

Segmentation of gliomas and prediction of patient overall survival: a simple and fast procedure.

Elodie Puybareau, Guillaume Tochon, Joseph Chazalon and Jonathan Fabrizio

EPITA Research and Development Laboratory (LRDE), France
elodie.puybareau@lrde.epita.fr

Abstract. In this paper, we propose a fast automatic method that segments glioma without any manual assistance, using a fully convolutional network (FCN) and transfer learning. From this segmentation, we predict the patient overall survival using only the results of the segmentation and a home made atlas. The FCN is the base network of VGG-16, pre-trained on ImageNet for natural image classification, and fine tuned with the training dataset of the MICCAI 2018 BraTS Challenge. It relies on the "pseudo-3D" method published at ICIP 2017, which allows for segmenting objects from 2D color images which contain 3D information of MRI volumes. For each n^{th} slice of the volume to segment, we consider three images, corresponding to the $(n - 1)^{\text{th}}$, n^{th} , and $(n + 1)^{\text{th}}$ slices of the original volume. These three gray-level 2D images are assembled to form a 2D RGB color image (one image per channel). This image is the input of the FCN to obtain a 2D segmentation of the n^{th} slice. We process all slices, then stack the results to form the 3D output segmentation. With such a technique, the segmentation of a 3D volume takes only a few seconds. The prediction is based on Random Forests, and has the advantage of not being dependant of the acquisition modality, making it robust to inter-base data.

Keywords: glioma · tumor segmentation · fully convolutional network · random forest · survival prediction.

1 Introduction

1.1 Motivation

Gliomas are the most common brain tumors in adults, growing from glial cells and invading the surrounding tissues [8]. Two classes of tumors are observed. The patients with the more aggressive ones, classified as high-grade gliomas (HGG), have a median overall survival of two years or less and imply immediate treatment [12, 15]. The less aggressive ones, the low-grade gliomas (LGG), allow an overall survival of several years, with no need of immediate treatment. Multimodal magnetic resonance imaging (MRI) helps practitioners to evaluate the degree of the disease, its evolution and the response to treatment. Images are analyzed based on qualitative or quantitative measures of the lesion [7, 20].

Developing automated brain tumor segmentation techniques that are able to analyze these tumors is challenging, because of the highly heterogeneous appearance and shapes of these lesions. Manual segmentations by experts can also be a challenging task, as they show significant variations in some cases.

During the past 20 years, different algorithms for segmentation of tumor structures has been developed and reviewed [1, 5, 6]. However, a fair comparison of algorithms implies a benchmark based on the same dataset, as it has been proposed during MICCAI BraTS Challenges [14]. One aim of the tumor segmentation is to predict the overall patient survival to adapt the treatment.

1.2 Context

This work has been done in the context of the MICCAI 2018 Multimodal Brain Tumor Segmentation Challenge (BraTS)¹. The aim was to provide a fully automated pipeline for the segmentation of the glioma from multi modal MRI scans without any manual assistance and to predict the patient overall survival.

Despite the relevance of glioma segmentation, this segmentation is challenging due to the high heterogeneity of tumors. The development of an algorithm that can perform fully automatic glioma segmentation and overall prediction of survival would be an important improvement for patients and practitioners.

We received data of 286 patients, with associated masks to develop our method. The data, available online, have been annotated and preprocessed [2–4]. The volumes given are T1, T1ce, T2 and FLAIR. Our method is then evaluated on new volumes: a validation set released by the organizers without the manual segmented masks (these masks will not be released), to obtain preliminary results. For the challenge evaluation, a new set will be released and we are asked to send the segmentations within 48h after receiving the data. This challenge will establish a fair comparison to state-of-the-art methods, and a release of a large annotated dataset.

As the segmentation is not an aim but a tool, this challenge also evaluate the survival prediction.

1.3 Related Works

As machine learning really improved the results of some segmentation tasks, the use of such strategy seems meaningful in the context of medical image segmentation.

In a work published in the IEEE Intl. Conf. on Image Processing (ICIP) in 2017 [21], 3D brain MR volumes are segmented using fully convolutional network (FCN) and transfer learning. The network used for transfer learning is VGG (Visual Geometry Group) [17], pre-trained on the ImageNet dataset. It takes as input a 2D color image that is here a 3D-like image, or *3D-like* image, composed of 3 consecutive slices of the 3D volume (see Fig. 1). This method uses only one modality, and reaches good results for brain segmentation.

¹ <https://www.med.upenn.edu/sbia/brats2018.html>

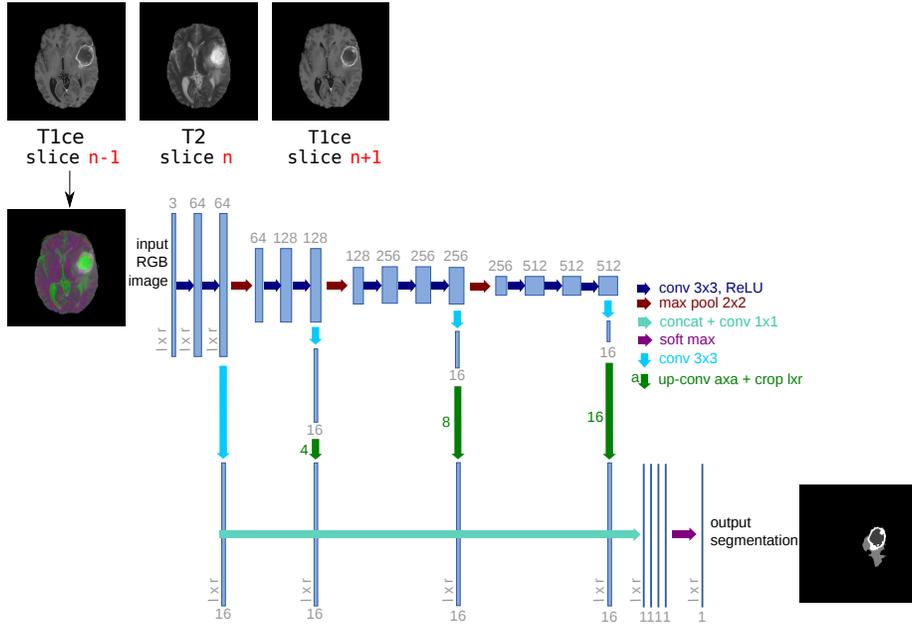


Fig. 2. Architecture of the proposed network. We fine tune it and combine linearly fine to coarse feature maps of the pre-trained VGG network [17]. Note that each input color image is built from the slice n and its neighbouring slices $n - 1$ and $n + 1$.

natural images. As it was a success, we adapted it to glioma segmentation. We rely on the 16-layer VGG network [17], which was pre-trained on millions of natural images of ImageNet for image classification [10]. For our application, we keep only the 4 stages of convolutional parts called “*base network*”, and we discard the fully connected layers at the end of VGG network. This base network is mainly composed of convolutional layers: $z_i = w_i \times x + b_i$, Rectified Linear Unit (ReLU) layers for non-linear activation function: $f(z_i) = \max(0, z_i)$, and max-pooling layers between two successive stages, where x is the input of each convolutional layer, w_i is the convolution parameter, and b_i is the bias term. The three max-pooling layers divide the base network into four stages of fine to coarse feature maps. Inspired by the work in [11, 13], we add specialized convolutional layers (with a 3×3 kernel size) with K (e.g. $K = 16$) feature maps after the convolutional layers at the end of each stage. All the specialized layers are then rescaled to the original image size, and concatenated together. We add a last convolutional layer with kernel size 1×1 at the end. This last layer combine linearly the fine to coarse feature maps in the concatenated specialized layers, and provide the final segmentation result. The proposed network architecture is schematized in Fig. 2.

The architecture described above is very similar with the one used in [13] for retinal image analysis, where the retinal images are already 2D color images.

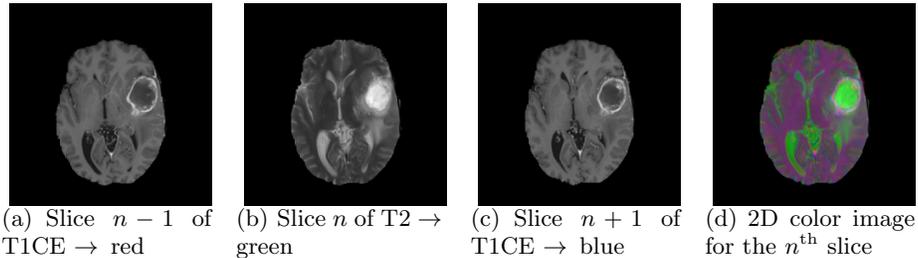


Fig. 3. Three successive slices (a-c) are used to build a 2D color image (d) from for example T1CE and T2 images.

Using such a 2D representation avoids the expensive computational and memory requirements of fully 3D FCN.

For the training phase, we use the multinomial logistic loss function for a one-of-many classification task, passing real-valued predictions through a softmax to get a probability distribution over classes. During training, we use the classical data augmentation strategy by scaling and rotating. We rely on the ADAM optimization procedure [9] (AMSGrad variant [16]) to minimize the loss of the network. The relevant parameters of the methods are the following: the learning rate is set to 0.002 (we did not use learning rate decay), the beta_1 and beta_2 are respectively set to 0.9 and 0.999, and we use a fuzz factor (epsilon) of 0.001.

At test time, after having pre-processed the 3D volume (requantization), we prepare the set of 2D color images and pass every image through the network.

We run the train and test phase on an NVIDIA GPU card. The testing one lasts less than 10 seconds.

The output of the network for one slice during the inference phase is a 2D segmented slice. After treating all the slices of the volume, all the segmented slices are stacked to recover a 3D volume with the same shape as the initial volume, and containing only the segmented lesions.

Post-processing We simply regularize the segmented volumes using the morphological closing.

2.2 Patient survival prediction

The second task of the MICCAI 2018 BraTS challenge is concerned with the prediction of patient overall survival from pre-operative scans (only for subjects with gross total resection (GTR) status). Note that, as precognized by the evaluation framework, the classification procedure is conducted by labeling subjects into three classes: short-survivors (less than 10 months), mid-survivors (between 10 and 15 months) and long-survivors (greater than 15 months).

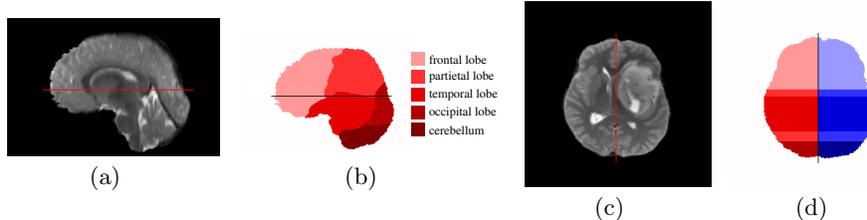


Fig. 4. (a) Sagittal and (c) axial slices from the T2 modality of a brain and (b,d) corresponding rescaled brain atlas.

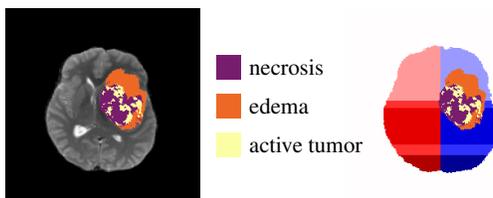


Fig. 5. T2 slice (left) and corresponding atlas slice (right) with segmented tumor overlaid.

Definition and extraction of relevant features The first step of the prediction task is the definition and extraction of relevant features impacting the survival of patients. Beside the patient age, we decided to focus on the tumor size and its localization within the brain. More specifically, we denote by S_i the segmented volume predicted by our Deep FCN architecture, as described in section 2.1 for the i^{th} patient. Voxels in S_i are labeled by 1, 2 and 4 (corresponding to ,  and  in Fig. 5, respectively), depending whether they were classified as necrosis, edema or active tumor, respectively.

Thus, we define the relative size of each class in S_i with respect to the total brain size (the number of non-zero voxels in the patient T2 modality) as the features related to the tumor size.

In order to describe the tumor position, we created a crude brain atlas divided in 10 regions accounting for the frontal, parietal, temporal and occipital lobes and the cerebellum for each hemisphere, as displayed by Figs. 4(b) and 4(d). The 3D atlas was first shaped to the average bounding box dimensions of all patients with GTR status, *i.e.* $170 \times 140 \times 140$ pixels. It is then adjusted to each patient bounding box dimensions by nearest-neighbors interpolation, and finally masked by all non-zero voxels in the patient T2 modality. Finally, we retrieve the centroid coordinates of the region within the atlas that is affected the most by the necrosis (*i.e.*, the region that has the most voxels labeled as necrosis in S_i with respect to its own size) relatively to the brain bounding box as well as the relative centroid coordinates of the necrosis + active tumor and defined those as the relevant features accounting for the tumor position.

In summary, each patient brain/tumor is defined by the following 6 criteria:

1. the patient age.
2. the relative size of necrosis with respect to brain size.
3. the relative size of edema with respect to brain size.
4. the relative size of active tumor with respect to brain size.
5. the relative coordinates of the region in the atlas that is the most affected by necrosis with respect to the brain bounding box.
6. the relative coordinates of the binarized tumor (only considering necrosis and active tumor) with respect to the brain boundingbox.

This leads to a total of **10 features per patient** (since both centroids coordinates are 3-dimensionals).

Training phase For the training phase, we first extract the feature vector $\mathbf{x}_i \in \mathbb{R}^{10}$ of each of the N patients in the training set (with $N = 59$), as described in section 2.2 above. All those feature vectors are stacked in a $N \times 10$ feature matrix $\mathbf{X}_{\text{train}}$ on which a principal component analysis (PCA) is performed. The feature-wise mean \mathbf{m}_{PCA} and standard deviation $\boldsymbol{\sigma}_{\text{PCA}}$ vectors computed during the scaling phase of the PCA, as well as the projection matrix \mathbf{V}_{PCA} are stored for further use. Finally, the PCA output is normalized again, yielding the $N \times 10$ matrix $\mathbf{Y}_{\text{train}}$. Finally, we train N_{RF} random forest (RF) classifiers [19] on all rows of $\mathbf{Y}_{\text{train}}$, using the true label vector $\mathbf{y}_{\text{label}}$ as target values, and store those RFs. The whole training phase is depicted by the workflow in Fig. 6(a). Each RF is composed of 10 decision trees, for which splits are performed using 3 features randomly selected among the 10 available, and based on the Gini impurity criterion [18]. Here, we arbitrary fixed $N_{\text{RF}} = 50$ in order to account for the stochastic behavior of RF classifiers.

Test phase The test phase is summarized by the workflow in Fig. 6(b). Features are computed in a similar manner for a patient belonging to the test data set as they are for the training set. The feature vector \mathbf{x}_{test} is then normalized using \mathbf{m}_{PCA} and $\boldsymbol{\sigma}_{\text{PCA}}$ and further projected in the PC space with \mathbf{V}_{PCA} , learnt during the PCA step of the training stage. The resulting vector \mathbf{y}_{test} is then fed to the N_{RF} RF classifiers, leading to N_{RF} independent class label predictions. The final label prediction y_{pred} (1, 2 and 3 for short-, mid- and long-survivors, respectively) is eventually obtained by majority voting.

3 Experiments and Results

This section presents the experiments and results obtained, during the development of our method (using the training dataset).

3.1 Experiments

In this part, we used only the training scans provided during the challenge.

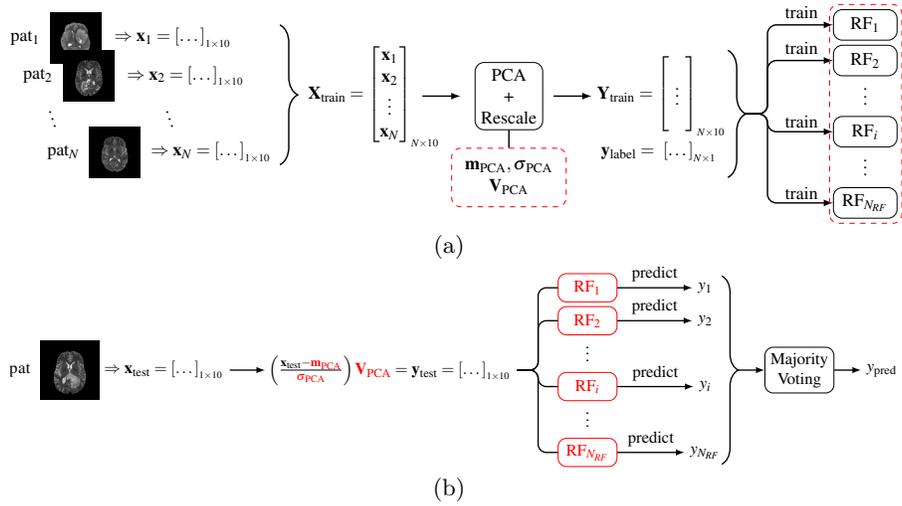


Fig. 6. (a) Training and (b) test procedures. The stored information after the training phase is encircled in dashed red in the training workflow (a).

Modalities We formed multi-modality pseudo-3D images using T1ce and T2 modalities: for each slice n , we combined the slice n of T2 with its $n - 1$ and $n + 1$ from T1ce.

Training and testing. We trained our model on randomly chosen scans. The model was trained using the parameters described in the previous section. We tested on the remaining scans and cross-validated.

Axis and combination Our method deals with 2D RGB images that are pseudo-3D. A clue to improve segmentation results and to get a better contours detection is to associate three networks, each network being trained on a particular axis (axial, sagittal and coronal), and to combine their results to obtain the final segmentation.

We first trained 3 networks, one in each axis. The inference was done according to the axis, so for one volume we obtained 3 segmentation. These segmentations are then combined: for each voxel, the final segmentation is the result of the majority voting procedure.

3.2 Results

Note The preliminary result is a dice of 0.87 for the whole tumor, computed on a fixed test set (10% of training set). We do not have yet the details of our performances during the challenge for the segmentation part.

For the prediction part, we reached the 2nd place during the challenge. Quantitative results will be added for the camera ready version. The Docker image will be available on <https://www.lrde.epita.fr/wiki/NeoBrainSeg>

4 Conclusion

We propose in this article a method to segment glioma in few seconds based on transfer learning from VGG-16, a pre-trained network used to classify natural images. This method takes the advantage of keeping 3D information of the MRI volume and the speed of processing only 2D images, thanks to the pseudo-3D concept.

This method can also deal with multi-modality, and can be applied to other segmentation problems, such as in [23], where a similar method is proposed to segment white matter hyperintensities, but pseudo-3D has been replaced by an association of multimodality and mathematical morphology pre-processing to improve the detection of small lesions. Hence, we might also try to modify our inputs thanks to some highly non-linear filtering to help the network segment tumors, precisely some mathematical morphology operators [22].

The strength of this method is its modularity and its simplicity. It is easy to implement, fast, and does not need a huge amount of annotated data for training (in the work on brain segmentation [21], there is only 2 images for training for some cases).

From a segmentation result, we propose a simple and efficient method to predict the patient overall survival, based on Random Forests. This method only needs as input a segmentation, a brain atlas and a brain volume for atlas registration. It means that our method is robust to the different acquisitions, does not need a special modality or setting, yielding to a method robust to inter-base variations.

Acknowledgments The authors would like to thank the organizers of the BraTS 2018 Challenge and the MICCAI Brainles Workshop, and Dr. Marie Donzel from Claude Bernard University Lyon 1 medical school for the useful discussions regarding the definition of relevant brain features for the survival prediction. The GPU card “Quadro P6000” used for the work presented in this paper was donated by NVIDIA Corporation.

References

1. Angelini, E.D., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Duffau, H.: Glioma dynamics and computational models: a review of segmentation, registration, and in silico growth algorithms and their clinical applications. *Current Medical Imaging Reviews* **3**(4), 262–276 (2007)
2. 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* **286** (2017)
3. 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* **286** (2017)

4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data* **4**, 170117 (2017)
5. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of mri-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology* **58**(13), R97 (2013)
6. Bonnín Rosselló, C.: Brain lesion segmentation using Convolutional Neuronal Networks. B.S. thesis, Universitat Politècnica de Catalunya (2018)
7. Eisenhauer, E.A., Therasse, P., Bogaerts, J., Schwartz, L.H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., et al.: New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European journal of cancer* **45**(2), 228–247 (2009)
8. Holland, E.C.: Progenitor cells and glioma formation. *Current opinion in neurology* **14**(6), 683–688 (2001)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR abs/1412.6980* (2014)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. pp. 1097–1105 (2012)
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. pp. 3431–3440 (2015)
12. Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., Kleihues, P.: The 2007 who classification of tumours of the central nervous system. *Acta neuropathologica* **114**(2), 97–109 (2007)
13. Maninis, K.K., Pont-Tuset, J., Arbeláez, P., Gool, L.V.: Deep retinal image understanding. In: *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Part II. Lecture Notes in Computer Science*, vol. 9901, pp. 140–148. Springer (2016)
14. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging* **34**(10), 1993 (2015)
15. Ohgaki, H., Kleihues, P.: Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *Journal of Neuropathology & Experimental Neurology* **64**(6), 479–489 (2005)
16. Reddi, S.J., Kale, S., Kumar, S.: On the convergence of adam and beyond. In: *International Conference on Learning Representations* (2018)
17. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014)
18. Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T.: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics* **8**(1), 25 (2007)
19. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P.: Random forest: a classification and regression tool for compound classification and qsar modeling. *Journal of chemical information and computer sciences* **43**(6), 1947–1958 (2003)
20. Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorensen, A.G., Galanis, E., DeGroot, J., Wick, W., Gilbert, M.R., Lassman, A.B., et al.: Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Journal of clinical oncology* **28**(11), 1963–1972 (2010)

21. Xu, Y., Géraud, T., Bloch, I.: From neonatal to adult brain MR image segmentation in a few seconds using 3D-like fully convolutional network and transfer learning. In: Proceedings of the 23rd IEEE International Conference on Image Processing (ICIP). pp. 4417–4421. Beijing, China (September 2017)
22. Xu, Y., Géraud, T., Najman, L.: Connected filtering on tree-based shape-spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38**(6), 1126–1140 (June 2016)
23. Xu, Y., Géraud, T., Puybareau, E., Bloch, I., Chazalon, J.: White matter hyperintensities segmentation in a few seconds using fully convolutional network and transfer learning. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries— 3rd International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Revised Selected Papers, Lecture Notes in Computer Science, vol. 10670, pp. 501–514. Springer (2018)