A General-Purpose Learning-Based Wrapper Method to Correct Systematic Errors in Automatic Image Segmentation: Consistently Improved Performance in Hippocampus, Cortex and Brain Segmentation

H. Wang[1], S. R. Das[1], M. Altinay[1], J. Pluta[1], J. Suh[1], C. craige[1], B. Avants[1], and P. Yushkevich[1]
[1]PICSL, Department of Radiology, University of Pennsylvania, Philadelphia, PA, United States

Introduction

Accurate automatic segmentation is highly desirable in a large number of neuroimaging applications. Many software tools that address specific segmentation problems are available to today's researcher. However, end users of these tools are not always able to achieve the high levels of segmentation accuracy reported by the tool developers. Many factors may contribute to this discrepancy. Firstly, the manual segmentation protocols used by the tool developers and tool users may be different. Thus, the automatic method may be performing just fine on the end user's data, but the end user's definition of the ground truth may differ from that of the tool developer. Secondly, modern automatic segmentation methods are largely knowledge-based and incorporate expert knowledge, which is often constructed based on some specific dataset that may be consistently different from the end user's data.

One way to address this problem is that the end user retrains and retunes the automatic segmentation method on his/her own data, using his/her own segmentation protocol. However, this approach is not universally available, and may require scientific and technical expertise far beyond the level needed to apply the segmentation method. We propose a simpler alternative approach that can be easily applied to any existing segmentation tool. Given example manual segmentations and imaging data provided by the end user, our main contribution is offering a *wrapper algorithm* that automatically adapts out-of-the-box automatic segmentation software to perform optimally on the user's data. A reference implementation of our wrapper method is provided as open source software (at http://www.nitrc.org/projects/segadapter).

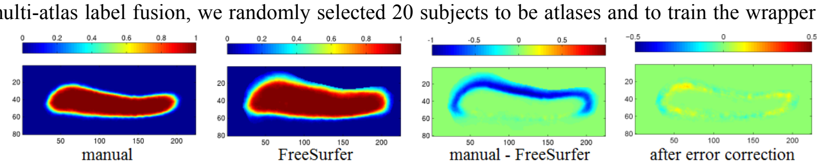
Materials and Methods

We demonstrate the usage of our wrapper algorithm through experiments on three different segmentation problems: segmentation of the hippocampus, brain extraction and brain tissue segmentation. For the experiments of hippocampus segmentation, we use the 3 T MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. In our study, we only use data from mild cognitive impairment (MCI) patients (n=82) and controls (n=57). For brain extraction and brain tissue segmentation experiments, we use the dataset containing 18 T1-weighted MR brain images from the Internet Brain Segmentation Repository (IBSR).

Two wrapper algorithms were implemented. The first method, *explicit error correction* (EEC) contains two sequential components: error detection and error correction. Error detection detects mislabeled voxels produced by the host method and error correction reassigns new labels to the voxels marked as mislabeled by bias detection. Both components were implemented using the AdaBoost algorithm [3] trained with the spatial features and features extracted from the intensity image and segmentations produced by the host method. The second method, *implicit error correction* (IEC) combines the two components and directly learns to perform the segmentation task using the AdaBoost algorithm with the same set of features. Since it aims at directly modifying the segmentation produced by a host segmentation method toward the corresponding manual segmentation, it implicitly corrects the errors produced by the host method. For binary segmentation problems, simply switching the label of the mislabeled voxels fixes the error. Hence, there is no need to train the error correction classifier for EEC as EEC is equivalent to IEC.

We tested the wrapper algorithms with four host automatic segmentation algorithms. FreeSurfer [2] and a multi-atlas label fusion algorithm [1] were applied separately as the host method for hippocampus segmentation. For each cross-validation experiment with FreeSurfer, we randomly selected 70 subjects for training and the remaining 69 for testing. For the cross-validation experiment with multi-atlas label fusion, we randomly selected 20 subjects to be atlases and to train the wrapper algorithm, and another 20 subjects for testing. The BET algorithm [4] and the Fast algorithm [5] were used as the host method to perform brain extraction and three-tissue segmentation, respectively. For these experiments, 9 subjects were randomly selected for training and 9 subjects for testing. For each host method, 10 cross-validation experiments were conducted.

Since providing user-supplied manual segmentation for training is a non-trivial task, it may be difficult to have the same amount of training data in practice. To establish the performance of the wrapper algorithm w.r.t. the size of training set, we conducted experiments using different sizes of training datasets for each host method.



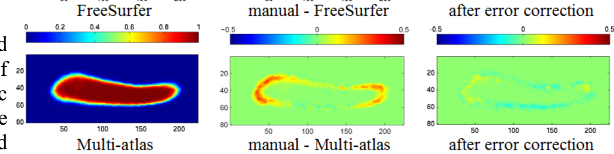
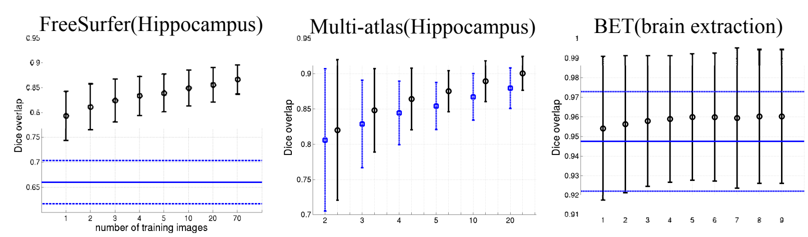


Fig 1. Normalized label distribution of manual and automatic segmentation of the hippocampus. The third and fourth columns show the consistent spatial differences between manual and the two host methods before and after applying the error correction algorithm.

Results

Fig. 1 shows the normalized spatial label distribution of manual and automatic hippocampus segmentation. FreeSurfer tends to over-segment the hippocampus and the multi-atlas label fusion approach tends to produce under-segmentations. For both host methods, the wrapper algorithm successfully corrected most errors. The segmentation results in terms of Dice overlap for all four experiments are summarized in Table 1. Fig. 2 shows the error correction performance w.r.t. the size of the training set. The wrapper algorithm makes consistent improvement for all host methods.



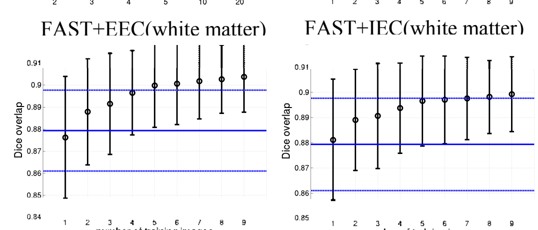


Fig 2. The average Dice overlap before (blue) and after (black) error correction vs. the size of the training set (error bars at ±1 s.d.).

Hippocampus	control / MCI	BET (brain extraction)	0.95
FreeSurfer (FS)	0.67 / 0.65	**BET + Wrapper**	**0.96**
FS + Wrapper	**0.88 / 0.86**	FAST (gray/white matter)	0.94 / 0.86
Multi-atlas (MA)	0.89 / 0.87	**FAST + EEC**	**0.95 / 0.90**
MA+ Wrapper	**0.91 / 0.89**	**FAST + IEC**	**0.95 / 0.91**

Table 1. Results as average Dice overlap before after applying the wrapper method. Hippocampus segmentation are shown for controls and MCI patients. Since the IBSR data have inaccurate CSF labels, three tissue segmentation are shown for gray/white matter only.

Discussion

We presented a method that can improve existing segmentation algorithm performance on the end-user's data using training data provided by the user. Although providing manual segmentations for training may be a difficult task, our experiments show that the wrapper algorithm can make significant improvement with only one or two manual segmentations. Hence, one does not need a large number of manual segmentations to benefit from our technique. Furthermore, our experiment using the multi-atlas approach demonstrates the ability to improve segmentation accuracy without requiring additional training data above that used by the multi-atlas algorithm itself.

Reference

[1] Artaechevarria et al, *IEEE Trans.on MI*, 28:1266–77, 2009. [2] Fischl et al, *Neuron*, 33:341–355, 2002. [3] Freund & Schapire, *ECCLT*, 1995. [4] Zhuang et al, *NeuroImage*, 32:79–92, 2006. [5] Zhang et al, *IEEE Trans. on MI*, 20:45–57, 2001.

Acknowledgements

This work is supported by NIH grants K25 AG027785, R01 AG037376 and grant 10295 from the Penn-Pfizer Alliance.