Learning-Based Meta-Algorithm for MRI Brain Extraction

Feng Shi¹, Li Wang¹, John H. Gilmore², Weili Lin³, and Dinggang Shen¹

¹ IDEA Lab, ³ MRI Lab, Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, USA

² Department of Psychiatry, University of North Carolina at Chapel Hill, USA
dgshen@med.unc.edu

Abstract. Multiple-segmentation-and-fusion method has been widely used for brain extraction, tissue segmentation, and region of interest (ROI) localization. However, such studies are hindered in practice by their computational complexity, mainly coming from the steps of template selection and template-to-subject nonlinear registration. In this study, we address these two issues and propose a novel learning-based meta-algorithm for MRI brain extraction. Specifically, we first use exemplars to represent the entire template library, and assign the most similar exemplar to the test subject. Second, a meta-algorithm combining two existing brain extraction algorithms (BET and BSE) is proposed to conduct multiple extractions directly on test subject. Effective parameter settings for the meta-algorithm are learned from the training data and propagated to subject through exemplars. We further develop a level-set based fusion method to combine multiple candidate extractions together with a closed smooth surface, for obtaining the final result. Experimental results show that, with only a small portion of subjects for training, the proposed method is able to produce more accurate and robust brain extraction results, at Jaccard Index of 0.956 ± 0.010 on total 340 subjects under 6-fold cross validation, compared to those by the BET and BSE even using their best parameter combinations.

1 Introduction

Brain extraction, also called as skull stripping, is a fundamental step in the brain MR image preprocessing, aiming to remove the non-brain tissues, such as skull, scalp, and dura. Accurate brain extraction is crucial since it is non-reversible for the subsequent steps to recover the wrongly-removed brain tissues, and also the residual non-brain tissue (especially dura) in the brain extraction result could overestimate the local brain volume or cortical thickness. Accordingly, a number of automated brain extraction algorithms have been developed, using morphology, morphology combined with edge detection (BSE) [1], deformable model (BET) [2], graph cut, watershed, and other hybrid techniques. Each algorithm has its merits and pitfalls. For example, it is reported that BET usually remains some non-brain tissue, while BSE may remove extra brain tissue [3]. Thus, when applied to a large cohort with varying scanning parameters and diagnosis types, manual adjustment of program parameters and manual editing of extraction results are inevitable.

Multiple-segmentation-and-fusion method has been recently developed with promising results in brain extraction [4], tissue segmentation [5], and region of interest (ROI) localization such as hippocampus [6]. Generally, the method choose a subset of the most similar templates from the template library to propagate their labels to the test subject, and then fuse them together for a final labeling. Studies suggest that combining complementary results may substantially improve the robustness of final brain extraction. However, the applications of this method are largely hindered by the high computational complexity, mainly coming from the first two steps, namely template selection and template-to-subject propagation (usually by nonlinear registration). For example, in MAPS method [4], it takes about 19 hours for brain extraction of a single subject.

In this study, we address these issues by proposing a novel learning-based meta-algorithm for MRI brain extraction. Specifically, first, we do not need exhaustive template-subject comparison to select the appropriate templates; instead, we use exemplars to represent the entire template library, and thus the test subject just needs to be compared with exemplars. Second, without need of the template-to-subject warping, we use a meta-algorithm, i.e., combining two existing individual brain extraction algorithms (BET and BSE), to conduct multiple extractions directly on test data. Effective parameter settings for the meta-algorithm are learned from the training data and passed to subject through exemplars. We further develop a level-set based label fusion to combine multiple candidate extractions together with a closed smooth surface, for obtaining the final result. Thus, errors in voxel-wise label fusion (e.g., by simple majority voting) can be greatly eliminated. Methodological details and experimental results are provided in the sections below.

2 Materials and Method

2.1 Subjects and MRI Acquisition

Our data is obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) (www.adni-info.org), which is a large multi-site study acquired from multiple centers, and so it is ideal to test the accuracy and robustness of the proposed brain extraction algorithm. We use the baseline T1-weighted MRI scans of 340 subjects with age ranging from 55 to 90 years old, including 33 Alzheimer's disease (AD), 236 mild cognitive impairment (MCI), and 71 healthy controls. These MRI scans were acquired from 1.5T scanners made by GE, Philips, and Siemens, using the scanning parameters: TR/TI/TE=2300/1000/3.5 ms, flip angle=8°, FOV=240×240 mm². Two sizes of images were acquired: one is 160 sagittal slices with voxel resolution of 1.25×1.25×1.2 mm³, and another is 180 sagittal slices with voxel resolution of 0.94×0.94×1.2 mm³.

For preprocessing, all images are resampled into isotropic 1×1×1 mm³. Bias field is estimated with N3 algorithm [7] to correct the intensity inhomogeneity in all MR images separately.

2.2 Overview of the Proposed Learning-Based Meta-Algorithm

A learning-based meta-algorithm is proposed for brain extraction, with three major steps as illustrated in Fig. 1. A small number of subjects are selected as training data, and the rest are used as test data. The *first* step is the brain localization, where a brain mask is generated from the training data to roughly localize the brain region of the test subject. The *second* step is exemplar selection and parameter learning. Exemplars are chosen for representing the whole training data. Parameters for each individual algorithm (BET/BSE) are uniformly sampled and exhaustively combined, to learn the parameter map of brain extraction performance for each training subject. As for each exemplar, its parameter map can be obtained by averaging from all training subjects according to their contributions to that exemplar. The *third* step is the brain extraction and fusion for the test subject. In the application, each test subject is compared with all exemplars to find the best matching one, and then the corresponding set of parameter combinations are employed to conduct multiple brain extractions on the test subject. All candidate extractions are then fused together into the final result with a level-set based algorithm. Details are given in the following subsections.

Fig. 1. A conceptual diagram illustrating the three main steps in the proposed learning-based meta-algorithm for brain extraction

2.2.1 Step 1: Brain Localization

This step is to spatially normalize the training data and generate a mask to roughly localize the brain for the test subject, so that the search range can be reduced and individual brain extraction algorithms can perform better. For example, when applied to the original brain image, BET often keep some non-brain tissues in front of brain stem, and BSE may even remain some neck regions [8]. These problems can be partially solved by this step.

Each training subject contains two images, namely a with-skull image and a brain-extracted image, which is manually delineated and serves as the ground-truth. First, all the brain-extracted images of training subjects are affine aligned onto a widely used population template named as ICBM152 [9] for spatial normalization. The estimated transformations are used to bring their with-skull images also onto the common

space. Then, the with-skull images and the brain-extracted images are averaged across all subjects separately. The average brain-extracted image (Fig. 2a) is binarized into a brain mask by keeping all non-zero voxels, which is further dilated for a number of voxels d to generate a final brain mask for better brain localization (Fig. 2b). We conservatively set $d = 12$ in this paper to avoid removing brain tissues when normalization errors exist. The average with-skull image (Fig. 2c) is used to serve as an intensity template for aligning the test image (which only has the with-skull image) onto the same space (Fig. 2d). Finally, the warped test image is masked with the generated brain mask (Fig. 2e). By doing this, the non-brain tissues with a certain distance from the brain in the test image can be completely (e.g., neck regions) or partially (e.g., near the brain stem regions) removed.

Fig. 2. Illustration of the average brain-extracted image from training data (a), the generated brain mask (b), the average with-skull image from training data (c), a test image from an AD subject (ID 067_S_0110) after affine alignment (d), and the test image applied with brain mask (e). Red arrows indicate some non-brain tissues, which pose challenges to individual brain extraction algorithms, are totally or partially removed by the brain mask.

2.2.2 Step 2: Exemplar Selection and Parameter Learning

Exemplar Selection. We propose to use a small number of exemplars to represent the entire data, so that prior knowledge can be carried by these exemplars and then propagate to the test subject. By doing so, the traditional exhaustive template-subject comparisons can be substantially reduced. To do this, we mask the with-skull images of training data by the dilated brain mask. Then, we use the affinity propagation algorithm [10] to cluster the masked training data into a number of classes, and further determine their respective exemplars. Briefly, each pair of training data is computed for their intensity similarity using mutual information, and the results on all possible pairs form an N by N matrix, where N is the number of training data. The diagonal elements of the similarity matrix have influence on the number of selected exemplars, which are set as the median of the input similarities to allow all training data to have equal probability as exemplars. In each iteration, the elements in similarity matrix

exchange messages of how suitable one element to be the exemplar of another one. When the matrix updating converges, a number of training subjects are selected as exemplars.

Parameter Learning. We propagate only the parameter settings learned from the training data to conduct multiple extractions directly on the test subject. BET and BSE are employed in our meta-algorithm due to their well-recognized performance and speed. Each of them has a number of parameters. We uniformly sample the commonly-used parameters for the two algorithms, as specified in Table 1. Then, every combination of parameters is used to conduct brain extraction on each masked training data. The overlap rate is finally computed between automated segmentation A and manual segmentation M with Jaccard Index as $J = |A \cap M|/|A \cup M|$. Thus a parameter map is formed for each individual algorithm on each training data, e.g., a 15×6 map built for BET (Fig. 3a) and a $5 \times 4 \times 5$ map built for BSE, where each element represents the overlap rate with respect to the parameters used. For each exemplar, its parameter map (Fig. 3b) is defined as the average over all training data according to their contributions to that exemplar.

Table 1. Parameters and sampling strategies used for BET and BSE in the proposed meta-algorithm

Methods	Parameters	Default	Sampling
BET	-f (fractional intensity threshold)	0.5	0.1:0.05:0.8
	-g (vertical gradient)	0	-0.3:0.1:0.2
BSE	-d (diffusion constant)	25	15:5:35
	-n (diffusion iterations)	3	1:2:7
	-s (edge detection constant)	0.62	0.5:0.05:0.7

Fig. 3. Illustration of BET parameter maps on a training data (a) and an exemplar (b), as well as the thresholded exemplar parameter map showing top M highest ranking parameter combinations (c-d). X-axis and y-axis represent the two parameters (i.e., -f and -g) in BET.

2.2.3 Step 3: Brain Extraction and Fusion

Brain Extraction. For the test subject, the best match exemplar is found by comparing the intensity similarity of its masked with-skull image with that of all exemplars. Thus, the parameter map of the selected exemplar can be used to guide multiple brain extractions on the test subject. Top M effective parameter combinations are chosen from the parameter map (Fig. 3c), so that M instances of brain extractions are conducted on the test subject (Fig. 3d). Note that more instances of individual brain extractions will bring higher computational cost, although the accuracy might be improved. We empirically use $M = 20$ in this paper.

Fusion. When multiple candidate extractions are available, the next question is how to fuse them together. Majority voting is widely used; however, it can produce isolated false extractions and sometimes unsmooth boundary in the extracted brain region. To address this issue, we model the brain as a closed entity with a smooth surface, and thus use a level-set method [11] to find the brain boundaries from the average label map of the multiple brain extraction results, as shown in Fig. 4. For this method, two parameters are important in constraining the shape of the level-set based surface. The first parameter is to smooth the resulting level set with kernel s , and the second parameter is to dilate the final brain extraction result outward for t voxels to reduce the risk of cutting brain tissues, as recommend in many studies [4]. It is worth noting that these two parameters could also be learned from the training data, by uniformly sampling the parameter values and picking the one with the best overlap with ground truth. We use $s = 11$ and $t = 2$ in this paper. Finally, by applying the estimated level-set mask (Fig. 4d) onto the test subject, the brain extraction result can be obtained (Fig. 4e).

Fig. 4. Illustration of the level-set based fusion on the AD subject shown in Fig. 2. Averaged candidate extractions is shown in (a), averaged mask in (b), level-set based surfaces (red) using $s = 0$, $t = 0$ (c), or $s = 11$, $t = 2$ (d), and final extraction result (e).

3 Experimental Results

In our experiment, a 6-fold cross-validation is performed. All subjects were randomly partitioned into 6 folds. Each fold was used in the training process and the rest was left as testing set. Estimation of errors was the overall proportion of errors committed in all folds. Taking one case as example, 54 of 340 subjects (8 AD, 22 MCI, and 24 controls) are selected as training data, and the rest are used as test data. With the method described above, the exemplars we obtained come from 2 AD, 3 MCI, and 1 control, indicating in part the ability of our exemplars for representing the data distribution.

Automated extractions are compared with manual ones by calculating their overlap rates. The performance is further compared with that obtained by each of the four baseline methods, i.e., BET without/with our brain localization step (called Raw BET and BET, respectively), and also BSE without/with our brain localization step (called Raw BSE and BSE, respectively). For these four baseline methods, every combination of parameters is explored and the one with the highest average Dice ratio across the entire test data is chosen as the final result. Furthermore, majority voting is also used to replace the level-set based data fusion in the proposed pipeline, and serve as another control method. All results are shown in Fig. 5. It can be observed that, the proposed method significantly outperforms other five methods ($p < 0.001$) by achieving the average Jaccard Index of 0.956 ± 0.010 , in comparison with 0.903 ± 0.026 and 0.935 ± 0.010 by Raw BET and BET, 0.870 ± 0.029 and 0.883 ± 0.037 by Raw BSE and BSE, and 0.849 ± 0.036 by MajorityVote. Note that, by using our method, not only the accuracy is largely improved, the robustness is also increased as reflected by the lowest standard deviation. Meanwhile, our brain localization step shows its ability to improve the accuracy and robustness of brain extractions for both BET and BSE, compared with Raw BET and Raw BSE, respectively. Another observation is that the extraction result on abnormal brain such as AD has lower accuracy than that of control, which agrees with previous findings [3].

Fig. 5. Jaccard Index for automated brain extraction by the four baseline methods, majority voting, and the proposed method, under 6-fold cross validation

To visualize the location of extraction errors in 3 algorithms (BET with brain localization, BSE with brain localization, and the proposed method), we generate the projection maps for the false positive and negative voxels. Specifically, the 3D false positive (or negative) maps of all test data are first averaged, and then projected onto the sagittal, coronal, and axial directions, as shown in Fig. 6. BET has more false positive voxels, indicating the inclusion of more non-brain tissues in the brain extraction results. BSE has more false negative voxels, indicating wrongly removing more brain tissues. Our proposed method produces the lowest false positive and false negative, in comparison to both baseline algorithms.

Fig. 6. Projection maps of false positive (left panel) and false negative (right panel) for the brain extractions obtained by BET, BSE, and our proposed method on all test data. Brighter value indicates more false positive (or negative) results in that location.

4 Discussion and Conclusion

We have presented a novel learning-based meta-algorithm for MRI brain extraction. It has three novelties. First, we introduce exemplars to represent the whole template library, thus largely reducing the size of templates to match during the application stage. Second, we alternatively propagate the learned parameter settings from template to subject, without requiring nonlinear registration between them. Third, we develop a level-set based label fusion method to overcome both boundary discontinuity and isolated errors in voxel-wise label fusion methods. Experimental results have shown that our proposed method can substantially improve the accuracy and robustness of brain extraction. Also, the proposed meta-algorithm is not limited to BET and BSE; it can be extended for inclusion of other brain extraction algorithms.

The proposed method has demonstrated comparable brain extraction accuracy (Jaccard 0.956, tested on 340 subjects) with MAPS (Jaccard 0.954, tested on 839 subjects) [4]. More importantly, it has much shorter processing time (~ 10 mins) to perform brain extraction on a new subject than that of MAPS (~ 19 hours) [4]. For training stage, 6 hours are needed to process 54 training data on a standard PC for spatial normalization, parameter map learning, and exemplar selection.

Acknowledgments. This work was supported in part by NIH grants EB006733, EB008760, EB008374, EB009634, MH088520, MH070890, MH064065, NS055754, and HD053000.

References

1. Shattuck, D.W., Leahy, R.M.: Automated graph-based analysis and correction of cortical volume topology. *IEEE Trans. Med. Imaging* 20(11), 1167–1177 (2001)
2. Smith, S.M.: Fast robust automated brain extraction. *Hum. Brain Mapp.* 17(3), 143–155 (2002)
3. Fennema-Notestine, C., Ozyurt, I.B., Clark, C.P., Morris, S., Bischoff-Grethe, A., Bondi, M.W., Jernigan, T.L., Fischl, B., Segonne, F., Shattuck, D.W.: Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: effects of diagnosis, bias correction, and slice location. *Hum. Brain Mapp.* 27(2), 99–113 (2006)
4. Leung, K.K., Barnes, J., Modat, M., Ridgway, G.R., Bartlett, J.W., Fox, N.C., Ourselin, S.: Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. *NeuroImage* 55(3), 1091–1108 (2010)
5. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain MRI. *Neuroimage* 47(2), 564–572 (2009)
6. Leung, K.K., Barnes, J., Ridgway, G.R., Bartlett, J.W., Clarkson, M.J., Macdonald, K., Schuff, N., Fox, N.C., Ourselin, S.: Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer’s disease. *NeuroImage* 51(4), 1345–1359 (2010)
7. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17(1), 87–97 (1998)
8. Rex, D.E., Shattuck, D.W., Woods, R.P., Narr, K.L., Luders, E., Rehm, K., Stolzner, S.E., Rottenberg, D.A., Toga, A.W.: A meta-algorithm for brain extraction in MRI. *NeuroImage* 23(2), 625–637 (2004)
9. Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J.: A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* 2(2), 89–101 (1995)
10. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* 315(5814), 972–976 (2007)
11. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Image Process.* 10(2), 266–277 (2001)