An Hippocampal Segmentation Tool Within an Open Cloud Infrastructure

Nicola Amoroso[1,2(✉)], Sabina Tangaro[1], Rosangela Errico[2,3], Elena Garuccio[4], Anna Monda[2], Francesco Sensi[3], Andrea Tateo[2], and Roberto Bellotti[1,2] for the Alzheimer's Disease Neuroimaging Initiative

[1] Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Bari, Italy
nicola.amoroso@ba.infn.it
[2] Dipartimento Interateneo di Fisica, Università Degli Studi di Bari, Bari, Italy
[3] Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Genova, Italy
[4] Dipartimento di Fisica, Università Degli Studi di Siena, Siena, Italy

Abstract. This study presents a fully automated algorithm for the segmentation of the hippocampus in structural Magnetic Resonance Imaging (MRI) and its deployment as a service on an open cloud infrastructure. Optimal atlases strategies for multi-atlas learning are combined with a voxel-wise classification approach. The method efficiency is optimized as training atlases are previously registered to a data driven template, accordingly for each test MRI scan only a registration is needed. The selected optimal atlases are used to train dedicated random forest classifiers whose labels are fused by majority voting. The method performances were tested on a set of 100 MRI scans provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI). Leave-one-out results (Dice $= 0.910 \pm 0.004$) show the presented method compares well with other state-of-the-art techniques and a benchmark segmentation tool as FreeSurfer. The proposed strategy significantly improves a standard multi-atlas approach ($p < .001$).

Keywords: Segmentation · Quantitative image analysis · Imaging biomarkers · Magnetic resonance imaging · Machine learning

1 Introduction

The "Smart cities and communities and social innovations" national operative programs have outlined the need for an efficient reorganization of health-care

Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 193–200, 2015.
DOI: 10.1007/978-3-319-23222-5_24

both to ensure higher standards in terms of quality of life for patients and to rationalize the economic resources to be allocated. Accordingly, the connected health and the e-health technologies can be considered pillars of an innovative smart thinking of cities and communities. Hippocampal atrophy is an established bio-marker for several neurodegenerative diseases, such as the Alzheimer's disease [7], a disease characterized by an impressive social and economic impact. However, no segmentation tool is currently employed in clinical practice, especially because computational requirements of best performing algorithms such as [10,11] are difficult to fulfill.

With this aim, in this paper we present a novel machine learning tool for hippocampal segmentation which has been proven to yield consistent improvements with respect of recent studies [12]. In particular, the proposed segmentation workflow for the human hippocampus and its deployment *as a Service*, on the PRISMA cloud[1] which exploits the Bari ReCaS [2] computer center, are described. Both PRISMA and ReCaS are national operative programs, the first in particular is a smart city program dealing with the development of Open Source platforms for computing solutions dedicated to e-Health or e-Government, just to mention a few.

The proposed approach efficiently exploits the cloud computational resources requiring only a linear registration followed by a warp to segment a test image. After registration optimal atlases are adaptively selected. First, a shape analysis algorithm is used to detect peri-hippocampal volumes of interest (VOIs). Then, the optimal atlases are selected by measuring the pairwise Pearson's correlation and they are used to train supervised classifiers. The leave-one-out performances of the methodology are compared with the publicly available segmentation tool FreeSurfer [8] and a basic multi-atlas pipeline, *i. e.* consisting of registration and label fusion, showing a significant improvement.

2 Materials and Methods

A data set of 100 T1 MRI scans from the ADNI database, including 29 normal controls (NC), 34 mild cognitive impairment (MCI) and 37 Alzheimer's disease (AD) subjects, has been used in preparation of this article. The set is composed by male and female subjects aged between 60 and 90 years old. The relative hippocampal labelings were provided by the EADC-ADNI harmonized segmentation protocol[3] [4,5]. The ADNI set consists of MPRAGE MRI brain scans with a resolution of $1 \times 1 \times 1$ mm^3. According to this, in the following, voxels or mm^3 will be interchangeably used without further specifications.

2.1 Increasing Inter-subject Similarity

Registration processes are sensitive to initial conditions, accordingly the intensities of MRI scans are normalized and the bias field removed with the improved

[1] http://www.ponsmartcities-prisma.it
[2] http://www.pon-recas.it
[3] www.hippocampal-protocol.net

N3 MRI bias field correction algorithm [14]. The MRI scans are co-registered with the MNI152 template with the FSL libraries [9] and the warp fields $\mathcal{F}_i$ are stored for later use. The goal of this processing is to maximize inter-subject similarity in order to help the classifiers to learn the disease patterns. The proposed algorithm is schematically represented in Fig. 1.

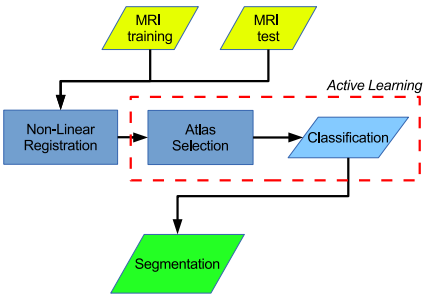

Fig. 1. A schematic overview of the proposed method: 1) non-linear registration, 2) Atlas selection and 3) classification with the latter two phases encompassed in a unique active learning framework.

After registration a gross peri-hippocampal region $\omega(VOI)_i$ is extracted with FAPoD [1], a fully automated hippocamapal shape analysis algorithm, from each scan. The $\omega(VOI)_i$ extracted by FAPoD contains a probable hippocampal region of about 17000 voxels distributed in a rectangular volume of interest of dimensions $50 \times 70 \times 70$ voxels. $\omega(VOI)_i$ are used for the atlas selection.

At this point, the data set is divided into a training subset $\mathcal{D}_t$ and a validation MRI scan a_v to perform a leave-one-out analysis. We use the Pearson's correlation to directly measure the similarity between the peri-hippocampal regions $\omega(VOI)_i$ of $\mathcal{D}_t$ and the $\omega(VOI)_v$ of the validation scan a_v. Pearson's correlation is then used as a ranking score to detect the first k optimal atlases. Machine learning approaches usually use training examples to build a shared base of knowledge and then learn a generalized model, on the contrary we adopt a substantial change of perspective. We try to learn different patterns from each $\omega(VOI)_i$, then, through image processing, we make the validation scan as similar as possible to training examples. Finally, we actively select the k most representative examples and use them for prediction, thus requiring the validation sample to reproduce these training patterns.

2.2 Classification and Segmentation

Each hippocampal $\omega(VOI)_i$ undergoes a voxel-wise feature extraction process which assigns to each voxel a set of 315 features [13]. Each voxel is represented as a vector whose elements represent information about position, intensity and texture. Texture information (contrast, uniformity, rugosity, regularity, etc.) is

expressed using Haar-like and some Haralick features. With this set of features, we train N random forest classifiers $\mathcal{C}_{\{1,\ldots,N\}}$, the labels for each training scan being the manual tracings of expert neuroradiologists.

For each validation scan we perform atlas selection as previously described in 2.1, then we pick the k models $\mathcal{C}_i$ of the optimal atlases and perform a voxel-wise prediction. For each voxel of the $\omega(VOI)_v$ the relative label is calculated as a weighted average of the k predicted v_k labels, the weight being the pairwise distance between the selected atlases and the target image. Finally, the inverse warps $\mathcal{F}_v^{-1}$ is applied and a 0.5 threshold is adopted to obtain a binary segmentation. The classification performances are measured in terms of Dice similarity index D and standard error ϵ defined as:

$$D = \frac{2\,|A \cap B|}{|A| + |B|} \tag{1}$$

$$\epsilon = \frac{\sigma}{N} \tag{2}$$

where A and B represent the regions being compared, cardinalities $|A|$, $|B|$ are intended as the measured volumes, σ being the standard deviation of the Dice distribution and $N = 100$ is the sample size.

2.3 Cloud Deployment as a Service

The field of medical imaging has seen in recent years an enormous development. Image databases, made of thousands of medical images, are currently available to supply clinical diagnosis, this is particularly true for brain diseases [2]. Medical image processing applications would greatly take advantage from open clouds deployment: run-time reduction, sharing of data collections and platform-hardware independent configurations are just a few examples.

The proposed algorithm requires an overall CPU time of 40 ± 10 minutes per scan. When dealing with large data sets this can represent a too much expensive computational cost to afford. Accordingly, cloud technology can tackle these computational issues by providing a user dedicated computing environment. Besides, offering the segmentation tool as a service (submitting images to be segmented to the related web portal[4]) can help its clinical adoption as no technical background is required to use it.

With this purpose the segmentation pipeline presented in this work has been encased within a virtualized wrapping framework, to fully automate not only the job submission and monitoring, but also the resource exploitation. In this way the segmentation pipeline can be accessed as a pure Software as a Service.

Once the end user proceeds to upload a brain scan to be segmented, a job management tool JST (Job Submission Tool) manages the submission and monitoring of the application. JST monitors the submission of all of jobs required by a given application thus hiding to the end user the complexity of operating in a heterogeneous and distributed computational environment. Moreover, JST is

[4] https://recasgateway.ba.infn.it

portable on different infrastructures like the EGI grid infrastructure, dedicated servers, local batch farms, IaaS/SaaS based cloud resources and as a consequence it allows to efficiently exploits the computational resources needed by the application.

3 Results

Active learning performances are stable, in fact they are not affected by the number of chosen atlases (no significant $p > .05$ difference can be found using $10 \sim 15$ atlases or more) as shown in Fig. 2. Statistical significance is assessed by means of non parametric Kruskal-Wallis test.

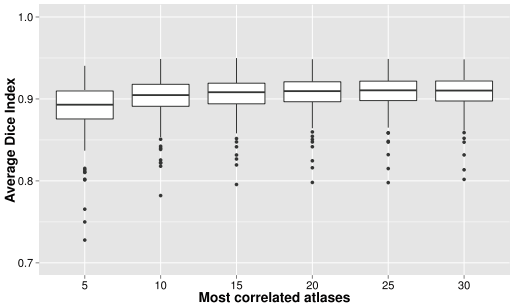

Fig. 2. The figure shows the Dice distribution obtained by averaging the corresponding left and right hippocampal performances and varying the number of selected atlases. Performances reach a plateau beyond $10 \sim 15$ atlases.

An analogous result shows that even basic multi-atlas performances reach a plateau when using ~ 10 atlases and beyond. Accordingly, to assess whether active learning can improve basic multi-atlas performance, with a fair comparison, we use 15 atlases for both cases. The proposed method improves the overall performances obtained by basic multi-atlas and FreeSurfer (see Fig. 3).

In fact, for left hippocampi median Dice index with the relative standard error is 0.908 ± 0.004, for right hippocampi 0.912 ± 0.003. Basic multi-atlas and FreeSurfer respectively achieve 0.845 ± 0.005 and 0.728 ± 0.005 for left hippocampi and 0.851 ± 0.004 and 0.733 ± 0.004 for right hippocampi. A Kruskal-Wallis test demonstrates the three Dice distributions are significantly different ($p < .001$).

Dice metric has an important drawback, in fact, it does not distinguish between false positive and false negative errors. As a consequence, two distinct segmentations can obtain the same Dice index performance, even if reproducing the manual tracing in one case with an excess of false positives, in the other with false negatives. This is why it is also important to perform an "agreement" measure, for example with a Bland-Altman analysis [3].

For the present work, we perform a Bland Altman analysis of the standardized manual and segmentation volumes, for both left and right hippocampi. The

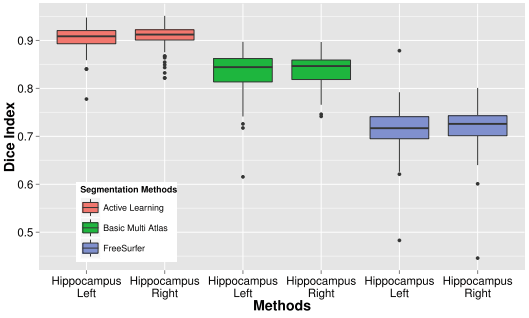

Fig. 3. The figure shows the Dice index performances varying with the different segmentation methods for both left and right hippocampi. Active learning performances (red) are significantly higher than those obtained through basic multi-atlas (green) or FreeSurfer segmentations (blue).

analysis confirms that less than 5% of standardized differences between segmented and manual volumes exceed the 95% confidence bounds, so that they can be considered statistical significant Fig. 4.

The correlation between segmented and manual volumes is 0.80 and 0.84 for respectively left and right hippocampal volumes. Therefore, active learning seems also to improve the agreement between manual and automated segmentations.

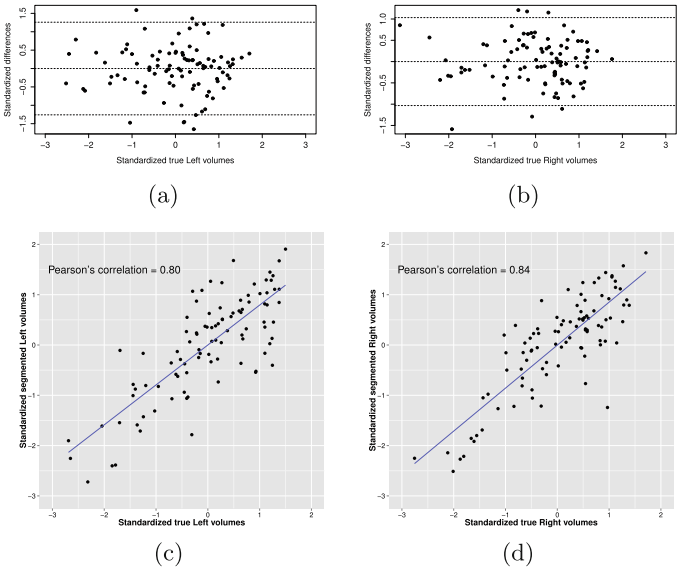

Fig. 4. The figure shows the Bland Altman analysis (measure agreement and correlation) for standardized volumes. Results for both left (a,c) and right (b,d) hippocampi are separately shown.

4 Conclusion and Future Work

In this study we present a novel segmentation algorithm based on a combined multi-atlas and machine learning strategy. A key role on the method is played by atlas selection. We select optimal atlases according to Pearson's correlation measurements between VOIs automatically detected to specifically contain the hippocampus. The performances obtained respectively for left and right hippocampi are 0.908 ± 0.004 and 0.912 ± 0.003.

This work demonstrates how active learning strategies, such as those presented, can bring substantial performance improvements. Nevertheless, which similarity metric to use should be further investigated. In fact, other similarity measurements, especially non linear techniques, such as Locally Linear embedding and Laplacian Eigenmaps, could be adopted for atlas selection. Besides, recent works suggest the use of warping fields for similarity measurements, accordingly a fair comparison should be performed.

It is worth noting that the method is computationally efficient, requiring a processing time of about 10 minutes per test scan. Moreover, the exploitation of cloud infrastructures potentially suggest it could be adopted for large clinical trials. With this regard, a limitation of the study is due to the absence of a clinical evaluation, even if the goal of this work lies far from this aspect. Future work will investigate how structural hippocampal properties obtained with this method can improve Alzheimer's disease diagnosis [6].

Acknowledgments. Nicola Amoroso, Rosangela Errico and Andrea Tateo acknowledge funding by the Italian MIUR grant PON PRISMA Cod. PON04a2_A from Università degli Studi di Bari, Italy. This research was also supported by Istituto Nazionale di Fisica Nucleare (INFN), Italy.

References

1. Amoroso, N., Bellotti, R., Bruno, S., Chincarini, A., Logroscino, G., Tangaro, S., Tateo, A.: Automated Shape analysis landmarks detection for medical image processing. In: Proceedings of the International Symposium, CompIMAGE (2012)
2. Bellotti, R., Pascazio, S.: Editorial: Advanced physical methods in brain research. The European Physical Journal Plus **127**(11), 145 (2012)
3. Bland, J.M., Altman, D.G.: Comparing methods of measurement: why plotting difference against standard method is misleading. The Lancet **346**(8982), 1085–1087 (1995)
4. Boccardi, M., Bocchetta, M., Apostolova, L.G., Barnes, J., Bartzokis, G., Corbetta, G., DeCarli, C., Firbank, M., Ganzola, R., Gerritsen, L., et al.: Delphi definition of the EADC-ADNI Harmonized Protocol for hippocampal segmentation on magnetic resonance. Alzheimer's & Dementia **11**(2), 126–138 (2015)
5. Boccardi, M., Bocchetta, M., Morency, F.C., Collins, D.L., Nishikawa, M., Ganzola, R., Grothe, M.J., Wolf, D., Redolfi, A., Pievani, M., et al.: Training labels for hippocampal segmentation based on the eadc-adni harmonized hippocampal protocol. Alzheimer's & Dementia **11**(2), 183–191 (2015)

6. Bron, E.E., Smits, M., van der Flier, W.M., Vrenken, H., Barkhof, F., Scheltens, P., Papma, J.M., Steketee, R.M.E., Orellana Méndez, C., Meijboom, R., Pinto, M., Meireles, J.R., Garrett, C., Bastos-Leite, A.J., Abdulkadir, A., Ronneberger, O., Amoroso, N., Bellotti, R., Cárdenas-Peña, D., Álvarez Meza, A.M., Dolph, C.V., Iftekharuddin, K.M., Eskildsen, S.F., Coupé, P., Fonov, V.S., Franke, K., Gaser, C., Ledig, C., Guerrero, R., Tong, T., Gray, K.R., Moradi, E., Tohka, J., Routier, A., Durrleman, S., Sarica, A., Di Fatta, G., Sensi, F., Chincarini, A., Smith, G.M., Stoyanov, Z.V., Sørensen, L., Nielsen, M., Tangaro, S., Inglese, P., Wachinger, C., Reuter, M., van Swieten, J.C., Niessen, W.J., Klein, S.: Standardized evaluation of methods for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. NeuroImage (in press)
7. Chincarini, A., Bosco, P., Gemme, G., Esposito, M., Rei, L., Squarcia, S., Bellotti, R., Minthon, L., Frisoni, G., Scheltens, P., et al.: Automatic temporal lobe atrophy assessment in prodromal ad: Data from the descripa study. Alzheimer & Dementia 1, 12 (2013)
8. Fischl, B.: FreeSurfer. NeuroImage **62**(2), 774–781 (2012)
9. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. NeuroImage **62**(2), 782–790 (2012)
10. Kim, M., Wu, G., Li, W., Wang, L., Son, Y.D., Cho, Z.H., Shen, D.: Automatic hippocampus segmentation of 7.0 Tesla MR images by combining multiple atlases and auto-context models. NeuroImage **83**, 335–345 (2013)
11. Lotjonen, J.M.P., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage **49**(3), 2352–2365 (2010)
12. Tangaro, S., Amoroso, N., Boccardi, M., Bruno, S., Chincarini, A., Ferraro, G., Frisoni, G., Maglietta, R., Redolfi, A., Rei, L., Bellotti, R.: Automated voxel-by-voxel tissue classification for hippocampal segmentation: Methods and validation. Physica Medica **30**(8), 878–887 (2014)
13. Tangaro, S., Amoroso, N., Brescia, M., Cavuoti, S., Chincarini, A., Errico, R., Inglese, P., Longo, G., Maglietta, R., Tateo, A., Riccio, G., Bellotti, R.: Feature Selection Based on Machine Learning in MRIs for Hippocampal Segmentation. Computational and Mathematical Methods in Medicine, Article ID(814104) (in press)
14. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE Transactions on Medical Imaging **29**(6), 1310–1320 (2010)