Progressive Label Fusion Framework for Multi-atlas Segmentation by Dictionary Evolution

Yantao Song[1,2], Guorong Wu[2], Quansen Sun[1], Khosro Bahrami[2], Chunming Li[3], and Dinggang Shen[2]

[1] School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing, Jiangsu, China
[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA
[3] School of Electronic Engineering, University of Electronic Science & Technology, Chengdu, Sichuan, China

Abstract. Accurate segmentation of anatomical structures in medical images is very important in neuroscience studies. Recently, multi-atlas patch-based label fusion methods have achieved many successes, which generally represent each target patch from an atlas patch dictionary in the image domain and then predict the latent label by directly applying the estimated representation coefficients in the label domain. However, due to the large gap between these two domains, the estimated representation coefficients in the image domain may not stay optimal for the label fusion. To overcome this dilemma, we propose a novel label fusion framework to make the weighting coefficients eventually to be optimal for the label fusion by progressively constructing a dynamic dictionary in a layer-by-layer manner, where a sequence of intermediate patch dictionaries gradually encode the transition from the patch representation coefficients in image domain to the optimal weights for label fusion. Our proposed framework is general to augment the label fusion performance of the current state-of-the-art methods. In our experiments, we apply our proposed method to hippocampus segmentation on ADNI dataset and achieve more accurate labeling results, compared to the counterpart methods with single-layer dictionary.

1 Introduction

Accurate and fully automatic segmentation is in high demand in many imaging-based studies. For instance, hippocampus is known as an important structure related with Alzheimer's disease, temporal lobe epilepsy and schizophrenia. Consequently, many neuroscience and clinical applications aim to seek for the imaging biomarker around hippocampus, which is indispensable of accurate segmentation of hippocampus from the MR brain images.

Recently, multi-atlas patch-based segmentation methods [1-5] have achieved many successes in medical imaging area. In current multi-atlas based methods,

N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 190–197, 2015.
DOI: 10.1007/978-3-319-24574-4_23

a set of patches, collected in a searching neighborhood and across all registered atlases, form a patch dictionary to represent the target image patch. In these methods, the assumption is that the representation profile obtained in the image (continuous) domain can be directly transferred to the (binary) domain of anatomical label. However, there is no evidence that such profile is domain-invariant. As a result, representation coefficients may not guarantee the optimal label fusion results.

To alleviate this issue, we propose a novel label propagation framework to progressively transfer the representation profile from the image domain to the anatomical label domain. To achieve it, we construct a set of intermediate dictionaries, which are eventually a sequence of milestones guiding the above domain transition. Then we apply the label fusion techniques (e.g., non-local mean [1, 2] and sparse representation [3, 6]) in a leave-one-out manner to obtain the representation profile for each atlas patch in each layer dictionary where all other instances are regarded as the atlas patches. Then, we can compute a label probability patch by applying the obtained representation profile to the respective label patches. Repeating the above procedure to all patches, we can iteratively construct the higher layer dictionaries, as the probability map within each label probability patch becomes sharper and shaper, until all label probability patches end up to the binary shapes of the corresponding label patches.

Given the learned multi-layer dictionary at each image point, the final weights for voting the label are also estimated in a progressive way. Starting from the initial layer, we gradually refine the label fusion weights by alternating the following two steps: (1) compute the representation profile of target image patch by using the patch dictionary in the current layer; and (2) refine the label probability map within the target image patch by applying the latest representation profile to the binary label patches, where the new probability patch is used as the new target in the next layer. In this way, we can gradually achieve the optimal weights for determining the anatomical label, under the guidance of the intermediate dictionary at each layer.

The contributions of our proposed multi-layer dictionary method include: (1) Since we harness the multi-layer dictionary to remedy the gap between patch appearance and anatomical label, our label fusion essentially seeks for the best label fusion weights, instead of only the optimal patch-wise representation; (2) The evolution of intermediate dictionaries allows us to use not only appearance features but also structure context information [7], which significantly improves the robustness in patch representation; (3) the framework of progressive patch representation by multi-layer dictionary is general enough to integrate with most of conventional patch-based segmentation methods and improve their segmentation performances instantly. Our proposed method has been evaluated in a specific problem of segmenting hippocampus from elderly brain MR images in the ADNI dataset. More accurate segmentation results have been achieved, with comparison to the state-of-the-art non-local mean [2] and sparse patch-based label fusion methods [6].

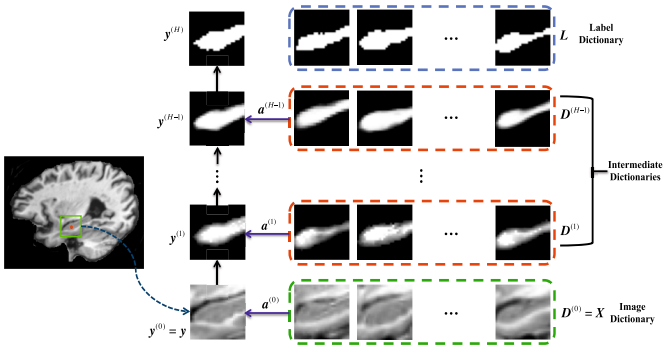

Fig. 1. The framework of proposed method. Given the image dictionary (green dash box) and the label dictionary (blue dash box). In order to overcome the significant gap between two different dictionaries, our method uses a set of intermediate dictionaries (red dash boxes) to gradually encode the transition from the representation coefficients in image domain to the optimal weights for label fusion. In the label fusion stage, we sequentially go through the intermediate dictionaries and obtain the final binary label via a set of probability maps, which become sharper and sharper as the layer increases.

2 Proposed Method

In general, multi-atlas patch-based segmentation aims to determine the label of each point in the target image T by using a set of N registered atlas images I_s as well as the registered label images L_s, $s = 1, \ldots, N$. For each voxel v in the target image, most of patch-based approaches construct a patch dictionary which consists of all patches extracted from the search neighborhood across all atlases. Without loss of generality, we assume there are K candidate atlas patches in the intensity patch dictionary $\boldsymbol{X} = [\boldsymbol{x}_k]_{k=1,\ldots,K}$, where we vectorize each patch into a column vector $\boldsymbol{x}_k$ and turn $\boldsymbol{X}$ into a matrix. Since each atlas patch has the label information, it is straightforward to construct a corresponding label patch dictionary $\boldsymbol{L} = [\boldsymbol{l}_k]_{k=1,\ldots,K}$, where each $\boldsymbol{l}_k$ is the column vector of labels coupled with $\boldsymbol{x}_k$. A lot of label fusion strategies have been proposed to propagate the labels from $\boldsymbol{L}$ to the target image voxel v, mainly driven by the patch-wise similarity α_k between each atlas patch $\boldsymbol{x}_k$ in $\boldsymbol{X}$ and the image patch $\boldsymbol{y}$ extracted at v. For example, non-local mean method [1,2] penalizes patch-wise appearance discrepancy in an exponential way as below

$$\alpha_k = \exp(-||\boldsymbol{y} - \boldsymbol{x}_\mathrm{k}||^2/2\sigma^2) \tag{1}$$

where σ controls the penalty strength. Instead of computing α_k independently, sparse patch based label fusion method (SPBL) [3,6] casts the optimization of weighting vector $\boldsymbol{\alpha} = [\alpha_k]_{k=1,\ldots,K}$ into the sparse patch representation scenario by

$$\arg min_{\boldsymbol{\alpha}} ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_1 \tag{2}$$

where λ controls the sparsity strength.

Hereafter, we call the weighting vector $\boldsymbol{\alpha}$ as the *patch representation profile*. Given the profile $\boldsymbol{\alpha}$ optimized based on appearance information, the latent label on v is assigned to the anatomical label which has the largest accumulated weights within $\boldsymbol{\alpha}$. However, there is a large gap between the intensity dictionary $\boldsymbol{X}$ and label dictionary $\boldsymbol{L}$ as shown in Fig.1. In order to make the patch representation profile $\boldsymbol{\alpha}$ eventually to be the optimal weighting vector for the label fusion, we construct a set of intermediate dictionaries to augment the single-layer dictionary $\boldsymbol{X}$ to the H-layer dictionary $\boldsymbol{D} = \{\boldsymbol{D}^{(h)} | h = 0, \ldots, H-1\}$ and gradually transform the representation profile from the purely appearance representation profile $\boldsymbol{\alpha}^{(0)}$ to the final optimal label fusion weighting vector $\boldsymbol{\alpha}^{(H-1)}$, where for each $\boldsymbol{\alpha}^{(h)}$ we get the corresponding probability map $\boldsymbol{y}^{(h)}$. As $\boldsymbol{\alpha}^{(h)}$ getting more and more reliable for label domain, the probability map becomes sharper and shaper, and eventually the probability map ends up to the binary shape $\boldsymbol{y}^{(H)}$ as shown in the top of Fig.1.

2.1 Dictionary Construction

To construct the multi-layer patch dictionary, we use the original image patch dictionary $\boldsymbol{X}$ to form the initial layer dictionary $\boldsymbol{D}^{(0)} = \boldsymbol{X}$ as shown in the bottom of Fig. 1, i.e., $\boldsymbol{D}^{(0)} = [\boldsymbol{d}_k^{(0)}]$, where $\boldsymbol{d}_k^{(0)} = \boldsymbol{x}_k$. From the first layer, we iteratively construct the intermediate dictionaries $\boldsymbol{D}^{(h)} (h = 1, \ldots, H-1)$ by alternating the following three steps.

First, starting from $h = 1$, for each instance $\boldsymbol{d}_k^{(h-1)}$ in the previous dictionary $\boldsymbol{D}^{(h-1)}$, we seek to use all the other instances $\boldsymbol{d}_j^{(h-1)} (j \neq k)$ in $\boldsymbol{D}^{(h-1)}$ to represent the underlying $\boldsymbol{d}_k^{(h-1)}$ by regarding that all instances in $\boldsymbol{D}^{(h-1)}$ form the instance-specific dictionary $\boldsymbol{B}_k^{(h)} = [\boldsymbol{d}_j^{(h-1)}]_{j=1,\ldots,K, j\neq k}$, where $\boldsymbol{B}_k^{(h)}$ has $K-1$ column vectors. Thus, we can obtain the patch representation profile $\boldsymbol{\beta}_k^{(h-1)}$ for $\boldsymbol{d}_k^{(h-1)}$ via current label fusion strategy, e.g., either non-local mean in Eq.(1) or sparse representation technique in Eq.(2). Note, $\boldsymbol{\beta}_k^{(h-1)}$ is the column vector of length $K-1$.

Second, since each atom in $\boldsymbol{B}_k^{(h)}$ is associated with one label patch in $\boldsymbol{L}$, we can build the surrogate label patch dictionary $\boldsymbol{L}_k = [\boldsymbol{l}_j]_{j=1,\ldots,K, j\neq k}$ by arranging the label patches with the same order as in $\boldsymbol{B}_k^{(h)}$. Then, we compute the label probability patch $\boldsymbol{p}_k^{(h)}$ by $\boldsymbol{p}_k^{(h)} = \boldsymbol{L}_k \cdot \boldsymbol{\beta}_k^{(h-1)}$.

Third, after repeating the above two steps for all instances $\boldsymbol{d}_k^{(h-1)}$, we evolve the intermediate patch dictionary $\boldsymbol{D}^{(h-1)}$ to the next level $\boldsymbol{D}^{(h)}$ by letting $\boldsymbol{D}^{(h)} = [\boldsymbol{d}_k^{(h)}]$, where $\boldsymbol{d}_k^{(h)} = \boldsymbol{p}_k^{(h)}$.

2.2 Multi-layer Label Fusion

Given the multi-layer dictionary $\boldsymbol{D}$, the conventional single-layer patch representation turns to the progressive patch representation where the weighting vector

$\boldsymbol{\alpha}$ is gradually refined from $\boldsymbol{\alpha}^{(0)}$ at the initial layer (optimal for the patch appearance representation only) to $\boldsymbol{\alpha}^{(H-1)}$ in the last layer (eventually optimal for the label fusion).

In the initial layer, we use the original intensity patch dictionary $\boldsymbol{D}^{(0)}$ to present the target image patch vector $\boldsymbol{y}^{(0)} = \boldsymbol{y}$ located at v and thus obtain the representation profile $\boldsymbol{\alpha}^{(0)}$ of the initial layer. Conventional label fusion methods stop here and then vote for the label via the weights in $\boldsymbol{\alpha}^{(0)}$. Instead, our progressive label fusion method computes the label probability vector $\boldsymbol{y}^{(1)}$ by letting $\boldsymbol{y}^{(1)} = \boldsymbol{L}\boldsymbol{\alpha}^{(0)}$. It is worth noting that the intensity target image vector $\boldsymbol{y}$ turns to the probability vector at this time. After the initial layer, we iteratively refine the probability map within the target image patch until it approaches the binary shape of labels. Specifically, we use $\boldsymbol{y}^{(1)}$ as the new target and continue to represent $\boldsymbol{y}^{(1)}$ by the intermediate dictionary $\boldsymbol{D}^{(1)}$ in the same layer, obtaining the new label probability vector $\boldsymbol{y}^{(2)}$. Then, we continue to represent $\boldsymbol{y}^{(2)}$ in the second layer through the intermediate dictionary $\boldsymbol{D}^{(2)}$, and so on. After repeating the same procedure until we reach the last layer $H-1$, the representation profile $\boldsymbol{\alpha}^{(H-1)}$ is regarded as the best weighting vector to determine the latent label on the target image point v.

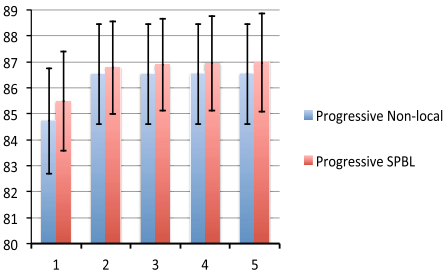

Fig. 2. The evolution curve of Dice ratio as the number of layers in the intermediate dictionary increases.

3 Experiments

In this section, we evaluate the performance of our proposed method on hippocampus segmentation. Specifically, we integrate two state-of-the-art label fusion methods, i.e., non-local [2] and SPBL [6], into our progressive label fusion framework. For comparison, conventional non-local and SPBL methods are used as reference, which only use the single-layer dictionary. Since our method computes the label fusion weights in H layers, the computation time is H times slower than the conventional single-layer method. However, we have used various strategies to speed up our algorithm, such as parallel programming and patch pre-selection.

3.1 Dataset and Parameters

We randomly select 64 normal subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (*www.adni-info.org*), where the hippocampus have been manually labeled for each subject. In our experiments, we regard those manual segmentations as ground truth. To label the target subject, we first aligned all the atlas images to the underlying target subject. In order to improve computational efficiency, atlas selection and patch selection strategies are applied.

In the following experiments, we fix the patch size as $5\times5\times5$ voxels, where the voxel size is $1mm$ in each direction, and the search window for constructing the dictionary as $5\times5\times5$ voxels. In non-local mean method, the penalty strength σ is set to 0.5, while in sparse patch based label fusion method the sparse constraint λ is set to 0.1. Here, we use the Dice ratio to measure the overlap between automatic segmentation and ground truth. Also as it is common in evaluation of label fusion method, all testing images were evaluated in a leave-one-out manner. Specifically, in each leave-one-out case, we use FLIRT in the FSL toolbox [8] with 12 degrees of freedom and the search range ±20 in all directions. For deformable registration, we use diffeomorphic Demons method [9] with smoothing kernel size 1.5. The iteration numbers in diffeomorphic Demons are 15, 10, and 5 in the low, middle, and high resolutions, respectively.

Table 1. The mean and standard deviation of Dice ratio (in %) in hippocampus labeling in the *linear* registration scenario

Method	Left	Right	Overall
Conventional Non-local	85.1±5.7	84.3±5.1	84.7±4.2
*Progressive Non-local	86.8±4.5	86.2±5.1	86.5±3.7
Conventional SPBL	85.8±4.5	85.1±4.8	85.5±3.7
Progressive SPBL	**87.1±3.2**	**86.7±5.3**	**86.9±3.3**

Table 2. The mean and standard deviation of Dice ratio (in %) in hippocampus labeling in the *deformable* registration scenario

Method	Left	Right	Overall
Conventional Non-local	86.8±4.9	86.6±2.9	86.7±3.2
*Progressive Non-local	87.9±4.0	88.1±3.2	88.0±3.0
Conventional SPBL	87.2±3.6	87.1±3.3	87.2±2.9
*Progressive SPBL	**88.2±3.6**	**88.5±3.1**	**88.3±2.8**

The evolution of segmentation accuracy with the number of layers used is shown in Fig.2. We can see that the improvement of our progressive method is

obvious after one layer (corresponding to the baseline methods), which offers more than 1% improvement of Dice ratio for both non-local and SPBL methods. Our progressive label fusion framework generally converges after the third layer. Considering the computation time, we use 4 layers ($H = 4$) in the following experiments.

3.2 Hippocampus Segmentation Results

Table 1 and Table 2 show the mean and standard deviation of Dice ratio on hippocampus (left, right and overall) in linear and deformable registration scenarios, respectively. Compared to the baseline methods (i.e., non-local and SPBL methods with single-layer dictionary), our progressive label fusion framework can improve the labeling accuracy with more than **1%** of Dice ratio. Maximum improvement is **1.8%** (conventional non-local vs. progressive non-local in linear registration case). The significant improvement of our method over the baseline method, with p-value less than 0.05 using paired t-test, is indicated with '*' in Table 1 and Table 2, respectively.

Table 3. The surface distance (in mm) on hippocampus labeling between automatic segmentations and ground truth with different number of layers

Number of Layers	H=1	H=2	H=3	H=4	H=5
Maximum Distance	2.83	2.00	1.52	1.22	1.21
Mean Distance	0.27	0.22	0.19	0.18	0.18

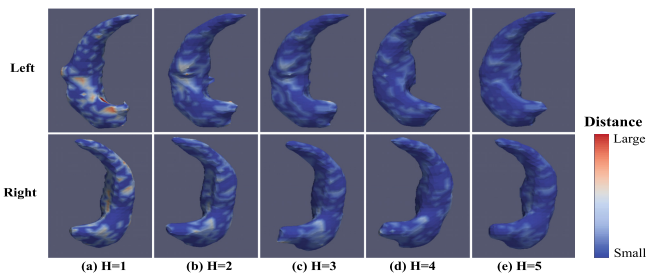

Fig. 3. The evolution of surface distance between the automatic segmentations and ground truth from the initial layer (a) to the last layer (e).

Furthermore, we calculate the surface distance between ground truth and the estimated hippocampus (left and right). SPBL method is used as the example to demonstrate the evolution of surface distance during the progressive label fusion procedure in Fig. 3. According to the color bar shown in the right side of Fig. 3,

the surface distances keep decreasing by increasing the number of layers in the intermediate dictionary. Table 3 shows the corresponding surface distance on the whole hippocampus in Fig. 3. In accordance with Fig. 3, as the layer number increases, the mean surface distance becomes smaller and smaller. When $H = 1$, which corresponds to the conventional one-layer SPBL method, the maximum distance is 2.83mm, and it significantly decreases to 1.21mm at $H = 5$ by our method.

4 Conclusion

In this paper, we proposed a progressive label fusion framework for multi-atlas segmentation by dictionary evolution. In our proposed methods, we constructed a set of intermediate dictionaries in a layer-by-layer manner to progressively optimize the weights for label fusion, instead of just using patch-wise representation as used in the conventional label fusion methods. We have applied our new label fusion method to hippocampus segmentation in MR brain images. Promising results were achieved over the state-of-the-art counterpart methods with the single-layer dictionary.

References

1. Coupe, P., Manjon, J.V., Fonov, V., et al.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. NeuroImage 54(2), 940–954 (2011)
2. Rousseau, F., Habas, P.A., Studholme, C.: A Supervised Patch-Based Approach for Human Brain Labeling. IEEE Trans. Medical Imaging 30(10), 1852–1862 (2011)
3. Tong, T., Wolz, R., Coupe, P., et al.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage 76(1), 11–23 (2013)
4. Wang, H., Suh, J.W., Das, S.R., et al.: Multi-atlas segmentation with joint label fusion. IEEE Trans. Pattern Anal. Mach. Intell. 35(3), 611–623 (2013)
5. Wu, G., Wang, Q., Zhang, D., et al.: A Generative Probability Model of Joint Label Fusion for Multi-Atlas Based Brain Segmentation. Medical Image Analysis 18(8), 881–890 (2014)
6. Zhang, D., Guo, Q., Wu, G., Shen, D.: Sparse patch-based label fusion for multi-atlas segmentation. In: Yap, P.-T., Liu, T., Shen, D., Westin, C.-F., Shen, L. (eds.) MBIA 2012. LNCS, vol. 7509, pp. 94–102. Springer, Heidelberg (2012)
7. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 32(10), 1744–1757 (2010)
8. Smith, S.M., Jenkinson, M., Woolrich, M.W., et al.: Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219 (2004)
9. Vercauteren, T., Pennec, X., Perchant, A., et al.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1), S61–S72 (2009)