

A 2D/3D Convolutional Neural Network for Brain White Matter Lesion Detection in Multimodal MRI

Leire Roa-Barco¹, Oscar Serradilla-Casado¹, Mikel de Velasco-Vázquez¹,
Asier López-Zorrilla¹, Manuel Graña^{1,2}(✉), Darya Chyzhyk³,
and Catherine Price³

¹ Computational Intelligence Group, UPV/EHU, San Sebastian, Spain

² ACPySS, San Sebastian, Spain

`manuel.grana@ehu.es`

³ McKnight Brain Institute, University of Florida, Gainesville, USA

Abstract. White matter hyperintensities (WMH) are characteristics of various brain diseases, so automated detection tools have a broad clinical spectrum. Deep learning architectures have been recently very successful for the segmentation of brain lesions, such as ictus or tumour lesions. We propose a Convolutional Neural Network composed of four parallel data paths whose input is a mixture of 2D/3D windows extracted from multimodal magnetic resonance imaging of the brain. The architecture is lighter than others proposed in the literature for lesion detection so its training is faster. We carry out computational experiments on a dataset of multimodal imaging from 18 subjects, achieving competitive results with state of the art approaches.

1 Introduction

White matter hyperintensities (WMH) can be caused by a variety of factors including ischemia, micro-hemorrhages, gliosis, damage to small blood vessel walls. Many patients showing WMH are idiopathic, however WMH have a strong relationship with age, arterial hypertension, demographic parameters such as gender, and some disease, such as diabetes, and biomarkers such as cholesterol [15]. It has been found associated with progressive cognitive impairment [5]. WMH are small size lesions compared with tumours and stroke lesions, lacking their structure of necrotic and inflamed tissues. They are mostly periventricular lesions, which primarily appear at the top of the horns of the lateral ventricles progressing around the ventricles. They may also appear as subcortical lesions [9, 25]. Several magnetic resonance image (MRI) modalities may be used for WMH detection and segmentation. They appear as hypointense in T1-weighted and as hyperintense in T2-weighted images [23]. The best modality is the fluid attenuated inversion recovery (FLAIR) imaging, where the lesions appear as hyperintense and with greater contrast, allowing to differentiate between

periventricular and subcortical lesions. Recent studies [17,21] also consider diffusion tensor imaging (DTI), specifically the scalar coefficients such as fractional anisotropy (FA), radial diffusivity (RD), and mean diffusivity (MD), which give the information about privileged directions of water diffusion, so they are sensitive to microstructural changes in white matter. In the last years, the interest in brain lesion image segmentation has increased, for example, public challenges have been carried out BRATS <http://braintumorsegmentation.org/> and ISLES <http://www.isles-challenge.org/> to advance the field. Most research on small lesion detection has been carried out for multiple sclerosis (MS) patients. Early approaches consisted in semiautomatic labellings in structural images [16] and FLAIR [11]. Early multimodal approaches applied voxelwise fuzzy expert systems [1] and Markov random fields (MRF) [20]. Machine learning supervised approaches have been also applied, such as Random Forest [8] and MRF regularized versions [22]. Unsupervised approaches have made advantage of the brain symmetry for big lesion detection [6]. Recently, Deep Learning approaches report great success in the segmentation of brain tumours, specifically Convolutional Neural Networks (CNN) [18,26] which is the approach that we are following in our own proposal. Processing 3D medical images by the CNNs can be done in 3 ways: (a) Considering each 2D slice of the 3D volume in some direction (sagittal, coronal or axial) as an independent input image that is feed to the CNN [18,26]. (b) Considering 3D windows of the volumetric image as input. (c) Considering hybrid 2D/3D inputs, i.e. feeding 2D slices and 3D windows of the volumetric image. This decision carries some implications in the CNN design, because a 3D input forces that hidden layers resulting of the filters have 3D structure [3,24]. This additional structural complexity has been found cumbersome to deal with large datasets, because the number of operations scalate cubically instead of quadratically. So the intended advantage of preserving 3D spatial relation information, is countered by convergence issues and computational time, so that the 3D windows are small, losing information of long distance spatial relations. Finally, the use of hybrid 2D and 3D input information [2,7] allows a good balance between the preservation of 3D spatial relations and the long distance relations that can be analysed in 2D data. In our architecture, we have used an hybrid 2D/3D NN where we use a small 3D cube and three different 2D windows, one for each of the 3 dimensional axis. The paper contents is as follows: first we present the dataset used for the experiments. Secondly, we discuss our architecture and the others used for comparison. Then we present our experimental results and, finally, some conclusions and future work.

2 Materials

The experimental evaluation of the proposed CNN architecture has been carried out in a set of 18 subjects MRI images corresponding to a previous study [19] where WHM was performed manually, thus providing the ground truth segmentation for the present work as 3D lesion masks. Each subject image includes a 3D T1-weighted, FLAIR image, and diffusion weighted images from which DTI

images, and subsequent FA coefficients, were computed using FSL software. T1-weighted volumes have been registered to 1 mm MNI template. The FLAIR and FA images have been coregistered to the MNI space by affine registration to normalized T1-weighted images. The lesion masks are also coregistered to MNI space. All the image intensities are normalized to the $[0,1]$ interval.

3 Tested CNN Architectures

Throughout the last years, Convolutional Neural Networks (CNNs) [13] have achieved excellent performance in many computer vision tasks. Several advances have solved convergence issues, and the advent of easy to exploit powerful Graphics Processing Units (GPUs) has speed up the training times by several orders of magnitude [4]. A CNN is a shared-weight neural network: all the neurons in a hidden layer share the same weights and bias. In fact, each layer implements a linear convolution filter whose kernel is learnt by gradient descent. Therefore, the output of the successive layers is a series of filtered/subsampled images which are interpreted as progressively higher level abstract features. Most CNN are applied to 2D signals, i.e. images, however in the medical image domain they are increasingly applied to 3D signals, i.e. volumetric imaging information. Specifically, two recent instances of CNNs have been successfully applied to brain lesion segmentation [10, 12] achieving remarkable success in the BraTs competition. Another recent segmentation example using a 2D/3D input data is [7], where authors trained two separate CNNs for each input dimensionality, performing a combination of their outputs by averaging.

3.1 Our Proposal: MPCNN

Our proposal is a Mixed Parallel CNN (MPCNN), which takes four inputs: three orthogonal big 2D windows on 3D image slices (one per spatial dimension) centered at the same voxel of the brain, and a 3D window, a cube whose sides are smaller than that of the 2D windows. Therefore, 2D data carry farther away spatial relations, while the 3D window carries 3D spatial relations. The MPCNN architecture consists of four parallel CNN, three dedicated to process the 2D window, and the fourth processing the 3D window. Furthermore, we use multi-modal MRI data, specifically T1, FLAIR and FA volumes, so that each voxel is in fact a three dimensional vector, much like an RGB image. In this sense, independent CNN filters at each layer are learnt for each image modality. The output is a couple of binary units that provide an estimation of the probability that the central pixel of the 2D and 3D windows is a WMH lesion voxel. Figure 1 shows a diagram of the MPCNN architecture. Each parallel subnetwork is a CNN, composed of a sequence of convolutional layers and max-pooling layers which reduce the dimensionality of the feature space after each convolution. In the version of the network tested in this paper the dimension of each of the input 2D windows is 35×35 , whereas the dimension of the input 3D cube is $11 \times 11 \times 11$. The activation function used to compute the output

of each neuron of the CNN is the Rectified Linear Unit (ReLU) [13,14] due to both its efficient computation and the fact that it solves the vanishing gradient problem. The architectures of the three 2D CNNs are identical, they are composed of three convolutions with kernels of size 3×3 . The number of convolutions increases along the layers, increasing the number of features accordingly. Moreover, a dimensionality reduction max-pooling layer with pool size of 2×2 is applied to the output of the second and third layers. The dimensions of the output of each layer are shown in Fig. 1. Thus, each 2D subnetwork's output layer has $6 \times 6 \times 55 = 1980$ neurons. The 3D CNN is composed only of two 3D convolutions (with kernel size of $3 \times 3 \times 3$), and one 3D max-pooling (with pool size of $2 \times 2 \times 2$) after the second convolution. Finally, all the subnetworks are merged (this results in $1980 \times 3 + 1485 = 7425$ nodes) and fully connected to the next layer, composed of 128 neurons. Finally, these 128 outputs are used to compute the final output of the network via the Softmax function. Hence, the two outputs will always be bounded between 0 and 1, and they will sum 1. This facilitates a probabilistic interpretation of the network output as a probability of lesion at the central voxel.

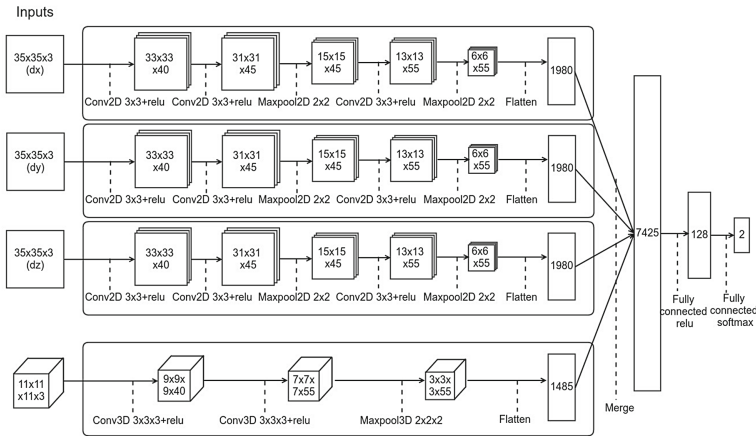


Fig. 1. The structure of the proposed for WMH lesion detection

3.2 ICCNN

For comparison, we have implemented a version of the Input Cascade CNN (ICCNN) architecture [10]. This network has two inputs: one for global context, and a smaller one for specific context. A convolution to the global context input is concatenated with the small input. Then, this data is divided into two parallel networks, each one analysing local and global features, by applying smaller and bigger kernels respectively. These networks are merged applying a final convolution, which ends up in a softmax layer. In our implementation of the network

we have reduced the dimension of the last layer to two neurons, which indicate whether the input represents a damaged voxel or not, and we have changed the training process, which has been done in one step with unbalanced data (10 negative cases per each positive damaged voxel). Moreover, we have changed the activation function to the ReLU, removed dropouts and used binary cross-entropy loss function for training. The main difference relative to MPCNN is that ICCNN only uses 2D slices as input.

3.3 DeepMedic

The other architecture tested for comparison is the DeepMedic [12], whose architecture has two main components; a 3D CNN and a fully connected 3D Conditional Random Field (CRF), which performs a postprocessing of the CNN output removing false positives. The CNN consists of four layers with $5 \times 5 \times 5$ kernels for feature extraction, and the classification layer is implemented as a convolutional layer with kernel of size $1 \times 1 \times 1$, allowing efficient dense-inference. The 3D CNN network has two pathways; one processes local information and the other processes larger contextual information, hence carrying out multi-scale processing of the data. Moreover, BN (Batch Normalization) is also applied to all the hidden layers, so that all Feature Maps obtained after each layer are normalized, preserving the signal, and avoiding serious weight convergence. After that, there are two hidden layers for combining the multi-scale parallel pathways. The full network is trained patch-by-patch and the size of the batches is selected automatically according to the neighborhood of the voxel in the input. The batches are built by extracting segments from the training images with 50% probability of being centered on a foreground or background voxel, which corrects the class-imbalance. The DeepMedic network training implementation downloaded from github was originally prepared for the ISLES and BraTS challenges, reporting state-of-the-art results on both performance on brain tumor and stroke lesion. However, since in our problem we only have 2 outputs not 5 as in the segmentation problems, in order to work with this network the last layer output has been reduced from 5 to 2 outputs.

Table 1. Results of the networks using holdout: TPR (True Positive Rate) and FPR (False Positive Rate)

#id	MPCNN		ICCNN		DeepMedic	
	TPR	FPR	TPR	FPR	TPR	FPR
#7	0.572	0.037	0.106	0.018	0.599	0.019
#15	0.503	0.024	0.259	0.063	0.650	0.013
#18	0.622	0.012	0.490	0.065	0.613	0.017
#21	0.245	0.011	0.280	0.034	0.463	0.012

4 Results

The MPCNN and ICCNN architectures have been implemented in Python using Keras with Tensorflow as backend. The DeepMedic implementation has been downloaded from github (<https://github.com/Kamnitsask/deepmedic>). The training and validation scripts have been executed in a desktop computer with RAM of 16 GB, and GPU NVIDIA GTX 1070 which has been used to speed up training. For validation, we apply holdout over the 18 available subject datasets: 14 have been used for training, and 4 for testing. To carry out the training in a limited reasonable time, we have subsampled the brain images as shown in Fig. 2 to obtain the training dataset. The brain image is decomposed in regular non-overlapping windows and a random voxel is picked from this window as

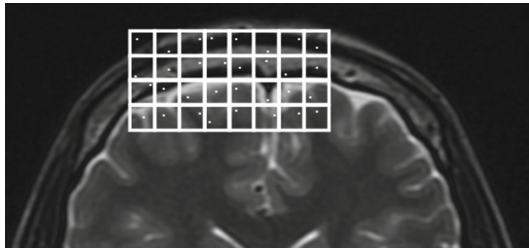


Fig. 2. Brain image subsampling to obtain the training dataset

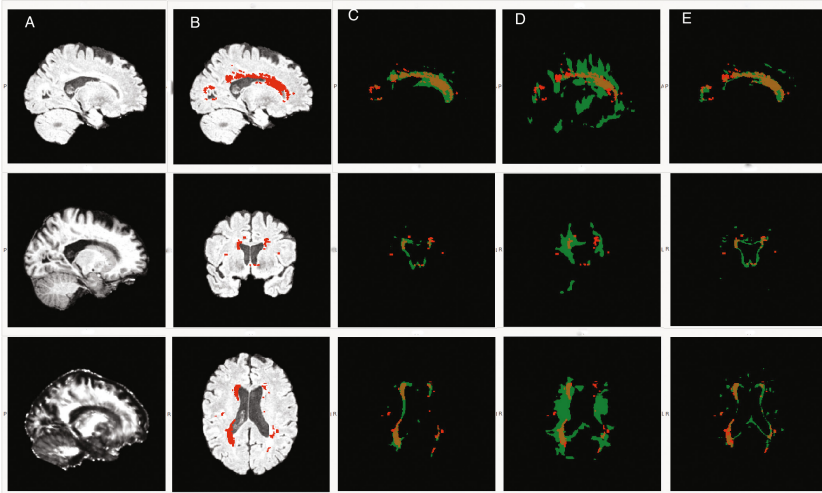


Fig. 3. Data and results of subject #18. A - Sample sagittal slides of T1, FA and FLAIR volumes. B - WMH ground truth lesion manually labeled overlaid on FLAIR slices C,D,E - prediction (green) and ground truth lesion (red), C for MPCNN, D for ICCNN, and E for DeepMedic

the center for the 2D/3D windows that conform the inputs. This process ensures a rather regular sampling interval and that the whole brain volume is sampled. Testing is carried out evaluating all the brain voxels in the test datasets. The problem is naturally imbalanced, i.e. there are many more healthy than lesion voxels, therefore we need to respect this imbalance in the training dataset. After some experimentation with a small CNN carrying crossvalidation on a reduced dataset, we set the imbalance ratio to 10 in the training dataset. In other words, we ensure that there is a ratio 10:1 of healthy to lesion voxels. We report True Positive Ratio (TPR) and False Positive Ratio (FPR) values, measuring how well the lesion is detected and the false alarms raised. Accuracy results for each test images are presented in Table 1. Overall, DeepMedic neural network has the best and most stable results, while ICCNN performs poorly. Our proposed MPCNN is faster to train than DeepMedic (a ratio 7:1) and has comparable results in two subjects (#7, #18), and slightly worse in another (#15). If we consider the maximum TPR achieved (0.65), seems that the architectures need to be improved, and that the success in tumour segmentation does not ensure success in WMH lesion detection. Figure 3 presents visual results of the experiment. From left to right, the first column shows images of the three modalities as an illustration of the dataset. The second column shows the lesion detected manually in three slices of brain #18 overlaid on the FLAIR image. Next columns illustrate the detection by MPCNN, ICCNN, and DeepMedic. It can be appreciated that all of them leave some lesion clusters undetected, and overestimate others. DeepMedic seems to create spurious lesion detection clusters, while our proposal MPCNN false alarms are more of the kind of cluster extensions, or connections between clusters. So, some qualitative differences of the response of the architectures can be appreciated which deserver further analysis and experimentation.

5 Conclusions and Future Work

We have proposed and tested a new 2D/3D CNN architecture for the detection of WMH lesions, which are smaller than other brain lesions (tumours and stroke lesions), lacking the necrotic and inflammation structures. We compare results with two other architectures published in the literature achieving competitive results. Qualitative assessment of the results, shows some advantage of our approach, which is closer to the manual segmentation in the sense that follows more closely the delineated voxel clusters, and creates less spurious detection clusters. The combination of 2D and 3D input windows allows to process the long distance spatial relations, while reducing the computational burden. Ongoing work improves the validation process computing a more complete cross-validation procedure, and more datasets will be included in the experiment. Our proposal may be also subject to changes in kernel parameters and other features of the CNN. Notice that no postprocessing to remove false alarms is done, contrary to DeepMedic, so additional work in postprocessing MPCNN results may provide enhanced results. In order to go ahead in this research area, we made the code available in github so that everyone can contribute to it.

References

1. Admiraal-Behloul, F., van den Heuvel, D., Olofsen, H., van Osch, M., van der Grond, J., van Buchem, M., Reiber, J.: Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. *NeuroImage* **28**(3), 607–617 (2005). <http://www.sciencedirect.com/science/article/pii/S105381190500460X>
2. de Brébisson, A., Montana, G.: Deep neural networks for anatomical brain segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 20–28, June 2015
3. Brosch, T., Yoo, Y., Tang, L.Y.W., Li, D.K.B., Traboulsee, A., Tam, R.: Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, Part III. LNCS, vol. 9351, pp. 3–11. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_1](https://doi.org/10.1007/978-3-319-24574-4_1)
4. Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in Neural Information Processing Systems, pp. 2843–2851 (2012)
5. Debette, S., Markus, H.S.: The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ* **341**, c3666 (2010). <http://www.bmj.com/content/341/bmj.c3666>
6. Erihov, M., Alpert, S., Kisilev, P., Hashoul, S.: A cross saliency approach to asymmetry-based tumor detection. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) MICCAI 2015, Part III. LNCS, vol. 9351, pp. 636–643. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4_76](https://doi.org/10.1007/978-3-319-24574-4_76)
7. Gao, X.W., Hui, R., Tian, Z.: Classification of ct brain images based on deep learning networks. *Comput. Methods Programs Biomed.* **138**, 49–56 (2017)
8. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage* **57**(2), 378–390 (2011). <http://www.sciencedirect.com/science/article/pii/S1053811911003740>
9. Grueter, B.E.: S.U.G.: age-related cerebral white matter disease (leukoaraiosis): a review. *Postgrad. Med. J.* **88**, 79–87 (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. *Med. Image Anal.* **35**, 18–31 (2017). <http://www.sciencedirect.com/science/article/pii/S1361841516300330>
11. Iorio, M., Spalletta, G., Chiapponi, C., Luccichenti, G., Cacciari, C., Orfei, M.D., Caltagirone, C., Piras, F.: White matter hyperintensities segmentation: a new semi-automated method. *Front. Aging Neurosci.* **5**(76) (2013). http://www.frontiersin.org/aging_neuroscience/10.3389/fnagi.2013.00076/abstract
12. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017)
13. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
14. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
15. Murray, A., Staff, R., Shenkin, S., Deary, I., Starr, J., Whalley, L.: Brain white matter hyperintensities: relative importance of vascular risk factors in nondemented elderly people. *Radiology* **237**, 251–257 (2005)

16. Payne, M.E., et al.: Development of a semi-automated method for quantification of MRI gray and white matter lesions in geriatric subjects. *Psychiatry Res. Neuroimaging* **115**(1), 63–77 (2002)
17. Pelletier, A., Periot, O., Dilharreguy, B., Hiba, B., Bordessoules, M., Chanraud, S., Pérés, K., Amieva, H., Dartigues, J., Allard, M., Catheline, G.: Age-related modifications of diffusion tensor imaging parameters and white matter hyperintensities as inter-dependent processes. *Front. Aging Neurosci.* **7**(255) (2016). http://www.frontiersin.org/aging_neuroscience/10.3389/fnagi.2015.00255/abstract
18. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Handels, H. (eds.) *BrainLes 2015. LNCS*, vol. 9556, pp. 131–143. Springer, Cham (2016). doi:[10.1007/978-3-319-30858-6_12](https://doi.org/10.1007/978-3-319-30858-6_12)
19. Price, C., Mitchell, S., Brumback, B., Tanner, J., Lamar, I.S.M., Giovannetti, T., Heilman, K., Libon, D.: MRI-leukoaraiosis thresholds and the phenotypic expression of dementia. *Neurology* **79**(8), 734–740 (2012)
20. Schwarz, C., Fletcher, E., DeCarli, C., Carmichael, O.: Fully-automated white matter hyperintensity detection with anatomical prior knowledge and without FLAIR. *Inf. Process. Med. Imaging* **21**, 239–251 (2009). Proceedings of the Conference
21. Tuladhar, A.M., van Dijk, E., Zwiers, M.P., van Norden, A.G., de Laat, K.F., Shumskaya, E., Norris, D.G., de Leeuw, F.E.: Structural network connectivity and cognition in cerebral small vessel disease. *Hum. Brain Mapp.* **37**(1), 300–310 (2016). <http://dx.doi.org/10.1002/hbm.23032>
22. Tustison, N., Wintermark, M., Durst, C., Avants, B.: Ants and árboles. In: *MICCAI BraTS Workshop*. Miccai Society, Nagoya (2013)
23. Uchiyama, Y., Kunieda, T., Hara, T., Fujita, H., Ando, H., Yamakawa, H., Asano, T., Kato, H., Iwama, T., Kanematsu, M., Hoshi, H.: Automatic segmentation of different-sized leukoaraiosis regions in brain MR images. In: *Proceedings of SPIE*, vol. 6915, pp. 69151S-1–69151S-8 (2008). <http://dx.doi.org/10.1117/12.770045>
24. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. In: *MICCAI BraTS (Brain Tumor Segmentation) Challenge*. Proceedings, Winning Contribution, pp. 31–35 (2014)
25. Yoshita, M., Fletcher, E., Harvey, D., Ortega, M., Martinez, O., Mungas, D.M., Reed, B.R., DeCarli, C.S.: Extent and distribution of white matter hyperintensities in normal aging, MCI, and AD. *Neurology* **67**(12), 2192–2198 (2006). <http://www.neurology.org/content/67/12/2192.abstract>
26. Zikic, D., Ioannou, Y., Brown, M., Criminisi, A.: Segmentation of brain tumor tissues with convolutional neural networks. In: *Proceedings MICCAI-BRATS*, pp. 36–39 (2014)