Stacked Models for Efficient Annotation of Brain Tissues in MR Volumes

Fabio Aiolli[1], Michele Donini[1], Enea Poletti[2] and Enrico Grisan[2]

[1] Department of Mathematics, University of Padova, Padova, Italy
[2] Department of Information Engineering, University of Padova, Padova, Italy

Abstract— **Magnetic resonance imaging (MRI) allows the acquisition of high-resolution images of the brain. The diagnosis of various brain illnesses is supported by the distinguished analysis of the different kind of brain tissues, which imply their segmentation and classification. Brain MRI is organized in volumes composed by millions of voxels (at least 65.536 per slice, for at least 50 slices), hence the problem of labeling of brain tissue classes in the composition of atlases and ground truth references, which are needed for the training and the validation of machine-learning methods employed for brain segmentation. We propose a stacking classification scheme that does not require any other anatomical information to identify the 3 classes, gray matter (GM), white matter (WM) and Cerebro-Spinal Fluid (CSF). We employed two different MR sequences: fluid attenuated inversion recovery (FLAIR) and double inversion recovery (DIR). The former highlights both gray matter (GM) and white matter (WM), the latter highlights GM alone. Features are extracted using a local multi-scale texture analysis, computed for each pixel of the DIR and FLAIR sequences. The 9 textures considered are average, standard deviation, kurtosis, entropy, contrast, correlation, energy, homogeneity, and skewness, evaluated on a neighborhood of 3x3, 5x5, and 7x7 pixels. A stacked classifier is proposed exploiting the a priori knowledge about DIR and FLAIR features. Results highlight a significative improvement in classification performance with respect to using all the features in a state-of-the-art single classifier.**

Keywords— **Stacked learning, Magnetic Resonance Imaging, Annotation, Textures, Brain**

I. Introduction

Magnetic resonance imaging (MRI) allows the acquisition of high-resolution images of the brain. The diagnosis of various brain illnesses, is supported by the distinguished analysis of the white matter (WM), gray matter (GM) and cerebro-spinal fluid (CSF). In this work we present a semi-supervised method to segment WM, GM, and CSF from MRI data that combines DIR and FLAIR scans, without exploiting any anatomical a priori information, and with the specific objective of preserving the lesions belonging to their correct tissue.

There exist widely available and commonly used brain tissue segmentation software, such as the segmentation tool in SPM [1] and FAST in FSL [2], which use both intensity and a priori anatomic information. However, having been designed for general use, they are not necessarily optimized for specific pulse sequences or for application to images from patients with a specific disease. For example, as observed in [3, 4], when used to segment MR images of MS patients, these tools occur in misclassification of MS lesions as GM due to overlapping intensities, which then requires time-consuming manual editing and introduces operator variability into the measurements.
Hence, manual delineation remains the *gold standard* procedure in studies where brain segmentation of MR data sets is required, especially when dealing with specific populations (e.g. [5, 6, 7, 8, 9, 10]). However, it is expert dependent, observer demanding and time consuming, and essentially not transferable. Automated techniques are necessary to overcome these obstacles, especially when large cohorts of data sets are involved [5]. Moreover, given the growing interest in translational studies in neuroscience, the need for building annotated *gold standard* segmentation and atlases on non human data ([11, 12]) has further stressed the demand for automatic techniques or fast annotation methods.

We propose a supervised classification method that exploits the texture information of the brain tissue provided by the two sequences FLAIR and DIR [13] (1). The former is characterized by the suppression of CFS and by the consequent enhancement of both GM and WM, which are however difficult to be distinguished one from the other; the latter has two inversion recoveries that allow suppressing the contribution of both CSF and WM, thus enhancing GM. We thus avoid using T1-w sequence, which, even if characterized by high spatial resolution, proved inadequate for tissue segmentation when brain lesions are present. In addition, the method does not need population-derived location-based priors, registration to template space, or explicit bias field modeling.

II. Material

Twenty-four slices (256×256), from $z = 20$ to $z = 44$, from both DIR and FLAIR sequences acquired on a patient

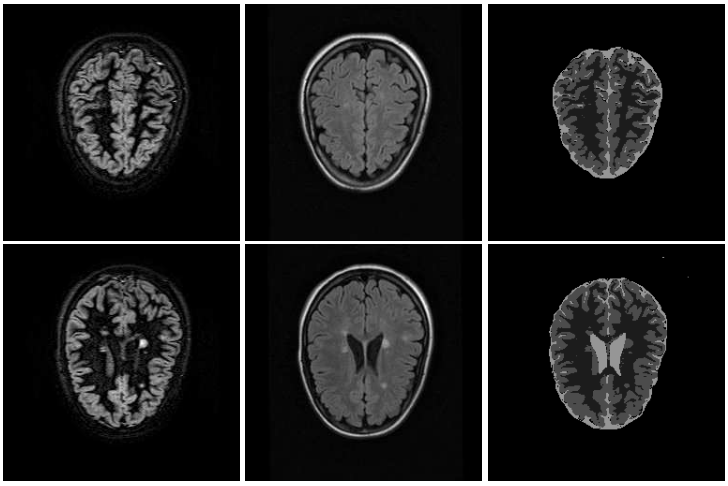

Fig. 1: (Leftmost column) two slices of the DIR sequence; (central column) the corresponding FLAIR slices; (rightmost column) manual ground truth segmentations provided by the experts

affected by MS [14], have been taken into account. In each slide, the three classes GM, WM and CSF were manually labeled.

III. Methods

A. Feature Extraction

The rationale of the feature extraction approach is to use the peculiar texture characteristic of a pixel neighborhood in order to obtain information about the pixel tissue class [13]. Image texture analysis has been subject of intense study and has been employed in a variety of applications; however, there is no general agreement upon definition of texture. For our specific application, we assume that a region in an image has a constant texture if a set of local statistics or other local properties of the picture function are almost constant.

Depending on the number of pixels defining the local feature, the statistical methods can be respectively classified as 1st-order, 2nd-order and higher-order statistics. 1^{st}-order statistics measure the likelihood of observing a specific gray value at a random location in the image (hence directly computable from the image histogram). 2^{nd}-order statistics measure the likelihood of observing a specific pair of gray values in a randomly placed dipole of pixels (computable from the gray level co-occurrence matrices (GLCM) [15]). Method proposed in [16] employed three 1st-order statistics: skewness-, median-, and median absolute deviation-based textures, which, on T1-w images, are approximately independent of bias field and of scanner gain. In order to increase the discriminability of the classes, and at the same time couple at best with the double source of information at disposal (i.e., the DIR and the FLAIR sequences), we opted to employ as features four 1^{st}-order statistics and five 2^{nd}-order statistics. The 1^{st}-order statistics considered in this work are mean, standard deviation, skewness, and kurtosis, while the 2^{nd}-order ones are contrast, correlation, homogeneity, entropy and energy.

We extract the local texture information at 3 different scales from blocks of $N \times N$ pixel, with N = 3, 5, and 7. For each 2^{nd}-order texture, 4 GLCMs are constructed, with $d = (d_x, d_y) \in \{(0,1),(1,1),(1,0),(1,-1)\}$. Then, to make the textures invariant to rotation, the obtained matrices are averaged over the 4 angles. Since the feature extraction is performed on both DIR and FLAIR images, the final feature vector associated to each pixel is composed by 56 values (2 original sequence pixel values, plus 9 textures $\times$ 3 scales $\times$ 2 sequences).

B. Classification

State-of-the-art algorithms typically cast the multi-class problem of classifying CSF, GM and WM with a *one against all* technique. The classifier system consists of three binary classifiers. The classifier h_c is trained with all the available labeled examples giving a positive label to examples of class c and a negative label to examples of other classes. All the features for each image (DIR and FLAIR) are used to build these three classifiers. The prediction is made by comparing the scores of the three classifiers and predicting the class whose corresponding classifier maximizes this score.

However, once considering the *a priori* knowledge we have about our specific problem, thinking in this way can be counter intuitive. In fact, using the approach described above,

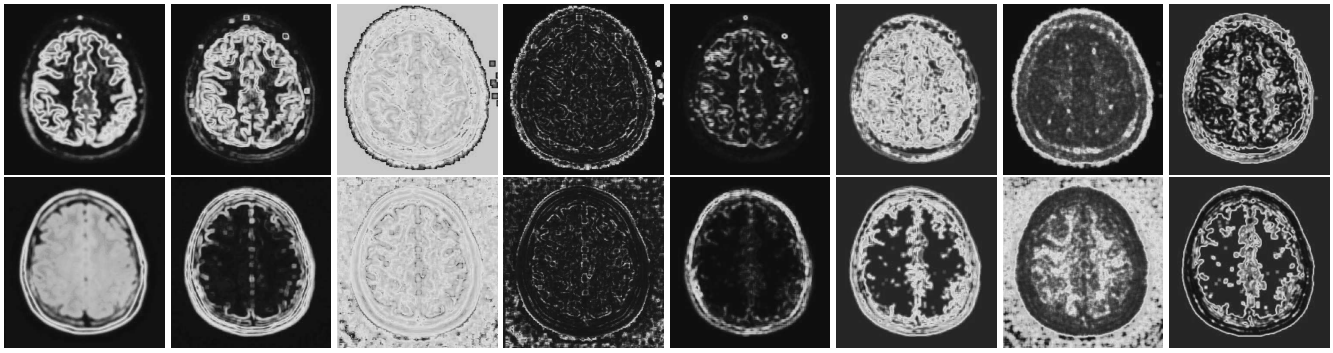

Fig. 2: *Upper row:* 8 textures extracted from the DIR slice. *Lower row:* 8 textures extracted from the FLAIR slice. *Columns, from left to right:* mean, standard deviation, skewness, kurtosis, contrast, homogeneity, entropy, energy. The scale employed is 5 (5×5 pixel block analysis). Original DIR and FLAIR slices are the one in Fig. 1; the slice is at $z = 33$.

we are not exploiting the fact that different types of features (DIR and FLAIR) contain diverse information and each one is naturally tailored to a more specific task. FLAIR based features highlight both gray matter (GM) and white matter (WM), then they can be more useful to discriminate CSF and not-CSF (GM or WM). On the other side, DIR based features highlight GM alone and thus can be more useful to discriminate between GM and WM when we already know that a particular voxel is not of class CSF.

Here, we consider a *two-step stacked* system that creates, in a first step, a binary classifier h_{CSF} using the FLAIR image features only to select the subset of voxel corresponding to CSF. In a second step, it creates a binary classifier h_{WM} that selects the WM from the not-CSF part of voxel. The remaining part of voxels are classified as GM.

C. Experimental Setting

We compared three classification settings using a dataset consisting of ten manually labeled slices. Specifically,

1. One-Against-All (OAA, [17]), where each binary classifier is trained with all features (DIR and FLAIR)
2. Stacked All Features (SAF, [18]), where stacking is performed as described above and all the features are used in both levels
3. Stacked Disjoint Features (SDF), where stacking is performed as described above. FLAIR features used in the first level only and DIR features used in the second level.

An *SVM-like* classifier [19] with *RBF* kernel ($\gamma = 0.01$) has been used for all different settings. On each experiment:

1. We randomly select few (5) labelled examples for each class (CSF, GM and WM).
2. We train all SVM binary classifiers using these labeled data as training data.
3. We classify all the unlabeled data.

We have repeated the steps above for 1000 times to increase the significance of our experiments and we calculated the average accuracy and standard deviation.

IV. Results

Results for all the three settings are summarized in the following table.

Algorithm	Accuracy	StdDev
One against all	68.821%	0.05243
Two-step (all the features)	68.870%	0.04961
Two-step (*a priori* knowledge)	70.403%	0.05742

We can see that our two-step stacking algorithm has a significantly better accuracy than state-of-the-art methods. We have also demonstrated that a significant improvement can be obtained by using *a priori* knowledge on the task at hand. Moreover, the baseline (OAA) method requires the training of three binary classifiers and all the features, while our two-step algorithm needs only two binary classifiers, each one working with a halved number features. The proposed approach provided both better results and better computational performance.

In order to support further our proposal on how to deal with this kind of *a priori* knowledge, another experiment was performed reversing the order of the two classifiers in our two-step algorithm. As expected, we obtained a significant decrease in performance in this case.

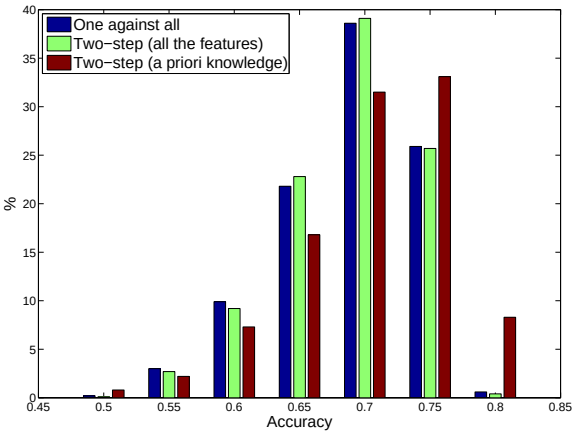

Fig. 3: Histogram of accuracy

V. Discussion and Conclusion

We have shown an effective way to inject a priori knowledge about the different nature of MR sequences in a stacking model for brain segmentation. In the future, we plan to improve our two-step algorithm in two principal ways. Firstly, by exploiting the existing topology among voxels given by their physical closeness. For this task we can create a graph representation of the brain containing the topological information. Preliminary experiments have shown an improvement in the results even using horizontal topological information only. Secondly, we also plan to study *active learning* algorithms to guide the initial selection of manual labeling. For example, we could study an active learning algorithm that selects the best voxels for manual labeling. The histogram in Fig. 3 shows how the accuracy strongly depends on the initial choice of voxels. Interestingly, in our two-step algorithm, we observe a larger number of cases in which the accuracy is over 80% with respect to the other methods. So we hope to give active learning algorithms able to choose the best voxels eligible for the training set, in an unsupervised manner.

References

1. Ashburner J., Friston K.J.. Unified segmentation *NeuroImage.* 2005;26:839-851. cited By (since 1996) 1330.
2. Smith S.M., Jenkinson M., Woolrich M.W., et al. Advances in functional and structural MR image analysis and implementation as FSL *NeuroImage.* 2004;23:S208-S219. cited By (since 1996) 1935.
3. Nakamura K., Fisher E.. Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients *NeuroImage.* 2009;44:769-776. cited By (since 1996) 22.
4. Gelineau-Morel R., Tomassini V., Jenkinson M.. The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis *Hum Brain Mapp.* 2011. cited By (since 1996) 4.
5. Gousias Ioannis S, Rueckert Daniel, Heckemann Rolf a, et al. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *NeuroImage.* 2008;40:672–84.
6. Shi Feng, Yap Pew-Thian, Fan Yong, Gilmore John H, Lin Weili, Shen Dinggang. Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. *NeuroImage.* 2010;51:684–93.
7. Tang Yuchun, Hojatkashani Cornelius, Dinov Ivo D, et al. The construction of a Chinese MRI brain atlas: a morphometric comparison study between Chinese and Caucasian cohorts. *NeuroImage.* 2010;51:33–41.
8. Habas Piotr a, Kim Kio, Corbett-Detig James M, et al. A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation. *NeuroImage.* 2010;53:460–70.
9. Gholipour Ali, Akhondi-Asl Alireza, Estroff Judy a, Warfield Simon K. Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly. *NeuroImage.* 2012;60:1819–31.
10. Serag Ahmed, Aljabar Paul, Ball Gareth, et al. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *NeuroImage.* 2012;59:2255–65.
11. McLaren Donald G, Kosmatka Kristopher J, Oakes Terrance R, et al. A population-average MRI-based atlas collection of the rhesus macaque. *NeuroImage.* 2009;45:52–9.
12. Calabrese Evan, Badea Alexandra, Watson Charles, Johnson G Allan. A quantitative magnetic resonance histology atlas of postnatal rat brain development with regional estimates of growth and variability. *NeuroImage.* 2013;71C:196–206.
13. Poletti E., Veronese E., Calabrese M., Bertoldo A., Grisan E.. Supervised classification of brain tissues through local multi-scale texture analysis by coupling DIR and FLAIR MR sequences *Proceedings of SPIE*. 2012;8314:83142T.
14. Veronese E., Poletti E., Calabrese M., Bertoldo A., Grisan E.. Unsupervised Segmentation of Brain Tissues using Multiphase Level Sets on Multiple MRI Sequences in *Intelligent Systems and Control/742: Computational Bioscience*ACTA Press 2011.
15. Haralick Robert M.. STATISTICAL AND STRUCTURAL APPROACHES TO TEXTURE. *Proc IEEE*. 1979;67:786-804. cited By (since 1996) 1640.
16. Vovk A., Cox R.W., Stare J., Suput D., Saad Z.S.. Segmentation priors from local image properties: Without using bias field correction, location-based templates, or registration *NeuroImage.* 2010. cited By (since 1996) 1.
17. Allwein Erin L., Schapire Robert E., Singer Yoram. Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers *Journal of Machine Learning Research.* 2000;1:113-141.
18. Cohen William W.. Stacked sequential learning in *International Joint Conference on Artificial Intelligence*:671-676 2005.
19. Aiolli F., Martino G. Da San, Sperduti. A.. A Kernel Method for the Optimization of the Margin Distribution (Prague, Czech Republic) 2008.

Author: Fabio Aiolli
Institute: Dept of Mathematics, University of Padova
Street: Via Trieste, 63
City: Padova
Country: Italy
Email: aiolli@math.unipd.it