Medical Image Analysis 18 (2014) 1262–1273

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Encoding atlases by randomized classification forests for efficient multi-atlas label propagation

D. Zikic [a,*], B. Glocker [b], A. Criminisi [a]

[a] Microsoft Research, 21 Station Road, Cambridge CB1 2FB, United Kingdom
[b] Biomedical Image Analysis Group, Imperial College London, 180 Queen's Gate, London SW7 2AZ, United Kingdom

ARTICLE INFO

Article history:
Received 24 January 2014
Received in revised form 23 June 2014
Accepted 25 June 2014
Available online 2 July 2014

Keywords:
Randomized forest
Multi-atlas label propagation
Brain
Segmentation

ABSTRACT

We propose a method for multi-atlas label propagation (MALP) based on encoding the individual atlases by randomized classification forests. Most current approaches perform a non-linear registration between all atlases and the target image, followed by a sophisticated fusion scheme. While these approaches can achieve high accuracy, in general they do so at high computational cost. This might negatively affect the scalability to large databases and experimentation. To tackle this issue, we propose to use a small and deep classification forest to encode each atlas individually in reference to an aligned probabilistic atlas, resulting in an *Atlas Forest* (AF). Our classifier-based encoding differs from current MALP approaches, which represent each point in the atlas either directly as a single image/label value pair, or by a set of corresponding patches. At test time, each AF produces one probabilistic label estimate, and their fusion is done by averaging. Our scheme performs only one registration per target image, achieves good results with a simple fusion scheme, and allows for efficient experimentation. In contrast to standard forest schemes, in which each tree would be trained on all atlases, our approach retains the advantages of the standard MALP framework. The target-specific selection of atlases remains possible, and incorporation of new scans is straightforward without retraining. The evaluation on four different databases shows accuracy within the range of the state of the art at a significantly lower running time.

1. Introduction

Labeling of healthy human brain anatomy is a crucial prerequisite for many clinical and research applications. Due to the involved effort (a fully manual labeling of a single brain takes 2–3 days (Klein and Tourville, 2012)), and increasing database sizes (e.g. ADNI, IXI, OASIS), a lot of research has been devoted to develop automatic methods for this task. While brain labeling is a general segmentation task (with a high number of labels), the standard approach for this task is multi-atlas label propagation (MALP) – see (Landman and Warfield, 2012) for an overview of the state of the art. With the *atlas* denoting a single labeled scan, MALP methods first derive a set of label proposals for the target image, each based on a single atlas, and then combine these proposals into a final estimate.

Currently, there are two main strategies for estimating atlas-specific label proposals. The first and larger group of methods non-linearly aligns each of the atlas images to the target image, and then – assuming one-to-one correspondence at each point – uses the atlas labels directly as label proposals, cf. e.g. (Rohlfing et al., 2004; Warfield et al., 2004; Heckemann et al., 2006). The second group of patch-based methods has recently enjoyed increased attention (Coupé et al., 2011; Rousseau et al., 2011; Wu et al., 2012). Here, the label proposal is estimated for each point in the target image by a local similarity-based search in the atlas. Patch-based approaches relax the one-to-one assumption, and aim at reducing the computational times by using linear instead of deformable alignment (Coupé et al., 2011; Rousseau et al., 2011), resulting in labeling running times of 22–130 min per target on the IBSR dataset (Rousseau et al., 2011). The fusion step, which combines the atlas-specific label proposals into a final estimate, aims to correct for inaccurate registration or labelings. While label fusion is a very active research topic, it is not the focus of this work. Additionally, some approaches perform further refinement, e.g. by learning classifiers for fine-scale class-based correction (Wang et al., 2012).

While current state of art techniques can achieve high levels of accuracy, in general they are computationally demanding. This is primarily due to the *non-linear registration between all atlases and the target image*, combined with the long running times for the best performing registration schemes for the problem (Klein et al.,

* Corresponding author.
E-mail address: darko.zikic@tum.de (D. Zikic).

http://dx.doi.org/10.1016/j.media.2014.06.010
1361-8415/

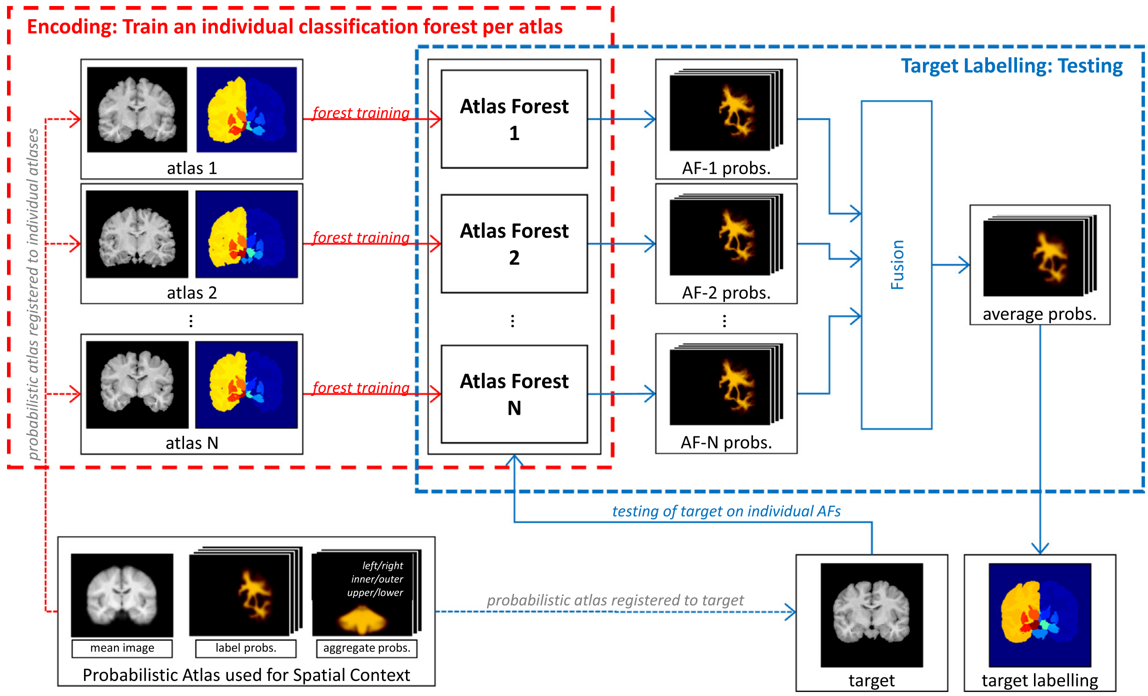

Fig. 1. Framework overview. A single atlas is encoded by training a corresponding atlas forest on the samples from that atlas only. The labeling of a new target is performed by the testing step on the trained atlas forests, and the following fusion of the probabilistic estimates by averaging. For the entire method, the intensity images are augmented by label priors as further channels, obtained by registering a probabilistic atlas.

2009). Current methods state running times of 2–20 h per single registration (Landman and Warfield, 2012). Furthermore, sophisticated fusion schemes can also be computationally expensive. State of the art approaches report fusion running times of 3–5 h (Wang et al., 2012; Asman and Landman, 2012a; Asman and Landman, 2012b).

While the major drawback of high computational costs is the scalability to large and growing databases, they also limit the amount of possible experimentation during the algorithm development phase.

Our method differs from previous MALP approaches in the way how label proposals for a single atlas are generated, and is designed with the goal of low computational cost at test time and experimentation. In this work, we focus on the question of how a single atlas is encoded. From this point of view, methods assuming one-to-one correspondence represent an atlas directly as an image/label-map pair, while patch-based methods encode it by a set of localized patch collections. Variations of the patch-based encoding include use of sparsity (Wu et al., 2012), or use of label-specific kNN search structures (Wang et al., 2013).

In contrast to previous representations, we encode a single atlas together with its relation to label priors by a small and deep classification forest – which we call an *Atlas Forest* (AF). Given a target image as input (and an aligned probabilistic atlas), each AF returns a probabilistic label estimate for the target. Label fusion is then performed by averaging the probability estimates obtained from different AFs. Please see Fig. 1 for an overview of our method. While patch-based methods use a static representation for each image point (i.e. a patch of fixed size), our encoding is spatially varying. In the training step, our approach learns to describe different image points by differently shaped features, depending on the point's contextual appearance.

Compared to current MALP methods, our approach has the following important characteristics:

1. *Only one registration per target is required.* This registration aligns the probabilistic atlas to the target. Since only one registration per target is required, the running time is independent of the database size in this respect. This differs conceptually from patch-based approaches, where the efficiency does not come from reducing the number of registrations, but from using affine instead of non-linear transformations.
2. *Efficient generation of atlas proposals and their fusion.* For proposal generation one AF per atlas is evaluated. Due to the inherent efficiency of tree-based classifiers at test time, this is significantly more efficient than current approaches.
3. *Efficient Experimentation.* A leave-one-out cross-validation of a standard MALP approach on n atlases requires registration between all images, thus scaling with n^2. In contrast, the training of the single AFs, which is the most costly component of our approach for experimentation, scales with n (this assumes that generating the probabilistic atlas is not part of experimentation).

Besides being efficient, experiments on 4 databases in Section 3 indicate that our scheme also achieves accuracy within the range of the state of the art.

Being based on discriminative classifiers, our approach is also related to a number of works which employ machine learning techniques. Compared to the use of multi-atlas label propagation techniques discussed above, the use of machine learning for brain labeling is still relatively limited. In (Tu et al., 2008), a hybrid model is proposed, which combines a discriminative probabilistic-boosting tree (PBT) classifier (Tu, 2005) with a PCA-based generative shape model of the individual anatomical structures. In (Tu and Bai, 2010), the Auto-Context framework with the PBT classifier was applied to brain labeling, and shown to outperform (Tu et al., 2008). Recently, the use of classifiers to correct systematic mistakes of labeling methods in a post-processing step has been shown to improve accuracy (Wang et al., 2011, 2012).