



Published in final edited form as:

Conf Comput Vis Pattern Recognit Workshops. 2014 June ; 2014: 3111–3117. doi:10.1109/CVPR.2014.398.

Learning-Based Atlas Selection for Multiple-Atlas Segmentation

Gerard Sanroma, Guorong Wu, Yaozong Gao, and Dinggang Shen

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA

Gerard Sanroma: gerard_sanroma@med.unc.edu; Guorong Wu: guorong_wu@med.unc.edu; Yaozong Gao: yzgao@cs.unc.edu; Dinggang Shen: dinggang_shen@med.unc.edu

Abstract

Recently, multi-atlas segmentation (MAS) has achieved a great success in the medical imaging area. The key assumption of MAS is that multiple atlases encompass richer anatomical variability than a single atlas. Therefore, we can label the target image more accurately by mapping the label information from the appropriate atlas images that have the most similar structures. The problem of atlas selection, however, still remains unexplored. Current state-of-the-art MAS methods rely on image similarity to select a set of atlases. Unfortunately, this heuristic criterion is not necessarily related to segmentation performance and, thus may undermine segmentation results. To solve this simple but critical problem, we propose a learning-based atlas selection method to pick up the best atlases that would eventually lead to more accurate image segmentation. Our idea is to learn the relationship between the pairwise appearance of observed instances (a pair of atlas and target images) and their final labeling performance (in terms of Dice ratio). In this way, we can select the best atlases according to their expected labeling accuracy. It is worth noting that our atlas selection method is general enough to be integrated with existing MAS methods. As is shown in the experiments, we achieve significant improvement after we integrate our method with 3 widely used MAS methods on ADNI and LONI LPBA40 datasets.

1. Introduction

Multiple-atlas segmentation (MAS) has recently gained popularity for labeling the anatomical structures of a target image [1, 2, 10, 11, 13]. It segments an unknown target image by transferring the labels from a population of annotated exemplars (*i.e.*, atlases) to the target image domain, where the transformation is usually derived from image registration. In atlas-based segmentation, we assume that if two anatomical structures have similar shape and location, they should bear the same label (or tissue type). Since a population of atlases encompasses large anatomical variability, MAS has a greater chance of finding appropriate atlases to label an underlying target image than single atlas-based segmentation. To this end, selection of the best atlases that contribute to achieve high segmentation performance is critical before applying any state-of-the-art MAS method.

Most common approaches for atlas selection are based on image similarity measurements such as sum of squared differences (SSD) or normalized mutual information (NMI) [1]. More advanced methods use the distance in the manifold, instead of in the Euclidean space [13, 5]. The drawback of these methods is that they are driven by image similarity which is

not necessarily correlated with final segmentation accuracy. More critically, many of them require the images to be non-rigidly registered in order to compute their distance, which is computationally demanding.

In order to overcome these limitations, we propose a supervised learning approach to model the relationships between a pair of images (*i.e.* atlas and target images) and the relevance of their contribution to segmentation. Figure 1 shows an overview of the approach. As input, our method receives a set of atlases linearly aligned to a common space, *e.g.*, a template image. In the training stage, we first compute the relevance score between each pair of atlases, using one of the atlas as the target image and the other as the atlas. We define the relevance score as the Dice ratio (DR) obtained by using the atlas image to segment the target image (TR1 in Figure 1). Next, we identify a number of key regions in the entire image domain, in order to obtain efficient representations (TR2.a). Then, we extract HOG features [4] to describe the anatomical characteristics in these key regions and compute their squared differences in order to get the pairwise features between each pair of images (TR2.b). After that, we employ SVM-Rank [8] to learn the latent relationship between the pairwise HOG features and the relevance segmentation score (TR3). In the testing stage, we first *linearly* align the to-be-segmented target image to the atlas domain (TS1). Next, we extract selected HOG features from the key regions and compute the pairwise feature vectors between the target image and each atlas (TS2). Finally, we evaluate their eventual segmentation performance using the learned SVM-Rank model (TS3) and select the best atlases to be used for MAS according to the predicted scores (TS4).

As opposed to heuristics such as image similarity, selection by our method is directly related to the segmentation performance. Our learning-based atlas selection method can boost the performance of current state-of-the-art MAS methods. As we show in the experiments, significant improvement is achieved on ANDI and LONI LPBA40 datasets, after we equip the majority voting [10], local weighted voting [2], and non-local patch-based MAS [11] methods with our atlas selection approach.

The overview of the paper is as follows: in Section 2 we describe the proposed method, in Section 3 we detail experimental evaluation and present the results, and in Section 4 we give some concluding remarks.

2. Method

Consider a set of atlases composed of intensity images $\mathcal{A} = \{A_i, i \in I\}$ with $I = \{1, \dots, N\}$ as index set, and their corresponding label images $\mathcal{L} = \{L_i, i \in I\}$ obtained by expert annotation. We assume that all atlases have been linearly aligned to a common space, *e.g.*, a template image. For each voxel in the atlas domain $x \in \Omega_i$, $A_i(x)$ indicates the intensity value at location x , and label $L_i(x) \in \{0, 1\}$ indicates whether the structure of interest is present at location x or not. As we will explain later, our method is straightforward to apply to the case of multiple labels.

Given a target image T , multiple-atlas segmentation aims to locate the anatomical structure in the target image by transferring the label information from the atlases to the target image domain. It consists of two steps: First, spatial correspondence between atlases and target

image is obtained by non-rigid registration. In this way, we obtain a set of registered atlases $\tilde{\mathcal{A}} = \{\tilde{A}_i, i \in I\}$ and their label images $\tilde{\mathcal{L}} = \{\tilde{L}_i, i \in I\}$. Next, a label fusion procedure is often used to determine the label on each target image point based on the label information from the registered atlases $\tilde{\mathcal{L}}$. Many label fusion strategies have been developed. To name a few, majority voting (MV) [10] assigns to each target voxel the label that most frequently appears among the corresponding atlas voxels. Local weighted voting (LWV) [2] weights the contribution of each atlas according to the image patch similarity. Non-local weighted voting (NLWV) [11] introduces the idea of non-local average by performing neighborhood search for similar patches, which gives some extra robustness to possible mis-alignments.

Atlas-selection is an important step affecting the accuracy of MAS. On one hand, using small subsets of atlases may produce less accurate segmentations since some relevant information from other atlases may be left out. On the other hand, using large subsets of atlases may undermine the segmentation performance because of the high amount of irrelevant information introduced. Given a target image T , the most common strategies are to select the K most similar atlases to T , according to image similarity measurements such as normalized mutual information (NMI) [1]. Even though similarity-based selection performs significantly better than random selection, it is still loosely related to the final score which is the relevance of an atlas A_i for segmenting the target image T . Given the target image T and a set of atlases \mathcal{A} and \mathcal{L} , the process of MAS can be represented as

$$\hat{L}_T = MAS(T; \mathcal{A}_S, \mathcal{L}_S) \quad (1)$$

where \hat{L}_T is the estimated segmentation for target image T , and $\mathcal{A}_S, \mathcal{L}_S$, with $S \subseteq I$, is the subset of selected atlases.

Dice ratio (DR) is widely used to measure the degree of overlap between the target structure and the deformed atlas structure. It is defined as

$$DR(L_T, \tilde{L}_i) = \frac{2 \times vol(L_T \cap \tilde{L}_i)}{vol(L_T) + vol(\tilde{L}_i)} \quad (2)$$

where $vol(\cdot)$ denotes volume. Supposing that we know the labels for the target image, this scoring function induces a selection of the best K atlases for target image T , denoted as S_T^* , so that,

$$DR(L_T, \tilde{L}_i) > DR(L_T, \tilde{L}_j), \forall i \in S_T^*, j \notin S_T^* \quad (3)$$

where the cardinality of the set is equal to K , denoted as $|S_T^*| = K$. This scoring function presents two problems to be used for atlas selection in our method: (1) the target labels L_T are unknown, and (2) the deformed atlas labels to the target space L_i are also unknown since one of our goals is to avoid computationally expensive non-rigid registration before the atlas selection step. Figure 2 demonstrates the superiority of using DR, rather than NMI, to select

the best atlases for MAS, where black and blue curves show segmentation performances (assessed by the Dice ratio) by increasing the number of the best atlases selected by equation (2) and NMI, respectively. This shows the potentially large room for improvement of the atlas selection strategy targeted at equation (2) with respect to the widely used criterion based on simple image similarity.

Our aim is to learn a scoring function f that correlates pairwise appearance information of a target image and each unregistered atlas image with segmentation performance in terms of DR. To make our learning approach tractable, we use a linear model for mapping the pairwise representation obtained by the feature extraction process to our final score f , in the following way

$$f(T, A_i) = \vec{w}^\top \Phi(T, A_i) \quad (4)$$

where $\Phi(\cdot, \cdot)$ is the feature vector derived from a pair of images and \vec{w} is the weighting vector where each element measures the contribution of a particular pairwise feature in predicting the segmentation score.

In the next two sections, we describe the process of learning the weights \vec{w} and compute the pairwise features $\Phi(\cdot, \cdot)$, respectively.

2.1. Learning the Relationships between Pairwise Appearance and Segmentation Score

In this section we focus on the computation of the weights \vec{w} while assuming that we have the pairwise features between images (which we will explain in section 2.2).

Consider an atlas in the training set as the target A_t , $t \in I$, and the rest as the atlases A_i , $i \neq t$. According to equation (3), we focus on the separation between the set of the best K atlases S_t^* and the rest $\{I \setminus \{S_t^*, t\}\}$. That is, we want to find the weights \vec{w} that fulfill the following relationships

$$\vec{w}^\top \Phi(A_t, A_i) > \vec{w}^\top \Phi(A_t, A_j), \quad \forall i \in S_t^*, j \notin \{S_t^*, t\} \quad (5)$$

It is worth noting that $\Phi(A_t, A_j)$ denotes the features extracted from a pair of linearly aligned intensity images without applying any non-rigid registration between them. This type of problem, in which we seek to fulfill certain order relationships between pairs of elements A_i, A_j with respect to a given reference A_t , is known as *learning to rank* and there exist several algorithms in the literature aimed at solving it. We use SVM-Rank¹ [8] because it has superior performance than other methods [3]. To this end, we compute a set of constraints for each target A_t , constraining the pairs of atlases A_i, A_j so that the i -th atlas should be ranked higher than the j -th atlas according to ground-truth DR. Consider S_t^* as the ground-truth selection of the best K atlases for target A_t , as defined in equation (3). The set of constraints for target A_t according to ground-truth selection is now defined as follows,

¹we use <http://olivier.chapelle.cc/primal/>

$$r_t^* = \{(i, j) | \forall i \in S_t^*, \forall j \notin \{S_t^*, t\}\} \quad (6)$$

where $(i, j) \in r_t^*$ means that the i -th atlas should be ranked higher than the j -th atlas for the segmentation of target A_t . Using SVM-Rank, we pose this problem as a constrained optimization problem in which we aim to find the weights w that maximize the margin between the scores of the relevant and non-relevant atlases. More specifically,

$$\begin{aligned} & \text{minimize: } \frac{1}{2} \|\vec{w}\|^2 + \eta \sum \xi_{t,i,j} \\ & \text{subject to:} \\ & \forall t \forall (i, j) \in r_t^*: \vec{w}^\top \Phi(A_t, A_i) \geq \vec{w}^\top \Phi(A_t, A_j) + 1 - \xi_{t,i,j} \\ & \forall t \forall i \forall j: \xi_{t,i,j} \geq 0 \end{aligned} \quad (7)$$

where the objective function represents a trade-off between a regularization term and margin size, controlled by the parameter η , and the margin is dynamically set to $1 - \xi_{t,i,j}$ where $\xi_{t,i,j}$ is the slack variable controlling the amount of margin violation regarding each triplet A_t, A_i, A_j .

As part of the training process, we need to perform $(N - 1)^2$ pairwise non-rigid registrations between the atlases in order to compute the constraints of equation (6).

2.2. Pairwise Feature Computation

As mentioned in equation 4, our relevance score f is a function of the pairwise representation features between a target T and an atlas A_i . In order to find the compact and accurate representation to describe the connection between T and A_i , the calculation of $\Phi(T, A_i)$ consists of three steps, namely, (1) key region detection, (2) pairwise HOG computation, and (3) feature selection, as detailed below. It is worth noting that we compute the pairwise features only after linear registration, both in the training and testing stages. However, the Dice ratio in the training stage (sec 2.1) is computed based on the non-rigid registration results, which essentially reflect the goal of our approach which is to predict the segmentation score based only on the rigid registration results.

Key Region Detection—In MAS, the segmentation label at each target point is decided based on the labels of the aligned atlases at that point. Regions with high label variability, such as label boundaries, are the source of most variability in segmentation results. Therefore, we use the appearance in these regions as cues to predict segmentation performance. Since we already know the label information in the training set, we can obtain the set of boundary locations $B(L_i) \subset \Omega_i$, from label image L_i . We define the set of sensible locations as the union of boundary locations in all the training set, $\mathcal{B} = \cup_i B(L_i)$. Figure 3 shows an example structure, its boundary and the union of the boundaries from all pre-registered atlases.

Pairwise HOG Computation—We use HOG features to obtain the appearance descriptors of the key regions, although other features can also be applied here. HOG computes a bi-dimensional histogram of gradient occurrences along spatial and orientation

bins. Spatial bins constitute a division of the image into a grid of 3D cells (since we use 3D images) whereas orientation bins constitute a division of the orientation range (1 ... 360) into intervals. Discretization into bins provides robustness to both displacement and orientation errors after linear alignment. We construct a feature vector $\vec{\theta}$ summarizing the HOG values at bins within a distance ρ of the nearest boundary point in \mathcal{B} . We use the locations of the bin centers in order to compute the distance to the boundary region. Overlaid on the intensity image, Figure 3.(d) shows centers of the selected bins.

After this process, for each individual image, we obtain a vector $\vec{\theta}$ containing a number M of selected HOG features in the boundary region. We use the squared difference between the features from the individual images in order to compute the pool of pairwise features between each atlas and a target image, as follows: $\Theta_{t,i}(j) = (\theta_t(j) - \theta_i(j))^2$, where j denotes the j -th feature and t, i denote the indices of the target and atlas images, A_t, A_i , respectively. In the following, we describe how to select the final set of compact features $\Phi(A_t, A_i)$ that will be used both for training and testing.

Feature Selection—In order to select a compact set of features, we sort the pool of features Θ according to the maximum-relevance minimum-redundancy criterion (MRmR) [9]. This criterion encourages the selection of features that are highly correlated with the target score while maintaining a low redundancy. After sorting, we obtain a sequence $F_1 \subset F_2 \subset \dots \subset F_m \dots \subset F_{M-1} \subset F_M$, such that F_m contains the indexes of the best m features according to MRmR. Suppose we already know F_{m-1} . The m -th feature to be added to F_{m-1} in order to create F_m is decided according to the following formula:

$$\max_{j \notin F_{m-1}} \left(Rel(\Theta(j), \Psi) - \frac{1}{m-1} \sum_{k \in F_{m-1}} Red(\Theta(j), \Theta(k)) \right) \quad (8)$$

where $\Theta(j) = [\Theta_{t,i}(j), \forall t, i]$ is a vector composed of the j -th pairwise feature from all target-atlas pairs, and $\Psi = [DR(L_t, L_i), \forall t, i]$ is a vector of ground-truth DR between all target-atlas pairs. As relevance function $Rel(\cdot)$, we use the absolute value of the Spearman's rank correlation coefficient which measures the correlation between the ranks induced by the features and the ground-truth score. As redundancy function $Red(\cdot)$, we use the absolute value of the Pearson correlation coefficient which measures the correlation between pairs of features. This formula encourages features correlated with the target score and penalizes features correlated with already chosen features.

We finally select a compact set of m' features with the highest MRmR. We found that $m' = 1000$ was enough in our experiments.

3. Experiments and Results

We have conducted experiments in the ADNI² and LONI-LPBA40³ datasets. ADNI dataset contains the segmentations of the left (L) and right (R) hippocampi, which have been

²<http://www.adni-info.org/>

³<http://www.loni.ucla.edu/Atlases/LPBA40>

obtained by a commercial brain mapping tool [6]. LONI-LPBA40 dataset is provided by the Laboratory of Neuro Imaging (LONI) at UCLA and contains 40 brain images, each with 56 manually-labeled ROIs. We regard these segmentations as the ground truth.

We use FLIRT [7] in order to linearly align all atlases to a template image prior to feature extraction. We use non-rigid registration by diffeomorphic demons [12] both for MAS and ground-truth DR computation. We apply the following label fusion methods: MV, LWV and NLWV.

We compare selection by our method (HOG+SVMRank) to selection by NMI. As baseline method, we use the negative sum of squared differences of the raw HOG features, Θ (HOG). This is represented by the scoring function $f(T, A_i) = -\sum_j (\theta_T(j) - \theta_{A_i}(j))^2$, where θ_T and θ_{A_i} are the pool of M HOG features within the boundary regions obtained from target image T and atlas image A_i , respectively. This corresponds to the similarity measurement originally used between HOG descriptors. All the methods have been applied to each structure independently. For NMI, we have used a rectangular image region containing the structure of interest. We conduct 5-fold cross-validation experiments in all the methods.

3.1. Parameter Setting

HOG features have two parameters, namely, the number of orientation bins O and cell size C (in voxels) of the spatial bins. We have found our method not to be very sensitive to the values of these parameters. Due to different sizes of the structures, setting these parameters to a fixed value may cause larger structures to generate an unnecessarily high number of features. In order to trim the number of features to a manageable size, we adaptively fix the HOG parameters C , O in order to get a reasonable number of features $M = 10^4$. Specifically, we start with initial values $C = 5$, $O = 9$ and iteratively augment the cell size C and decrease number of orientation bins O until the number of features is less than M . We set $\rho = 0$, thus selecting only features within the boundary region.

3.2. ADNI Results

For the ADNI dataset, we use 66 randomly selected individuals containing the segmentations of the left and right hippocampus in order to test the segmentation accuracy of different atlas selection methods. The atlas selection methods compared include: NMI-based method, the baseline method (HOG), and our proposed method (HOG+SVMRank). Segmentation accuracy is assessed by the Dice ratio. Figures 4, 5 and 6 show the average segmentation results in both hippocampi obtained by different selection methods using the label fusion methods MV, LWV and NLWV, respectively. Vertical axis shows the segmentation accuracy averaged over all queries in the 5 folds, and horizontal axis shows the number of atlases used.

Regarding the different atlas selection methods, best results by our method are $\sim 2\%$ better than best results by NMI for MV label fusion and $\sim 1\%$ for both LWV and NLWV label fusions. Best results for NMI selection in all the label fusion modalities are achieved when using $K \simeq 25$ best atlases. Comparable results by our method are obtained using as few as K

≈ 7 best atlases, thus representing a considerable save in computation time on deploying both non-rigid image registration and label fusion.

Results of the baseline method (green curves in Figures 4–6) clearly demonstrate that: (1) HOG features are more useful than image intensity in selecting the best atlases; and (2) our learning procedure for atlas selection can substantially improve the segmentation results by providing atlases which are more correlated with segmentation performance.

It is worth mentioning the decrease in performance experienced by MV after selecting a few best atlases. This is because, contrarily to other methods, in MV all atlases have the same importance for predicting the final label. This causes a negative impact as less significant atlases are progressively added after a few best atlases. On the other hand, weighted voting methods use patch similarity measurements in order to filter out poor contributions. Their behavior when using large number of atlases depends on their internal parameters and the number of atlases used, but their performance is more stable than MV.

In order to demonstrate that our scoring function targeted at the ground-truth Dice ratio generalizes well to other performance measurements, we compute the average surface distance errors (in mm) by the different atlas selection methods when using $K = 23$ atlases to segment the left and right hippocampi. The average surface distance errors obtained by HOG+SVMRank vs NMI are 0.4084 vs. 0.4693 (MV) ; 0.3728 vs. 0.4036 (LWV) ; and 0.3608 vs. 0.3819 (NLWV), respectively.

3.3. LONI Results

The LONI dataset contains 40 subjects with 56 manually labeled structures. We use the subset of L and R structures located in the cortical area in order to test different atlas selection methods. The structures used are: frontal gyrus, pre-central gyrus, orbitofrontal gyrus, post-central gyrus, parietal gyrus, occipital gyrus, temporal gyrus and parahippocampal gyrus.

Figures 7, 8 and 9 show the average DR obtained when selecting 25% of the best atlases, *i.e.*, $K = 7$, for MV, LWV and NLWV label fusions, respectively. Each figure shows the results for the left (L) and right (R) parts of the structures. The average improvement of the proposed method with respect to NMI for $K = 7$ are $\sim 1.1\%$, $\sim 1.5\%$ and $\sim 1.2\%$ for MV, LWV and NLWV label fusion, respectively. A circle and a triangle on top of each method's bar represents significant improvement, according to a paired t-test, with respect to NMI or the baseline method. Our method obtained significant improvements with respect to NMI in 9, 11 and 10 out of the 16 structures for MV, LWV and NLWV label fusions, respectively. This compares favorably to the baseline method which only obtains significant improvement with respect to NMI in 2, 4 and 3 out of the 16 structures for MV, LWV and NLWV label fusions, respectively. The most significant improvement of our method with respect to the others has been obtained in the orbitofrontal, occipital and temporal gyri. In the precentral and postcentral gyri, the baseline method obtains similar results to the proposed method, suggesting that the learning part is not having any impact. In the parietal gyrus, however, neither the baseline or the proposed method gets any improvement with respect to NMI.

Confirming what we observed in ADNI, the baseline method obtains intermediate results between NMI and the proposed method. Regarding the computational saving, the average proportion of atlases needed by our method in order to meet the best results by NMI are 40%, 64% and 42% for MV, LWV and NLWV label fusion, respectively.

4. Conclusion

We have presented a method aimed at selecting relevant atlases for multiple atlas segmentation. Our relevance score is related to the expected Dice ratio *after* non-rigid registration of an atlas and a target image. This measure is more related to segmentation accuracy than simple image similarity. We use a learning-based method in order to map the image features before non-rigid registration to our relevance score. In this way, we avoid the need for using the costly non-rigid registration prior to the selection step. Our method presents improvements with respect to NMI-based selection of $\sim 2\%$ for MV label fusion and $\sim 1\%$ for LWV and NLWV label fusions. Our method obtained significant improvement with respect to the NMI-based method in 11 and 10 out of the 16 structures from LONI, for LWV and NLWV label fusions, respectively. Selection by our method requires a third part of the atlases required by the NMI-based method in order to meet its best results. This considerably reduces the computational time required.

References

1. Aljabar P, Heckemann R, Hammers A, Hajnal J, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage*. 2009; 46(3):726–738. [PubMed: 19245840]
2. Artaechevarria X, Munoz-Barrutia A, de Solorzano CO. Combination strategies in multi-atlas image segmentation: Application to brain mr data. *Medical Imaging, IEEE Transactions on*. Aug; 2009 28(8):1266–1277.
3. Chapelle O, Keerthi SS. Efficient algorithms for ranking with svms. *Inf Retr*. Jun; 2010 13(3):201–215.
4. Dalal N, Triggs B. Histograms of oriented gradients for human detection. *International Conference on Computer Vision & Pattern Recognition*. Jun.2005 2:886–893.
5. Hoang Duc AK, Modat M, Leung KK, Kadir T, Ourselin S. Manifold learning for atlas selection in multi atlas-based segmentation of hippocampus. 2012
6. Hsu YY, Schuff N, Du AT, Mark K, Zhu X, Hardin D, Weiner MW. Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia. *Journal of magnetic resonance imaging : JMRI*. Sep; 2002 16(3):305–310. [PubMed: 12205587]
7. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*. 2002; 17(2):825– 841. [PubMed: 12377157]
8. Joachims T. Optimizing search engines using clickthrough data. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '02*. 2002:133–142.
9. Peng H, Long F, Ding C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell*. Aug; 2005 27(8):1226–1238. [PubMed: 16119262]
10. Rohlfing, T.; Brandt, R.; Menzel, R.; Russakoff, DB.; Maurer, CR, Jr. *The Handbook of Medical Image Analysis – Volume III: Registration Models*. Vol. chapter 11. Kluwer Academic/Plenum Publishers; Aug. 2005 Quo vadis, atlas-based segmentation?; p. 435-486.
11. Rousseau F, Habas PA, Studholme C. A supervised patch-based approach for human brain labeling. *IEEE Trans Med Imaging*. 2011; 30(10):1852–1862. [PubMed: 21606021]

12. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*. Nov.2008 45(1)
13. Wolz R, Aljabar P, Hajnal JV, Hammers A, Rueckert D. Leap: Learning embeddings for atlas propagation. *NeuroImage*. 2010; 49(2):1316–1325. [PubMed: 19815080]

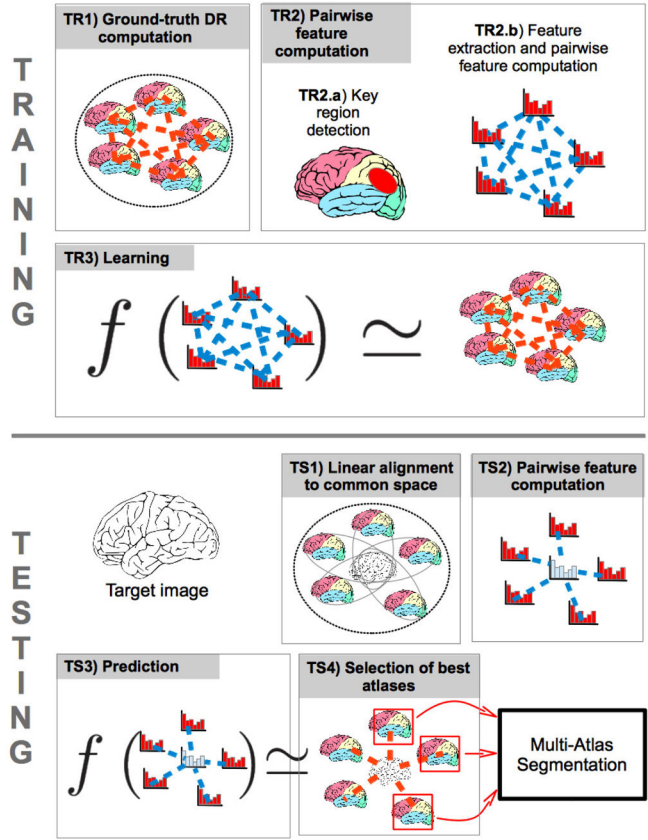


Figure 1. Overview of our method. **Training:** TR1) computation of Dice ratio (DR) between segmented atlases, TR2) computation of pairwise features from key regions between each pair of atlases, and TR3) learning of relationships between pairwise features and ground truth DR. **Testing:** TS1) linear alignment of target image to the common space, TS2) computation of pairwise features between target image and all the atlases, TS3) prediction of the segmentation performance using the learned model and, TS4) selection of atlases with the highest scores to be used for multi-atlas segmentation.

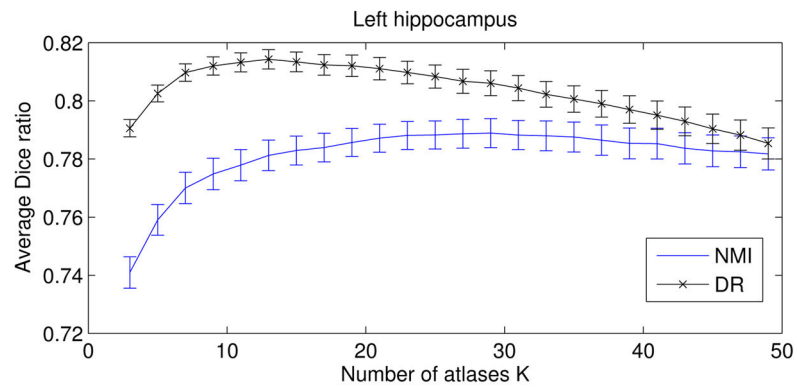


Figure 2. Average segmentation accuracy of 66 images using an increasing number of atlases selected by NMI and the ground-truth DR (equation (2)), respectively. Here we show the results of applying LWV to segment the left hippocampus in the ADNI dataset.

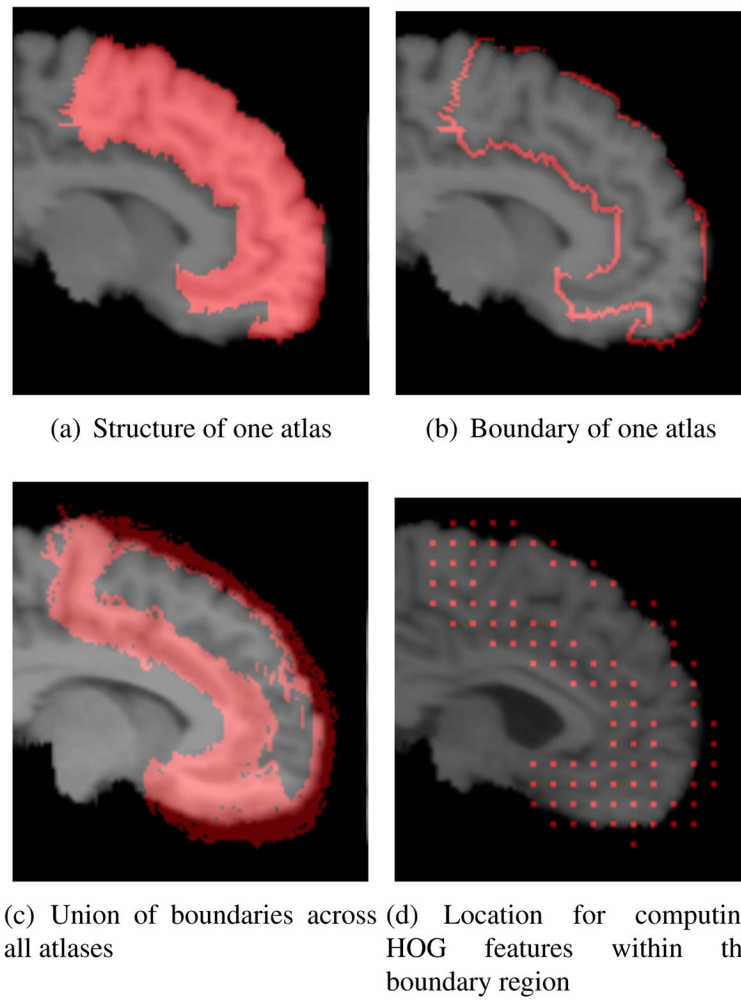


Figure 3.

(a) Frontal gyrus from one exemplar, (b) its boundary, (c) the union of the boundaries from the linearly aligned atlases, and (d) centers of the selected HOG spatial bins in the boundary region.

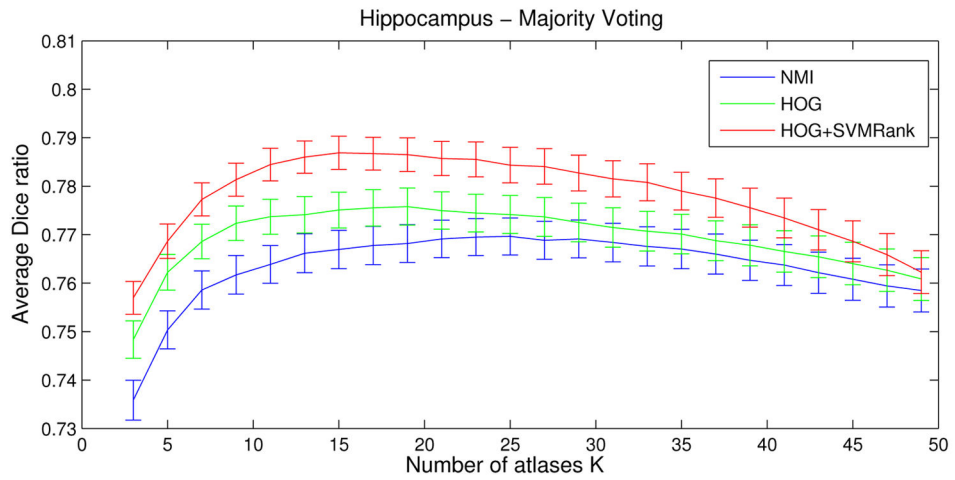


Figure 4.
Label fusion with majority voting

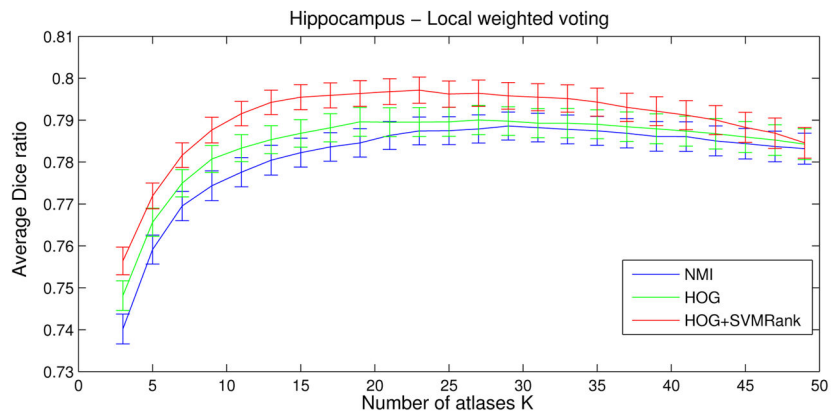


Figure 5.
Label fusion with local weighted voting

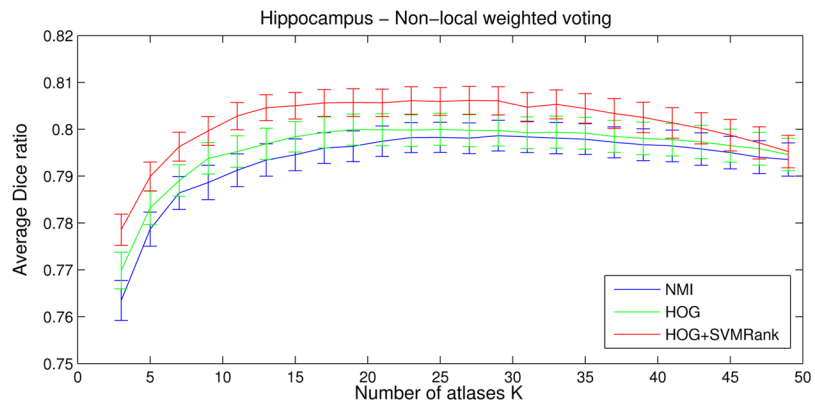


Figure 6.
Label fusion with non-local weighted voting

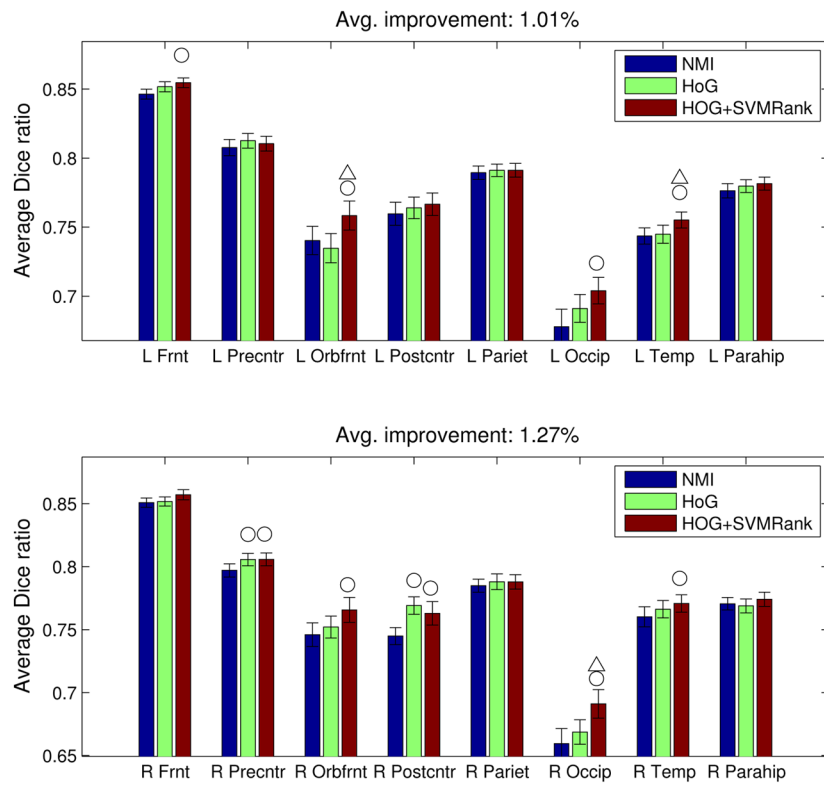


Figure 7.
Label fusion with majority voting

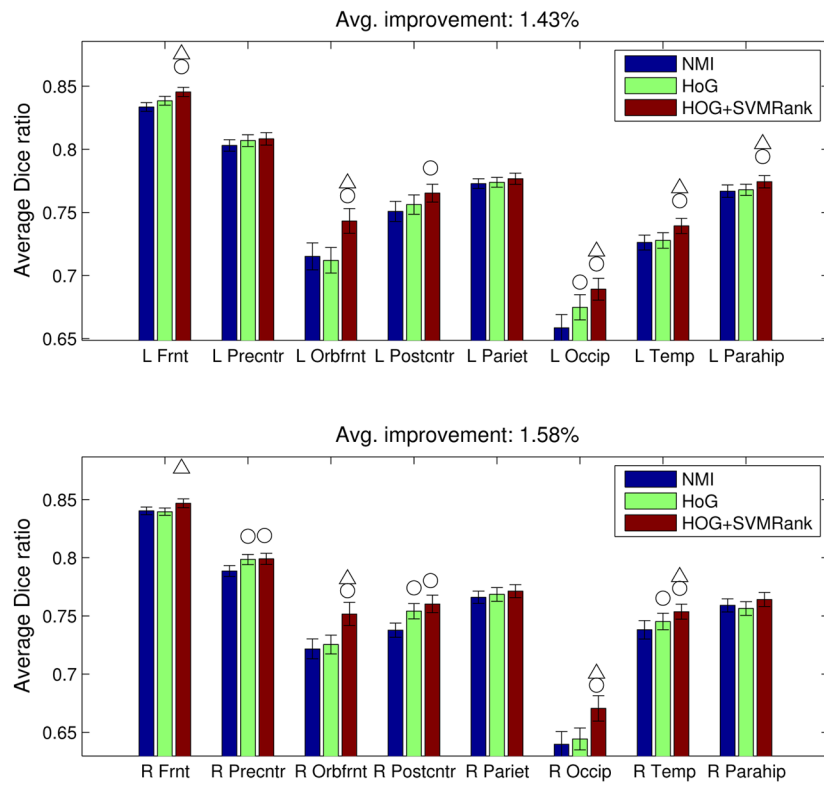


Figure 8.
Label fusion with local weighted voting

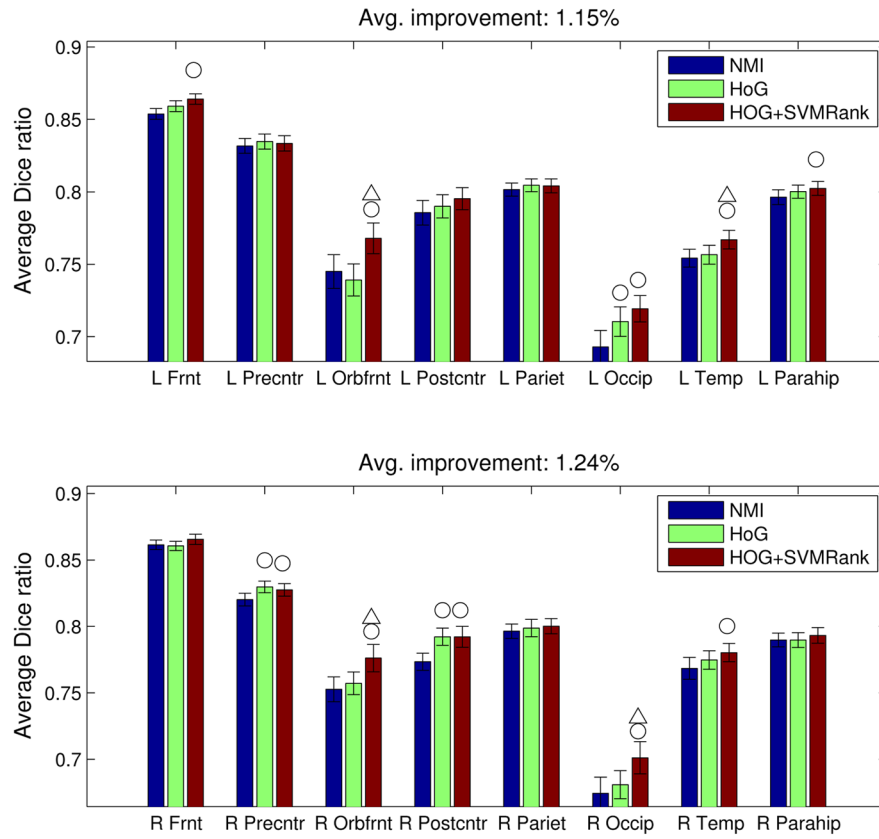


Figure 9.
Label fusion with non-local weighted voting