Automatic Hippocampus Labeling Using the Hierarchy of Sub-region Random Forests

Lichi Zhang[1], Qian Wang[1], Yaozong Gao[2,3], Guorong Wu[3], and Dinggang Shen[3(✉)]

[1] School of Biomedical Engineering, Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China
{lichizhang, wang.qian}@sjtu.edu.cn
[2] Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, USA
yzgao@cs.unc.edu
[3] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, USA
{grwu, dgshen}@med.unc.edu

Abstract. In this paper, we propose a multi-atlas-based framework for labeling hippocampus regions in the MR images. Our work aims at extending the random forests techniques for better performance, which contains two novel contributions: *First*, we design a novel strategy for training forests, to ensure that each forest is specialized in labeling the certain sub-region of the hippocampus in the images. In the testing stage, a novel approach is also presented for automatically finding the forests relevant to the corresponding sub-regions of the test image. *Second*, we present a novel localized registration strategy, which further reduces the shape variations of the hippocampus region in each atlas. This can provide better support for the proposed sub-region random forest approach. We validate the proposed framework on the ADNI dataset, in which atlases from NC, MCI and AD subjects are randomly selected for the experiments. The estimations demonstrated the validity of the proposed framework, showing that it yields better performances than the conventional random forests techniques.

1 Introduction

Accurate hippocampus labeling in the Magnetic Resonance (MR) brain images is a task of pivotal importance to the researches of many neural diseases including the Alzheimer's disease, schizophrenia and epilepsy [1]. The approaches are highly demanded, since it is infeasible to manually label a large set of 3D MR images. The main challenge in the segmentation of the hippocampus is that its grey intensity has close similarity to the surrounding region-of-interests (ROIs), such as amygdala, caudate nucleus, and thalamus [2]. Recent developments in this field concentrate on utilizing the information of manually labeled atlas images for the estimation of the test image. Currently a popular way among those techniques is the multi-atlas label propagation (MALP),

G. Wu et al. (Eds.): Patch-MI 2015, LNCS 9467, pp. 19–27, 2015.
DOI: 10.1007/978-3-319-28194-0_3

because of its robustness and simplicity for brain image labeling. Basically, there are two steps in the MALP approaches: (1) implement image registration to spatially align all images, and (2) use manual labels of the atlases to label the test image following certain label fusion strategies. There are many efforts aiming at improving the two steps. The detailed literature can be found in [3–6].

Random forest [7, 8] has been proven as a robust and fast multi-class classifier, which has been widely applied in many applications. A major contribution of the random forest is its bagging strategy, which adds randomness when training the classifiers. Zikic et al. [9] developed an "atlas forest" approach, which encodes each individual atlas and its corresponding label map via random forest. In the testing stage, atlas forests produce individual probabilistic labeling maps for the test image, the average of which is then regarded as the final labeling. Lombaert et al. [10] introduced Laplacian forests, in which the training images are re-organized and embedded into a low-dimensional manifold. Images having close similarities are grouped together. In the training stage, each tree is learned using only specific group of similar atlases following a guided bagging approach. The strategy concerning tree selection for the given test image is also proposed. The method is demonstrated experimentally to yield higher training efficiency and segmentation accuracy.

In this paper, we focus on the high variation across individual sub-regions in the ROI, and present a novel hierarchical learning framework to further improve the labeling performance. In the training stage, we develop a set of random forests, in which each forest is trained for a specific sub-region of the ROI. The random forests are placed into a hierarchical structure, which is derived from the registration-based auto-context technique. Specifically, for a higher level in the hierarchy, the random forests are trained with the context features that are extracted from the outputs of the lower level. Moreover, the lower-level outputs also guide all atlas and test images to be better registered for the higher level. In the testing stage, we select optimal forests for individual sub-regions of the test image. Therefore, the labeling results can be gradually improved by the hierarchy of random forests.

2 Method

In this section, we present the detailed description of the hierarchical learning framework, which consists of the training and the testing stages. The random forests implemented in this work are trained following a novel sub-region labeling strategy, while as in the testing stage a corresponding forest selection approach is applied for predicting the labeling information of the test image. In Sect. 2.1 we elaborate the strategy of sub-region random forests along with the forest selection method. Following the registration-based auto-context strategy, the method of sub-region random forest is extended as a hierarchical structure, in which each level contains a set of forests for image labeling. In Sect. 2.2, we describe the registration-based auto-context model, while as in Sect. 2.3 we present the methodologies of the proposed framework in the training and testing stages, respectively.

2.1 Sub-region Random Forest

The main idea of the sub-region random forest for labeling, as presented in Fig. 1, is that individual sub-regions should be trained and tested with the most suitable classifiers, as the variation across individual sub-regions can be high. In the training stage, we commence by randomly extracting numerous patches from the training images. Note that there is a higher priority of choosing patches in the boundary parts of the hippocampus. Suppose there are m patches selected, the set of 3D cubic patches is denoted as $P = \{p_1, p_2, \ldots, p_m\}$.

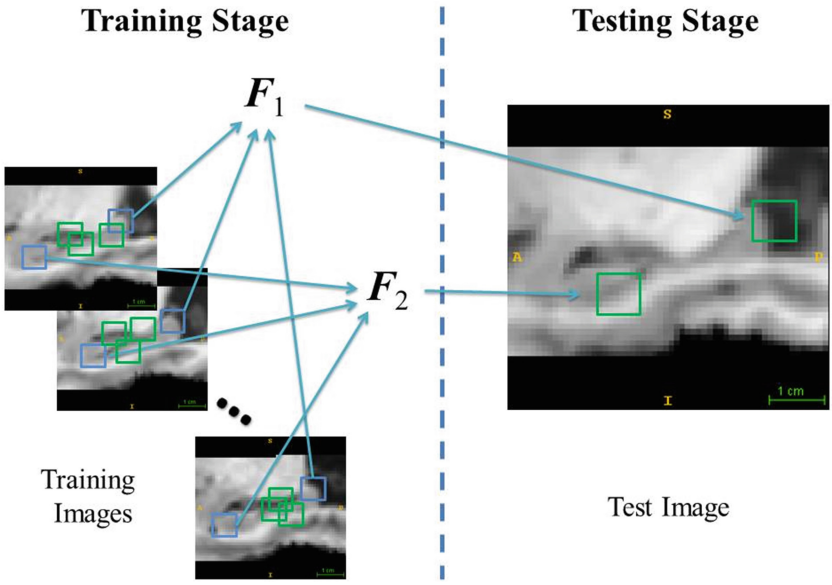

Fig. 1. The process of training and testing sub-region random forests. F_i is the trained forest specialized for labeling the corresponding sub-region in the test image.

Next, we cluster the selected patches based on their intensity similarities, which are measured by the mean of squared intensity differences between each pair of patches, given as: $S(p_i, p_j) = \sum_{x=1}^{r} \sum_{y=1}^{r} \sum_{z=1}^{r} [p_i(x,y) - p_j(x,y)]^2 / r^3$, where r is the patch length. The similarity measures are then utilized by the affinity propagation method [11] for clustering. When similar patches are grouped together, each cluster has its own forest to be trained, which can be used for predicting the labeling result of the corresponding sub-regions in test images.

In the testing stage, we follow the forest selection strategy to find the appropriate classifiers for labeling each voxel in the test image. It is implemented by utilizing a labeling prior (i.e., from tentative labeling output) as the reference. The estimation of the prior is presented in Sect. 2.3. Denoting the test image as I', and the prior as D', the i-th trained forest in the k-th level as F_i^k with their corresponding estimate D_i^k, we also

obtain the overall mask information M_i in each cluster, which is the union of all regions that the patches in that cluster have ever covered. Denote the set of patches in the i-th cluster as $P^i = \{p_1^i, p_2^i, \ldots, p_{m_i}^i\}$, we have: $M_i = \bigcup_{j=1}^{m_i} R(p_j^i)$, where $\mathrm{R}(p_j^i)$ is the region for p_j^i. The novel metric for the testing image is given as follows:

$$\mathrm{W}\left(I^{'}, F_i^k\right) = \mathrm{DSC}\left(D^{'}, D_i^k, M_i\right), \tag{1}$$

where $\mathrm{W}\left(I^{'}, F_i^k\right)$ is the score for the forest selection. $\mathrm{DSC}\left(D^{'}, D_i^k, M_i\right)$ is the function for measuring the Dice overlapping ratio between the two labeling estimates $D^{'}$ and D_i^k, as it only counts the voxels within the mask region M_i. When the forests with the top W scores are selected, their corresponding estimates D_i^k are combined together for producing the labeling result of the test image.

2.2 Registration-Based Auto-Context

Our aim in proposing the registration-based auto-context method is to further improve the robustness and performance of the sub-region random forest. To fulfill this goal, we consider the auto-context model in [13], which enables the higher-level forests to encode more comprehensive information provided by the lower-level classifiers. Following the auto-context model, the sub-region random forest approach is extended to be a hierarchical structure framework. In each level of the hierarchy, the forests are trained using *not only* the image appearance features from the training images, *but also* the context features computed from the estimates of the lower-level forests. The method is proved to be effective, and is considered suitable to be incorporated in this framework.

Besides, we extend the auto-context model with the aid of deformable registration approach, in order to solve the issue of spatial alignment in individual images, as mentioned in Sect. 1. Due to the variation of hippocampus across individual training/test images, the robustness of the sub-region random forest can be strongly challenged. As only the patches within a certain sub-region are provided for training the classifier, the forest would certainly give poor labeling predictions when it was applied to other mis-aligned sub-regions in test images. Therefore, the labeling performances can be greatly reduced.

In this paper, the tool of the diffeomorphic Demons [12] is applied to reduce the shape variations of the ROIs in the training/test images[1]. The conventional process of deformable registration is to align the intensity images into the template space that could be arbitrarily chosen, and then apply the computed warps to their corresponding label maps. However, the global registration process is ineffective in experiments, as the hippocampus is too small compared with the whole brain MR images.

[1] Settings for the diffeomorphic demons are: 15, 10 and 5 iterations in low, middle and high resolution. The smoothing kernel size is set as 2.0.

Therefore, we decide to implement a localized deformable registration in this work, which only focuses on aligning the ROI of hippocampus instead. This novel strategy is incorporated with the auto-context model, which is regarded as the registration-based auto-context method in this paper. The detailed methodology is given as follows:

(1) In the bottom level, we commence by initializing the context feature by averaging the initially-aligned label maps of all training atlases. It is shown in the experimental section that this strategy is proved valid when labeling the hippocampus region in the brain MR images. Since the shape and location variations of different atlases are stable, the average information of all label maps can be basically considered as the initial context feature for training the classifiers. The deformable registration in this level is also implemented following the traditional way, by registering whole intensity images directly.
(2) In the higher level, we first perform the localized registration to the training and test images. Since the labeling information for the test image is unknown, we use the estimated label maps in the lower level instead, to align with the label maps in the training images. In order to avoid the potential registration error, the label maps are smoothed by a Gaussian filter with $\sigma = 1$ mm. We then register the smoothed label maps to the template space, and apply the obtained warp information to the intensity images.
(3) Since the quality of registration is gradually improved when approaching the topmost of the hierarchy, the training and testing for individual levels of the hierarchy of random forests actually happen in different image spaces. Therefore, the estimated labeling is first warped back to the original image space using the inverted warp information in the lower level, and then registered to the template space in the higher level. Therefore, the tentative labeling results will get updated automatically through the increased levels of the hierarchy, enabling the higher-level forests to generate more robust and accurate labeling results.

2.3 Hierarchical Learning Framework

The methodologies in the training and testing stages for the proposed hierarchical learning framework are presented as follows:

Training Stage

(1) In the bottom level, we initialize the context feature and perform deformable registration by following the strategy of the registration-based auto-context method. Also note that since the localized registration strategy is not implemented in the bottom level as described in Sect. 2.2, the sub-region random forest is therefore not eligible to be applied in this level. Hence we instead apply the conventional random forest approach to the training images, where the bagging procedure is applied by the uniform sampling of the whole region in all training images.

(2) In the higher level, the estimates in the lower level are utilized for the deformable registration. Their corresponding probability label estimates for each hippocampus region are aligned to the template space using the computed warp information, and then considered as the context feature for training the forests in the higher level.
(3) Process of patch extraction and clustering is implemented, following the strategies introduced in Sect. 2.1. The procedure of the forest clustering and re-training is therefore implemented iteratively for completing the hierarchical structure of the forests in the end.

Testing Stage

(1) Given a new test image, we also commence by obtaining the context feature and implementing the registration following the same strategy as the training stage. In the bottom level, the forest selection strategy is not implemented, since the sub-region corresponds to the whole ROI and only one forest is available in this level, and the selection strategy requires prior labeling information.
(2) In the second and higher levels, the labeling estimate in the lower level is considered as the prior for the selection strategy, and also the source of the context features in the hierarchy. Using the introduced fusion strategy, iteratively the labeling results will be further refined, which is ended when reaching the top-most level in the hierarchy.

3 Experimental Results

In this section, we evaluate the proposed framework for labeling hippocampus in MR brain images. The dataset employed is the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset[2] [14], which provides a large number of adult brain MR images acquired from 1.5T MR scanners, along with the annotated left and right hippocampi.

We have randomly selected 101 ADNI images from the Normal Control (NC), Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) subjects, in which one atlas containing the closest label map similarity in overall with the rest is considered as the template for the registration process. The template itself is excluded from subsequent training and testing. We used the standard preprocessing procedures following the works introduced in [6] to ensure the validity of the estimation. Besides, the ITK-based histogram matching program was also applied to all the ADNI images for the experiments, which were then rescaled to the intensity range [0 255]. Next, we implemented the FLIRT program in the FSL library [15] for affine registration to bring all images into the template space.

We implemented 10-fold cross-validation experiments for demonstrating the validity of the proposed methods. State succinctly, the 100 images are equally divided into 10 folds. In each fold, we select one fold for testing, and the rest for training.

[2] http://adni.loni.ucla.edu.

Table 1. Quantitative comparison of performances in different configurations when labeling the left and the right hippocampi.

DSC	Bottom level	Second level
Left Hippo.	81.82 % ± 0.89 %	82.31 % ± 0.90 %
Right Hippo.	82.08 % ± 0.60 %	82.33 % ± 0.60 %
Overall	**81.95 % ± 0.66 %**	**82.32 % ± 0.67 %**

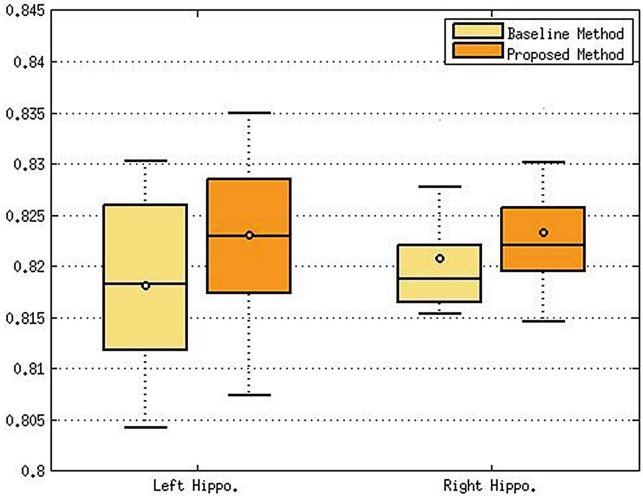

Fig. 2. The box plot for the labeling accuracies of different configurations on the left (left panel) and the right (right panel) hippocampi.

It is also noted that the same settings and parameters were used in all the experiments. Their values are decided by considering both aspects of computation costs and estimation performance. The maximum depth of each tree is 20, and each leaf node has a minimum of 8 samples. Each node chooses Haar-like features from a pool at the size of 1000. The Haar-like features are calculated from the 3D patches with the maximal size of 10 mm × 10 mm × 10 mm.

In this paper we construct a two-level hierarchical structure in the training stage for efficient computation. In the bottom level, one single random forest is trained using the conventional approach, which has 720 trees since the number of training atlases is large. In the second level, 20 patches with size of 15 × 15 × 15 are extracted in each training image. The number of forests are decided by the clustering process using affinity propagation based on the default settings, and each forest has 15 trees.

Our goal in this section is to demonstrate the improvements of the performance when the novel strategies are applied in the second level. Table 1 shows the comparison results between the labeling estimates and the groundtruth using the Dice ratio. Figure 2 presents the box plots that visualize the performances of the two levels. The left and the right panels show the information of the left and the right hippocampi, respectively. It can be observed that in both hippocampi, the DSC scores of the second

level are better than that in the bottom level, which represent the performances of the conventional random forest technique. It is also worth noting that the p-value in the two-tailed paired t-test between any two levels for both left and right hippocampi is lower than 0.05, indicating that the proposed strategies can significantly improve the labeling accuracy when applied to MR brain images. The average runtime of the labeling process is around 10 min using a standard computer (Intel Core i7-4770 K 3.50 GHz, 16 GB RAM), which is affordable in the applications of medical image analysis. It is also noted that in the future work, more levels will be developed in the current hierarchical structure of random forests.

4 Conclusion

In this paper, we present a novel framework using the random forests and several other techniques. While the conventional methods train the forests using whole region information, each forest obtained following the proposed training strategy is focused on *only* the specific sub-regions of the ROI. We also provide the novel registration-based auto-context method to aid our work, and also a forest selection strategy for choosing the most suitable classifiers. In the experimental section, we apply the proposed framework to the ADNI dataset, in which we observe some significant improvements in the labeling performance. In the future work, we will seek possibilities of applying the proposed frameworks to labeling of other ROIs in the brain images.

References

1. Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L.L., Augustinack, J., Dickerson, B.C., Golland, P., Fischl, B.: Automated segmentation of hippocampal subfields from ultra high resolution in vivo MRI. Hippocampus **19**, 549–557 (2009)
2. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron **33**, 341–355 (2002)
3. Jia, H., Yap, P.-T., Shen, D.: Iterative multi-atlas-based multi-image segmentation with tree-based registration. NeuroImage **59**, 422–430 (2012)
4. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging **23**, 903–921 (2004)
5. Wu, G., Wang, Q., Zhang, D., Nie, F., Huang, H., Shen, D.: A generative probability model of joint label fusion for multi-atlas based brain segmentation. Med. Image Anal. **18**, 881–890 (2013)
6. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage **54**, 940–954 (2011)
7. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). Springer
8. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found. Trends® Comput. Graph. Vis. **7**, 81–227 (2012)

9. Zikic, D., Glocker, B., Criminisi, A.: Atlas encoding by randomized forests for efficient label propagation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 66–73. Springer, Heidelberg (2013)
10. Lombaert, H., Zikic, D., Criminisi, A., Ayache, N.: Laplacian forests: semantic image segmentation by guided bagging. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 496–504. Springer, Heidelberg (2014)
11. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science **315**, 972–976 (2007)
12. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. NeuroImage **45**, S61–S72 (2009)
13. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 1744–1757 (2010)
14. Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L.: The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin. N. Am. **15**, 869–877 (2005)
15. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. Neuroimage **62**, 782–790 (2012)