Local Label Learning (L3) for Multi-Atlas based Segmentation

Yongfu Hao¹, Jieqiong Liu², Yunyun Duan³, Xinqing Zhang², Chunshui Yu⁴, Tianzi Jiang¹, and Yong Fan*¹

¹National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
Departments of ²Neurology and ³Radiology, Xuanwu Hospital of Capital Medical University, Beijing, 100053, China
⁴Department of Radiology, Tianjin Medical University General Hospital, Tianjin, China

ABSTRACT

For subcortical structure segmentation, multi-atlas based segmentation methods have attracted great interest due to their competitive performance. Under this framework, using deformation fields generated for registering atlas images to the target image, labels of the atlases are first propagated to the target image space and further fused somehow to get the target segmentation. Many label fusion strategies have been proposed and most of them adopt predefined weighting models which are not necessarily optimal. In this paper, we propose a local label learning (L3) strategy to estimate the target image’s label using statistical machine learning techniques. Specifically, we use Support Vector Machine (SVM) to learn a classifier for each of the target image voxels using its neighboring voxels in the atlases as a training dataset. Each training sample has dozens of image features extracted around its neighborhood and these features are optimally combined by the SVM learning method to classify the target voxel. The key contribution of this method is the development of a locally specific classifier for each target voxel based on informative texture features. The validation experiment on 57 MR images has demonstrated that our method generates segmentation results of hippocampal with a dice overlap of 0.908±0.023 to manual segmentations, statistically significantly better than state-of-the-art segmentation algorithms.

Keywords: Multi-Atlas based Segmentation, Local Label Learning, Hippocampal Segmentation, SVM

1. INTRODUCTION

Automatic segmentation of subcortical structures from structural MR images is a challenging task because of overlapped intensity distribution and blurred boundary between foreground and background¹,². Many segmentation techniques have been proposed to automatically segment subcortical structures and their performance hinges on how and to what extent the image appearance and structure shape information is modeled. Due to that the spatial positions of the anatomical structures stay relatively stable across subjects, atlas based methods have attracted great interest recently. Such methods achieve segmentation by first registering one atlas image to the target image so that its associated atlas label is propagated to the target image space, and then using the propagated segmentation label as the estimated segmentation of the target image. Since anatomical variability between the target image and the atlas images may degrade the segmentation performance, multi-atlas based methods have been favored. Similar to the single atlas based segmentation, multi-atlas based segmentation methods propagate each atlas label to the target image space and then fuse multiple labels to estimate the target label.

Given multiple propagated labels, how to fuse labels plays a very important role in the multi-atlas based segmentation. Several label fusion strategies focusing on optimizing weights of atlas labels have been developed and most of them use predefined weighting models with parameters to be estimated empirically. Majority Voting is the simplest one which gives each atlas an equal weight³. STAPLE uses EM strategy to iteratively estimate the performance of each atlas label and use its performance as a weight to fuse multiple labels⁴. Recent studies have demonstrated that the segmentation performance can be improved by locally weighting different atlases using intensity information of the target and atlas.

* yfan@nlpr.ia.ac.cn; yong.fan@ieee.org
images. However, the intensity itself may not be a robust and informative feature for segmenting subcortical structures due to their overlapped intensity distribution and blurred boundary between foreground and background.

To overcome limitations of the existing label fusion methods, we develop a machine learning strategy, referred to as local label learning (L3), to directly learn a classification function for the target image segmentation based on training data. For each of the target image voxel, we first find its neighboring voxels in the atlas images as training samples. Then, dozens of texture features are extracted from each training sample’s surrounding sub-volume and each training sample’s associated class label is obtained from the atlas label information. Finally, we use support vector machine (SVM) to construct a classifier on these training samples and the classifier is applied to the target image voxel for determining its segmentation label. The key contribution of this method is the development of massive locally specific classifier for each target image voxel with informative texture features which are optimally combined by the learning method. A fast SVM implementation, i.e., LSVM, makes the algorithm computationally manageable. The validation experiment of segmenting hippocampus for 57 MR images has demonstrated that our method obtains segmentation results with 0.908±0.023 Dice overlap to manual segmentations, statistically significantly better than state-of-the-art segmentation algorithms.

2. METHODS

We first briefly present the traditional label fusion methods, then detail the proposed method and highlight its difference from the existing techniques. We denote a target image by \(T^{img}\) and \(N\) registered atlases by \(S_1 = (S_1^{img}, S_1^{lab}), \ldots, S_N = (S_N^{img}, S_N^{lab})\), where \(S_i^{img}\) is the atlas image and \(S_i^{lab}\) is its associated segmentation label with value +1 indicating foreground and -1 indicating background.

2.1. Multi-atlas based Segmentation

The multi-atlas based image segmentation algorithm first spatially registers the atlas images to the target image \(T^{img}\) and propagates the atlas labels to the target image space, and then the multiple propagated labels are fused to generate a segmentation result \(T^{lab}\) using a specific label fusion strategy. For simplicity, we use \(S_i = (S_i^{img}, S_i^{lab})\) to donate the registered atlas which has been transformed to the target image space.

The label of each target image voxel \(x\) can be generally estimated as follow:

\[
T^{lab}(x) = \arg \max_{l \in \{+1, -1\}} \sum_{i=1}^{N} w_i(x) p(S_i^{lab}(x) = l) \tag{1}
\]

where \(w_i(x)\) is the weight assigned to the atlas label \(S_i^{lab}\) at position \(x\), \(l\) indexes the possible labels -1 or +1, and \(p(S_i^{lab}(x) = l)\) is the probability that \(S_i^{lab}\) belongs to label \(l\) at \(x\). One popular representation of \(p(S_i^{lab}(x) = l)\) can be formulated as:

\[
p(S_i^{lab}(x) = l) = \begin{cases} 1, & \text{if } S_i^{lab}(x) = l \\ 0, & \text{otherwise.} \end{cases} \tag{2}
\]

The difference among different fusion strategies lies in how to define the weight \(w_i(x)\). The majority voting strategy assigns equal weights to atlases, i.e., \(w_i(x) = 1/N\). However, due to anatomical variability and the registration accuracy between atlases and target image, atlases might contribute differently to the final segmentation of the target image. To take into consideration the atlas difference, one can use simultaneous truth and performance level estimation (STAPLE) which was proposed to fuse segmentation results of multiple raters by simultaneously estimating their performance so that different weights can be assigned to raters according to their performance in label fusion. One disadvantage of the above methods is that only the atlas labels are used for label fusion, however, the atlas images may contain information which useful for charactering the weights for each atlas. Recent studies have demonstrated that good segmentation performance can be obtained by making the weights locally adaptive and taking into consideration the image appearance similarity. Applying the zero-mean summed square distance (ZSSD) and a Gaussian model, \(w_i(x) = \exp\left\{-\frac{1}{\sigma(x)} \sum_{y \in N(x)} \|S_i^{img}(y) - T^{img}(y)\|^2\right\}\) can be used to define the weights, where \(N(x)\) is a neighborhood around \(x\) and \(\sigma(x)\) is the model parameter controlling the weight distribution.
Most of the existing image similarity based local weighting methods need to explicitly specify a weighting scheme which is not necessarily the best fit for the label fusion problem. So we propose to learn the link between image appearance and label by a supervised learning strategy using atlas images and labels as training set. Once all the atlases are registered to the target image space using a pair-wise non-linear image registration algorithm, instead of explicitly specifying an atlas weighting scheme and estimating weights for the atlases, a supervised learning method, referred to as local label learning (L3), is utilized to learn a classifier locally for each voxel in the target image space.

![Diagram showing the framework of our proposed local label learning (L3) method, illustrating how the voxel (blue) in the target image T is labeled by training samples extracted from the atlases $S_1 \cdots S_N$.](image)

**2.2. Local Label Learning (L3) strategy**

The procedure of our method consists of three steps: Candidate training set construction, K nearest neighbor (KNN) searching and SVM classification. Fig.1 schematically shows the pipeline of proposed method.

1) **Candidate training set construction**

A set of training samples will be extracted from the registered atlases for each voxel of the target image. Traditionally, most multi-atlas based segmentation methods directly take the corresponding voxel of each atlas for label fusion. However, such a strategy may not be appropriate as the image registration of atlases cannot achieve perfect alignment of all image voxels across images. In\(^{10,11}\), the authors proposed a local search algorithm to find the best match in each atlas for the target image voxels. However, the number of training samples is limited to the number of atlases as each voxel can only has one correspondence in each atlas.

Instead of constraining one-to-one correspondence between the target image and each atlas image, a local patch based method is adopted here\(^{12,13}\). Given one voxel $x$ of the target image, as illustrated as a blue square donated in Fig.1, voxels in its neighborhood $N(x)$ of all atlases are used as training samples. In this paper, we take a $L \times L \times L$ cube-shaped neighborhood and get a total number of $N \cdot L^3$ training samples $\{(\vec{f}_{i,j}, l_{i,j})|i = 1, ..., N, j \in N(x)\}$ from $N$ atlases, where $L$ is the width of neighborhood, $\vec{f}_{i,j}$ is a feature vector extracted from voxel $j$ of the $i^{th}$ atlas by feature extraction to be described in the following, and $l_{i,j} \in \{+1,-1\}$ is the associated segmentation label.

To the best of our knowledge, only the image intensity information has been used in atlas-based segmentation studies for label fusion. However, image intensity is not a proper feature for separating subcortical structures apart since most subcortical structures share similar intensity patterns in MR images, which has been demonstrated in several subcortical segmentation studies\(^{1,2}\). To better characterize the property of each training sample, a set of low level features are extracted in our method. The features extracted or each voxel include intensities in its $3 \times 3 \times 3$ neighborhood, positions, output of mean filters, gradient filters and curvature filters\(^{14}\). All the extracted features are concatenated as a vector $\vec{f}_{i,j}$ as mentioned above.
2) **KNN searching**

After the candidate training set construction step, we have a set of training samples for each voxel of the target image. As a local patch based method is used, the number of training samples is larger than the number of atlases, which makes the trained classifier more resistant to noises/outliers. However, there may also contain irrelevant samples in the training set which may have negative impact on the training classifier. Furthermore, the computation burden is high if all the samples are used as we need train a classifier for each voxel.

Based on the above consideration, a KNN strategy is adopted to find the most relevant samples for each testing voxel. From all the candidate samples, approximate $K$ training samples are extracted using the $K$ nearest neighbor rule based on their Euclidean distances to the testing voxel. The selected samples are illustrated in the dashed black circle in Fig. 1.

3) **SVM classification**

After KNN searching, a supervised learning method is used to train a classifier for segmentation. Particularly, SVM is adopted to build binary classifiers due to its superior performance. The SVM classifier automatically combines features to obtain an optimal separating hyperplane. Furthermore, SVM’s sample selection mechanism is a good fit for our training sample set which may contain irrelevant samples. As we need to train a classifier for each of the target voxels, a fast SVM algorithm is chosen. Particularly, we adopt Lagrangian support vector machine (LSVM) algorithm which constructs a classifier using unconstrained differentiable convex programming. A publicly available software package of LSVM is adopted in our algorithm implementation (research.cs.wisc.edu/dmi/lsvm/). The separating hyperplane found by LSVM is illustrated by the black line in the dash black circle as showed in Fig. 1. Once a classifier is available, it can be used to classify the target voxel.

2.3. **Initial segmentation with majority voting**

The computation time of the L3 method is high as we need to train a classifier for each voxel of the target image. So we propose to give an initial segmentation of the target image before applying our L3 method. The purpose of this procedure is to save the segmentation time of our proposed method without degenerate the segmentation performance. Such a procedure is based on the observation that the difficulty of segmenting voxels in different regions is different, voxels far away from the structure boundary are easy to be segmented and can be successfully segmented by simple and fast methods, while voxels around the structure boundary are difficult to be segmented.

Base on this consideration, we try to first estimate an initial segmentation by majority voting. For each voxel, the output of the majority voting based fusion is a probability value of the voxel belonging to the structure. Our experiment results show that voxels labeled as either 1 or -1 with 100% certainty can be correctly segmented with a percentage of close to 100%. On average only one voxel labeled with 100% certainty was misclassified in each image. Therefore, our algorithm can focus on the voxels with probability values greater than 0 and smaller than 1.

3. **EXPERIMENTAL RESULTS**

3.1 **Data**

The proposed method is validated using hippocampus segmentation on 57 T1-weighted MR images collected in a study of Alzheimer’s disease. Sagittal T1-weighted MR images were acquired using a 3.0 Tesla Siemens scanner with a magnetization prepared rapid gradient echo (MP-RAGE) sequence (TR/TE = 2000/2.6 ms; FA = 9°; slice thickness = 1 mm, no gap). These images (20 controls, 15 MCI patients and 22 AD patients) were manually segmented by a trained expert. Ten of them were also segmented by another expert and the inter-rater variability measured by relative overlap was $0.8907 \pm 0.002$.

Leave-one-out cross-validation is used to evaluate our method. Before further processing, “N3” bias field correction is applied to the images for reducing intensity inhomogeneity, then, images from all the datasets were registered to the MNI152 space using affine registration, and these aligned images were resampled with a voxel size of $1x1x1$ mm$^3$. For each of the left and right hippocampal, a bounding box was generated to cover the whole hippocampus in the MNI152 space.
The parameters of our algorithm are determined empirically, including the neighborhood length $\zeta$ for getting the candidate training samples, the value of $K$, and LSVM’s parameter $\nu$ which is equivalent to $C$ of the traditional SVM$^7$. We set $L$ to 3, $K$ to 50 and the LSVM’s parameter $\nu$ to 0.1 based on the consideration of computation speed and segmentation performance. Other parameters of LSVM are the default values (tolerance set as $10^{-5}$, max iteration set as 100 and $\alpha = 1.9/\nu$).

3.2 Atlas selection and Registration

For each testing image, 20 most similar images are selected from the remaining 56 images. As all images have been aligned to a MNI152 space, for each testing image, the top 20 most similar atlases are selected based on normalized mutual information (NMI) between the testing image and the atlas images$^{16}$. After atlas selection, The Symmetric Normalization (SyN) algorithm implemented by ANTS$^{17}$ is then used to register these atlases to the target testing image. Our proposed L3 method needs about 2 minutes to label one side of the hippocampal without considering the image registration time.

3.3 Comparison with state-of-the-art label fusion Methods

Three state-of-the-art label fusion algorithms are used as benchmarks. The first benchmark is majority voting. The second benchmark is STAPLE$^{8,9}$. The third benchmarks is LWGU$^5$, which fuses the atlas labels by local weighting based on intensity similarity between target image and the atlas images. There are two parameters need to determined, $\sigma$ is adaptively set as $\sigma(x) = \min_{y\in H(x)} \left\{ \| \hat{c}^{\text{img}}(y) - T^{\text{img}}(y) \|^2 + \varepsilon \right\}$, $\varepsilon$ (set as $1e-20$) is a small constant to ensure numerical stability. $N(x)$ is set to $3 \times 3 \times 3$ which is determined using Leave-one-out cross-validation.

The Dice overlap is chosen for the comparison of segmentation performance of different methods. By donating $A$ as the manual segmentation, $B$ as the automated segmentation, and $V(X)$ as the volume of segmentation $X$, the metrics are defined as: Dice overlap = $2 \frac{V(\text{A} \cap \text{B})}{V(\text{A}) + V(\text{B})}$.

Fig. 2 shows the dice overlap scores of both right and left hippocampus of the segmentation results generated by three benchmarks and our proposed method. The scores of the proposed method (0.908±0.023 for the right and 0.906±0.025 for the left) are statistically significantly higher than other methods’ results (measured using Paired t-tests). Table 1 shows the dice results of different hippocampal segmentation methods reported in recently published literature. It can be seen that our method performs favorably with those published results. For visual inspection of the proposed algorithm’s performance, representative segmentation results are shown in Fig. 3. Scatter plots for hippocampal volumes obtained from manual segmentation vs. L3 is shown in Fig. 4 with diagnostic groups denoted. Although large anatomical variability may degrade the segmentation performance as we use a dataset consisting of subjects of discrete diagnostic groups, our hippocampal segmentation results still compare favorably to the recent published results.
Table 1. The performance of different hippocampal segmentation methods reported in recent published literatures.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Dice</th>
<th>Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-atlas based method$^3$</td>
<td>0.82</td>
<td>30 Normal controls</td>
</tr>
<tr>
<td>Machine learning based method$^{18}$</td>
<td>0.85</td>
<td>15 subjects(No population description)</td>
</tr>
<tr>
<td>FreeSurfer initialized atlas based method$^{19}$</td>
<td>0.86</td>
<td>4 normal controls</td>
</tr>
<tr>
<td>Auto Context Model$^{20}$</td>
<td>0.860/0.852(Left/right)</td>
<td>7/7/7(Normal/MCI/AD)</td>
</tr>
<tr>
<td>Multi-atlas based method$^{21}$</td>
<td>0.887</td>
<td>80 normal controls</td>
</tr>
<tr>
<td>Multi-atlas+error correction$^{22}$</td>
<td>0.908</td>
<td>57 normal controls</td>
</tr>
<tr>
<td>Non-local patch based method$^{13}$</td>
<td>0.884</td>
<td>80 normal controls</td>
</tr>
<tr>
<td>L3</td>
<td>0.906/0.908(left/right)</td>
<td>20/15/22(Normal/MCI/AD)</td>
</tr>
</tbody>
</table>

Fig.3. Segmentation results generated by Majority Voting (MV), STAPLE, Local Weighting (LW) and our proposed method (L3) on three cases. Red: Manual Segmentations, Blue: Automated Segmentations, Green: Overlap between Manual and Automated Segmentations.

Fig.4. Scatter plots for hippocampal volumes obtained from the manual segmentation and our proposed automated segmentation method L3.
4. CONCLUSION

In this paper, a supervised learning method is used to improve the performance of multi-atlas based segmentation method. Instead of finding the optimal weights to fuse the labels in the traditional multi-atlas based framework, we propose to learn the target label from training samples. Particularly, we use the target voxel’s local neighbors in the atlases as training samples and texture features are extracted to capture intensity information, resulting powerful SVM classifiers for optimally fuse labels. Direct comparison with state-of-the-art label fusion methods on the same dataset and indirect comparison with the recent published results demonstrate that L3 can achieve competitive performance for hippocampal segmentation.

ACKNOWLEDGMENTS

This work was partially supported by the National Basic Research Program of China (973 Program) 2011CB707801, the Hundred Talents Program of the Chinese Academy of Sciences, and the National Science Foundation of China (Grant No. 30970770, 91132707, 60831004).

REFERENCES


