sagittal axis. To train the method, the learning rate was set to $1e - 4$ and it remained constant during the 35 epochs. We use sparse softmax cross entropy loss and minimize it using the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1e - 8$.

2.4 Validation and test

To test the model, the 46 volumes from the validation dataset and the 146 from the test dataset were evaluated with the network. The only pre-processing step applied to the data is the individual normalization per MRI described in section 2.3. In these stages, the patches are extracted at uniform intervals in the images. No additional post-processing was done to the volumes. Lastly, the whole volume was reconstructed using the segmented patches. In this testing stage a new MRI takes less than 15 seconds in producing the prediction.

2.5 Evaluation Metrics

We consider the four complementary performance metrics proposed in the challenge for quantitative evaluation.

Dice Coefficient: for every model, the Dice-Coefficient (Equation 1) is calculated as performance metric. This measure states the similarity between clinical ground truth annotations and the output segmentation of the model. Afterwards, we calculate the average of the results to obtain the overall dice coefficient of the models.

$$DC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

Hausdorff Distance: the Hausdorff Distance (Equation 2) is mathematically defined as the maximum distance of a set to the nearest point in the other set [18]. In other words, it measures how close are the segmentation's and the expected output's boundaries. This metric is used to assess the alignment between the contours of the segmentations.

$$H(A, B) = \max\{\min\{d(A, B)\}\} \quad (2)$$

Sensitivity and Specificity: are statistical measures used to evaluate the behavior of the predictions and the proportions of True Positives (TP), False Negatives (FN), False Positives (FP) and True Negatives (TN). The Sensitivity (Equation 3), also known as True Positive Rate, gives the proportion of true positives predicted correctly. The specificity (Equation 4), also known as True Negative Rate, measures how well the true negatives are predicted.

$$Sensitivity = TPR = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = TNR = \frac{TN}{TN + FP} \quad (4)$$

3 Experimental Results

we performed extensive experiments to find the optimal number of paths needed to solve this task. We tested the architecture for three, four and five pathways with multiple resolutions in the validation dataset. In Table 1 we present the results:

Table 1. Dice coefficient, sensitivity, specificity and Hausdorff distance of our neural network for Enhanced Tumor, Whole Tumor and Core Tumor; evaluated over the validation dataset from BraTS 2017.

	Dice			Sensitivity			Specificity			Hausdorff		
	Enh.	Wh.	Core	Enh.	Wh.	Core	Enh.	Wh.	Core	Enh.	Wh.	Core
3 paths	0,68	0,86	0,69	0.74	0.87	0.69	0.99	0.99	0.99	10.34	14.74	14.12
4 paths	0,71	0,88	0,68	0.72	0.86	0.68	0.99	0.99	0.99	6,12	9,63	11,38
5 paths	0,68	0,88	0,69	0.74	0.86	0.69	0.99	0.99	0.99	7.43	7.99	12.86

As demonstrated by Table 1, the four-pathway architecture obtains the best results in the previous experiment. The 5 path approach (with additional down-samples of three, four, six and eight) gets a similar performance with a minimal decrease. However, this method uses more trainable parameters and therefore takes longer to train. For this reason, we choose the four-pathway architecture that takes advantage of a multi-resolution approach, in order to get information of the location of the tumor and, at the same time, acquire local data that helps to differentiate the structures of the lesions in a reasonable time. In Table 2 we present the results on the training, validation and test datasets using the evaluation metrics explained in 2.5:

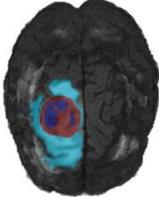
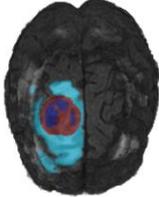
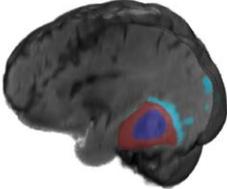
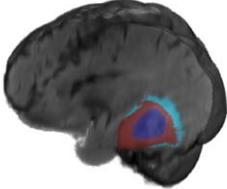
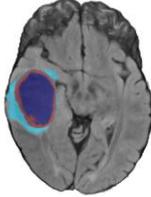
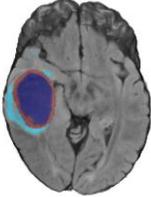
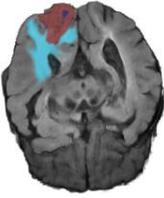
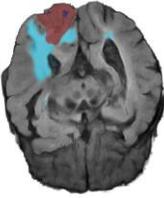
Table 2. Dice coefficient, sensitivity, specificity and Hausdorff distance of our neural network for Enhanced Tumor, Whole Tumor and Core Tumor; evaluated over the training, validation and test datasets from BraTS 2017. Note: Sensitivity and specificity measures were not provided for the evaluation in test dataset.

	Dice			Sensitivity			Specificity			Hausdorff		
	Enh.	Wh.	Core	Enh.	Wh.	Core	Enh.	Wh.	Core	Enh.	Wh.	Core
Train	0,74	0,89	0,87	0.83	0.91	0.89	0.99	0.99	0.99	5,85	15,99	11,18
Val	0,71	0,88	0,68	0.72	0.86	0.68	0.99	0.99	0.99	6,12	9,63	11,38
Test	0,65	0,86	0,67	-	-	-	-	-	-	51,70	10,39	36,20

Overall, our approach reached a competitive result. The usage of different levels of resolution and fully connected layers proved to be an effective way to obtain a detailed and accurately located segmentation of brain tumors in MRIs. This behavior is exhibited in the average dice coefficients obtained for the whole tumor segmentation task, as shown on Table 2. Our method reached a result of

89% in the training phase. Furthermore, when introducing completely new data we get a similar performance (88% in validation and 86% in test). The minimal change between these results demonstrates the robustness of the model.

Table 3. Visual comparison between the ground truth against some results obtained by our neural network

Ground truth	Our segmentation
	
	
	
	

The sensitivity and specificity were also measured for the train and validation datasets. We obtain high sensitivity results, ranging between 68% and 91% for all the evaluated tasks. Therefore, the method has a good recall. In medical problems, this is a specially important measure due to the interest in finding every ailment affecting the patient, in order to prevent any complication. Addition-

ally, our neural network obtains a specificity of 99% for all the evaluated tasks in both training and validation, which shows that the method is able to predict with great certainty which parts of the brain are healthy. Another strong point of our network is the ability to produce fast segmentations, as it takes around fifteen seconds to process the four modalities and produce the new segmentation. In critical and time-sensitive clinical situations, the efficiency of our approach represents a faster diagnosis and a higher probability of survival for the patient.

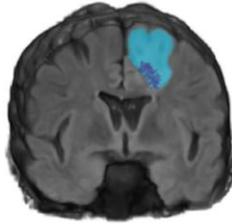
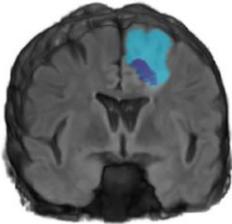
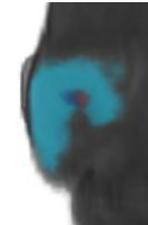
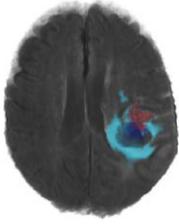
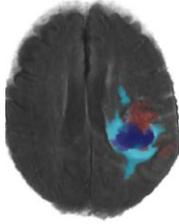
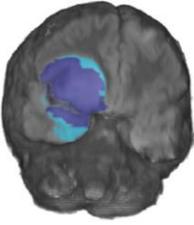
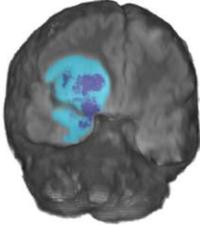
Table 4. Dice coefficient, sensitivity, specificity and Hausdorff distance of our neural network and the Deepmedic implementation for Enhanced Tumor, Whole Tumor and Core Tumor; evaluated over validation datasets from BraTS 2017.

	<u>Dice</u>			<u>Sensitivity</u>			<u>Specificity</u>			<u>Hausdorff</u>		
	Enh.	Wh.	Cor.	Enh.	Wh.	Cor.	Enh.	Wh.	Cor.	Enh.	Wh.	Cor.
Deepmedic	0.69	0.86	0.68	0.72	0.86	0.64	0.99	0.99	0.99	10.1	25.0	17.5
Ours	0,71	0,88	0,68	0.72	0.86	0.68	0.99	0.99	0.99	6,12	9,63	11,38

Table 4 shows a comparison between the results of our architecture against the implementation of Deepmedic (available in [13]) over the validation dataset of BraTS 2017. Our method reached a better performance in all the evaluated metrics, the high improvement in the Enhancing tumor and Whole tumor tasks (2 points in Dice coefficient measure) demonstrates that our approach is not only able to locate the area of the tumor better, but also has a greater capacity to identify correctly the internal structures of the lesion, information that can be vital when performing a diagnosis and treatment of a patient.

In Table 3, we present some examples of the predictions against the ground truth. It is important to emphasize the precision with which our method differentiates the structures that compose the tumor. We can see its capability to predict the exact area where the patient’s tumor occurs with minimal noisy activations in other areas. In Table 5, we present some examples of the predictions against the ground truth that show the limitations of our method. We found that it can identify and locate the tumor with a high degree of precision in all the evaluated cases. However, when differentiating between the inner parts of the tumor, it falls short in some examples. In general, as can be seen in Table 5, the prediction is correct for the location of the affected region, but it predicts false positives in the contours of the specific structures. Taking into account the difficulty of the problem, where even for the experts it is complicated to locate the tumor accurately and even more to identify their internal parts, our method obtains an robust performance in a reduced period time.

Table 5. Visual comparison between the ground truth against the worst results obtained by our neural network

Ground truth	Our segmentation
	
	
	
	

4 Conclusion

We propose a volumetric multimodality neural network. Our method receives as input 3D patches extracted from the dataset images. The architecture consist of four identical parallel pathways, to extract features on four specific resolution levels, each one with six convolutional layers and two residual connections. We then combine their results using fully connected layers. Finally, every pixel is classified into background or one of the three categories belonging to the tumor. In this paper, we have presented results in the 2017 BraTS Challenge dataset (Training, Validation and Test) reaching an average dice coefficient of 89% over

training dataset, 88% over validation dataset and 86% over test dataset for the whole tumor segmentation task (Table 2).

The use of multiple resolutions has proven to be an effective way to extract detailed information and coarse, semantic data from the images. However, most methods use a series of consecutive blocks and pooling operations for that purpose. As a consequence, deeper blocks lose some of the fine information obtained in early stages. In this paper, we showed that the use of parallel independent blocks to extract different levels of features allows us to obtain accurate and detailed results. Additionally, the use of a patch-wise approach has proven to be useful to deal with large amounts of data, which are highly imbalanced and need to be processed simultaneously to avoid the loss of 3D and multimodality information.

References

- [1] Mayo Clinic: Brain tumor - Symptoms and causes. <http://www.mayoclinic.org/diseases-conditions/brain-tumor/symptoms-causes/dxc-20117134> Online; accessed July 19, 2017.
- [2] kinderkrebsinfo.de: Brain tumours - tumours of the central nervous system. https://www.kinderkrebsinfo.de/diseases/brain_tumours/index_eng.html Accessed: 2017-08-24.
- [3] The Royal Marsden NHS Foundation Trust: Glioma. <https://www.royalmarsden.nhs.uk/your-care/cancer-types/paediatric-cancers/glioma> Accessed: 2017-08-24.
- [4] 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, L., 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., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Antonio Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S.J., Riklin-Raviv, T., Reza, S., Ryan, M., 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., Yee, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)". *IEEE Transactions on Medical Imaging* **34**(10) (2015) 1993,2024
- [5] Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. *Elsevier* **8**(3) (2004) 275–283
- [6] Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Joint tumor segmentation and dense deformable registration of brain mr images. *Springer* (2012) 651–658
- [7] Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.P., Reyes, M.: Segmentation of brain tumor images based on integrated hierarchical classification and regularization. *Proceedings MICCAI-BRATS* (2012)
- [8] Menze, B.H., Geremia, E., Ayache, N., Szekely, G.: Segmenting glioma in multimodal images using a generative-discriminative model for brain lesion segmentation. *Proceedings MICCAI-BRATS* (2012)

- [9] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. *ECCV* (2014) 818–833
- [10] Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. *CoRR* **abs/1605.06211** (2016)
- [11] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. *CoRR* **abs/1505.04597** (2015)
- [12] Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. *CoRR* **abs/1606.04797** (2016)
- [13] Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A., Criminisi, A., Rueckert, D., Glocker, B.: Deepmedic on brain tumor segmentation. Athens, Greece Proc. BRATS-MICCAI code:<https://github.com/Kamnitsask/deepmedic>.
- [14] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data* ([In Press]) (2017)
- [15] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: "segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection". *The Cancer Imaging Archive* (2017) doi:[10.7937/K9/TCIA.2017.KLXWJJ1Q](https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q).
- [16] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: "segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection". *The Cancer Imaging Archive* (2017) doi:[10.7937/K9/TCIA.2017.GJQ7ROEF](https://doi.org/10.7937/K9/TCIA.2017.GJQ7ROEF).
- [17] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
- [18] Rote, G.: Computing the minimum hausdorff distance between two point sets on a line under translation. *Information Processing Letters* **38**(3) (1991) 123–127