

Hierarchical Label Fusion with Multiscale Feature Representation and Label-Specific Patch Partition

Guorong Wu and Dinggang Shen

BRIC and Department of Radiology, University of North Carolina at Chapel Hill, USA
{grwu, dgshen}@med.unc.edu

Abstract. Recently, patch-based label fusion methods have achieved many successes in medical imaging area. After registering atlas images to the target image, the label at each target image point can be subsequently determined by checking the patchwise similarities between the underlying target image patch and all atlas image patches. Apparently, the definition of patchwise similarity is critical in label fusion. However, current methods often simply use entire image patch with fixed patch size throughout the entire label fusion procedure, which could be insufficient to distinguish complex shape/appearance patterns of anatomical structures in medical imaging scenario. In this paper, we address the above limitations at three folds. *First*, we assign each image patch with multiscale feature representations such that both local and semi-local image information can be encoded to increase robustness of measuring patchwise similarity in label fusion. *Second*, since multiple *variable* neighboring structures could present in one image patch, simply computing patchwise similarity based on the entire image patch is not specific to the particular structure of interest under labeling and can be easily misled by the surrounding *variable* structures in the same image patch. Thus, we partition each atlas patch into a set of new label-specific atlas patches according to the existing label information in the atlas images. Then, the new label-specific atlas patches can be more specific and flexible for label fusion than using the entire image patch, since the complex image patch has now been semantically divided into several distinct patterns. *Finally*, in order to correct the possible mis-labeling, we hierarchically improve the label fusion result in a coarse-to-fine manner by iteratively repeating the label fusion procedure with the gradually-reduced patch size. More accurate label fusion results have been achieved by our hierarchical label fusion method with multiscale feature presentations upon label-specific atlas patches.

1 Introduction

Many medical imaging based studies demand accurate segmentation of anatomical structures, in order to quantitatively measure structure differences across individuals or between two groups. To this end, automatic ROI (Region of Interest) labeling has been a hot topic in medical image processing areas, as evidenced by hundreds of labeling and label fusion methods that have been developed to improve both segmentation accuracy and robustness.

In order to deal with high structural variations in the population, multiple atlases with delineated labels are commonly used for labeling the latent ROIs of the target image [1]. The basic assumption behind multi-atlas based segmentation is that the target image should bear the same label as the atlas image if both of them present similar shape/appearance. Thus, all atlas images are required to be registered to the target image before label fusion. To alleviate the possible mis-registration, patch-based label fusion technique [1, 2] is also advocated by measuring the patchwise similarity at each point. Intuitively, the higher the similarity between target image and a particular atlas image is, the more confidence we assign the label on that atlas to the target image.

It is apparent that the patchwise similarity is the key in patch-based label fusion methods. Most of the current state-of-the-art methods only use the fixed patch size throughout entire label fusion procedure. For example, $7 \times 7 \times 7$ or $9 \times 9 \times 9$ cubic patches are usually used in the literature. In order to make the label fusion robust to noise, image patches are required to be large enough in order to capture sufficient image content. However, large image patch could raise a critical issue in labeling small anatomical structures, since the patchwise similarity could be dominated by the surrounding large structures in the image patch. The main reason for such dilemma is that the simple use of whole image patch lacks high-level knowledge to distinguish complex appearance patterns in medical imaging data.

Many efforts have been made to improve the discrimination power of image patches. For instance, sparse dictionary learning technique is used in [3] to find the best feature representations in label fusion. However, the dictionary is still confined in using the whole image patch with fixed size. In this paper, we address the above limitations in a new perspective of developing hierarchical and high-level feature representations for image patch. In general, our contribution has three folds.

First, we propose to adaptively treat each image point within the image patch by designing the image patch with multi-scale feature representations. We argue that image points close to the patch center should use fine-scale features to characterize the details of patch center, while the level of image features could gradually turn from fine to coarse as the distance toward the patch center increases. To this end, we assign the conventional image patch with the *layerwise* multi-scale feature representation by adaptively capturing image features in each layer with different scale.

Second, it is very common that the to-be-segmented ROI, e.g., hippocampus, is surrounded by other complex structures. Those surrounding *variable* structures may mislead the patchwise similarity measurement. In computer vision area, recognizing object could be much easier if the foreground pattern can be separated from the background clutters [4]. In light of this, we present a new concept of label-specific patch partition to enhance the discriminative power of each atlas patch in label fusion. Specifically, since each atlas patch bears the well-determined labels, such information can provide the valuable heuristic about anatomical structures and thus can be used to guide the splitting of each atlas patch into a set of new complementary label-specific (or structure-specific) image patches. It is worth noting that each label-specific image patch carries only the image information at selected locations with same label. Therefore, our label-specific partition *not only* enriches the representations for each atlas

patch *but also* encapsulates the high-level label information. To the best of our knowledge, such important label information is poorly used in the current label fusion methods. Afterwards, sparsity constraint is further used in our proposed label fusion method to deal with the increased number of label-specific image patches.

Third, current label fusion methods fix the patch size throughout the entire label fusion procedure. Here, we go one step further, e.g., propose to iteratively refine the labeling results by gradually reducing the patch size with the progress of label fusion. Specifically, we use the large image patches in the beginning, in order to make the label fusion robust. Sparsity constraint is used to allow only a small number of atlas patches for joining the label fusion. Then, for those selected atlas patches, we can reduce their patch size and repeat the label fusion procedure to refine their respective weights in the final label fusion.

We comprehensively evaluate the performance of our new label fusion method both in segmenting hippocampus in ADNI dataset and labeling 54 ROIs in LPBA40 dataset. More accurate labeling results are achieved, compared with the state-of-the-art label fusion methods.

2 Methods

Given the target image T , the goal of label fusion is to automatically determine a label map L_T for the target image T . To achieve it, we need to first register all atlas images as well as their labeled maps to the target image space. Here, we use $\mathbf{I} = \{I_s | s = 1, \dots, N\}$ and $\mathbf{L} = \{L_s | s = 1, \dots, N\}$ to denote N registered atlases and label maps, respectively. For each target image point x ($x \in T$), all the atlas patches¹ within a certain search neighborhood $n(x)$, denoted as $\vec{\beta}_{s,y}$ ($\vec{\beta}_{s,y} \subset I_s$, $y \in n(x)$), are used to compute the patchwise similarities w.r.t. the target image patch $\vec{\alpha}_{T,x}$ ($\vec{\alpha}_{T,x} \subset T$). It is worth noting that we arrange each patch, $\vec{\beta}_{s,y}$ and $\vec{\alpha}_{T,x}$, into a column vector.

Next, label fusion strategies, e.g., non-local averaging, can be used to calculate the weighting vector $\vec{w} = [w_{s,y}]_{s=1, \dots, N, y \in n(x)}$ for each atlas patch $\vec{\beta}_{s,y}$. As we will explain in Section 2.2, we adopt the sparsity constraint in our method by regarding the label fusion procedure as the problem of finding optimal combination among a set of atlas patches $\{\vec{\beta}_{s,y}\}$ for the target image patch $\vec{\alpha}_{T,x}$ [5, 6]:

$$\hat{\vec{w}} = \arg \min_{\vec{w}} \|\vec{\alpha}_{T,x} - \mathbf{B}\vec{w}\|^2 + \lambda \|\vec{w}\|_1, \text{ s.t. } \vec{w} > 0 \quad (1)$$

where the scalar λ controls the strength of sparsity constraint and \mathbf{B} is the matrix by assembling all column vectors $\{\vec{\beta}_{s,y}\}$ in a columnwise way. Assuming that we have M possible labels $\{l_1, \dots, l_m, \dots, l_M\}$ in the atlases, then the label on target image point x can be efficiently determined by:

$$\hat{L}_T(x) = \arg \max_{m=1, \dots, M} \sum_{s=1}^N \sum_{y \in n(x)} [w_{s,y} \cdot \delta(L_s(y), l_m)] \quad (2)$$

¹ Some label fusion methods use patch pre-selection to discard the less similar patches.

where Dirac function $\delta(L_s(y), l_m)$ is always zero except for the case when $L_s(y)$ bears the label l_m . In that case, $\delta(L_s(y), l_m)$ equals to 1.

It is clear that image intensities in the entire image patch are used in label fusion (Eq. 1). Since one image patch may contain more than one anatomical structures and the to-be-segmented ROI may have very complex shape/appearance pattern, current patch-based label fusion methods have high risk of being misled by the current definition of patchwise similarities that are computed based on the entire image patch. In the following, we propose three ways to improve the label fusion accuracy: **(1)** substantially upgrading the feature discrimination power by using multi-scale feature representations (Section 2.1); **(2)** adaptively building label-specific atlas patches by using the existing label information in the atlases (Section 2.2); and **(3)** hierarchically improving label fusion accuracy in a coarse-to-fine manner by gradually reducing the patch size (Section 2.3).

2.1 Multi-scale Feature Representations

In current patch-based label fusion methods, every point in the image patch uses its own intensity value and equally contributes in computing the patchwise similarity. Here, we allow each point to use adaptive scale for capturing local appearance characteristics. Specifically, we first partition the whole image patch into several nested non-overlapping *layers*, spreading from the center point to the bound of image patch. Next, we use small scale to capture the fine-scale features for the layer closest to the patch center. Gradually, we use larger and larger scale to capture the coarse-scale information as the distance to the patch center increases. Although advanced pyramid image technique can be applied for multiscale feature representation, we choose a more efficient way by replacing the intensity value with the average intensity in a certain neighborhood, due to the consideration of computational time. For example, for the points in the first layer that is the closest to the patch center (including the patch center and its 6 immediate neighboring points), we still keep using their original intensities. For each point in the second layer, we replace its intensity value with the average intensity value in its $3 \times 3 \times 3$ neighborhood. Similarly, we use intensity average in a larger neighborhood as the feature representation for the image points beyond the second layer. In this way, the image patch is now equipped with the multi-scale feature representation. Hereafter, $\vec{\alpha}_{T,x}$ and $\vec{\beta}_{s,y}$ denote the image patches after replacing the original intensities with the multi-scale feature representations.

2.2 Label-Specific Atlas Patch Partition

Since atlas image patches have label information, we can partition each atlas patch into a set of new label-specific atlas patches for encoding the label information. Given the atlas patch $\vec{\beta}_{s,y}$, we use $\vec{\gamma}_{s,y}$ to denote its associated labels. Suppose there are M kinds of labels in $\vec{\gamma}_{s,y}$. Then, the proposed label-specific atlas patch set $\mathbf{P}_{s,y}$ consists of M label-specific atlas patches, i.e., $\mathbf{P}_{s,y} = \{\vec{p}_{s,y}^m | m = 1, \dots, M\}$, where $\vec{p}_{s,y}^m$ is the column vector. Each element u in $\vec{p}_{s,y}^m$ keeps the intensity value $\vec{\beta}_{s,y}(u)$ if and only

if $\vec{\gamma}_{s,y}(u)$ has label l_m ; otherwise, $\vec{p}_{s,y}^m(u) = 0$. Mathematically, we have $\vec{p}_{s,y}^m(u) = \vec{\beta}_{s,y}(u) \cdot \delta(\vec{\gamma}_{s,y}(u), l_m)$, where $\delta(\cdot, \cdot)$ is the same Dirac function as used in Eq. 2.

Note that the number of image patches increases significantly after we partition each atlas patch into the label-specific atlas patch set. Thus, we propose to use the sparsity constraint again in label fusion, in order to select only a small number of label-specific atlas patch $\vec{p}_{s,y}^m$ to represent the target image patch $\vec{a}_{T,x}$. By replacing each conventional atlas patch with label-specific atlas patches, the matrix of atlas patches \mathbf{B} in Eq. 1 now expands to $\mathbf{P} = [\mathbf{P}_{s,y}]_{s=1,\dots,N,y \in n(x)}$. Then, the new energy function for label fusion can be reformulated as:

$$\hat{\xi} = \arg \min_{\xi} \|\vec{a}_{T,x} - \mathbf{P}\xi\|^2 + \lambda \|\xi\|_1, \text{ s.t. } \xi > 0, \quad (3)$$

where $\xi = [\xi_{s,y}^m]$ is the weighting vector for each label-specific atlas patch $\vec{p}_{s,y}^m$. Since each $\vec{p}_{s,y}^m$ is only related with a particular label l_m , each element $\xi_{s,y}^m$ in ξ represents the probability of labeling the center point x of the target image patch $\vec{a}_{T,x}$ by label l_m . Therefore, the labeling result on the target image point x can be obtained by:

$$\hat{L}_T(x) = \arg \max_{m=1,\dots,M} \sum_{s=1}^N \sum_{y \in n(x)} \xi_{s,y}^m \quad (4)$$

Fig. 1 demonstrates the construction of label-specific atlas patch set \mathbf{P} for the case with only two labels, i.e., $M = 2$. As displayed in Fig. 1(a), each atlas patch $\vec{\beta}_{s,y}$ is split into two label-specific atlas patches $\vec{p}_{s,y}^1$ and $\vec{p}_{s,y}^2$, where we use the black to denote the zero elements. For example, the zero elements in $\vec{p}_{s,y}^1$ have their label as l_2 , instead of l_1 . The objective function in Eq. 1 is to minimize the appearance difference between $\vec{a}_{T,x}$ and $\mathbf{B}\vec{w}$. In our method, we first divide each whole atlas patch into several label-specific patches and then recognize the structural patterns in $\vec{a}_{T,x}$ in a label-by-label manner. In this way, our method makes the representation of $\vec{a}_{T,x}$ more selective and flexible.

The advantage of using label-specific atlas patches is demonstrated by the toy example in Fig. 1(b), where we use red and blue to denote two different labels and numbers represent the intensity values. To be simple, only two atlas patches are used in this example. Apparently, the first atlas patch (first column in \mathbf{B}) and $\vec{a}_{T,x}$ belongs to the same structure since their intensity values are both in the ascending order. If we estimate the weighting vector \vec{w} based on the entire atlas patch by Eq. 1 ($\lambda = 0.01$), the weights for the first and second atlas patches are 0.43 and 0.49, respectively. According to Eq. 2, we have to assign the target point with the blue (incorrect) label. In our method, we first extend the matrix \mathbf{B} to label-specific atlas patch set \mathbf{P} , as shown in the bottom of Fig. 1(b) and then solve the new weighing vector ξ by Eq. 3. As suggested by ξ , the overall weights for red and blue labels are 0.885 (0.88+0.005) and 0.800 (0.69+0.11), respectively. Therefore, we can correctly assign the target point with red label. This example demonstrates the power of our method.

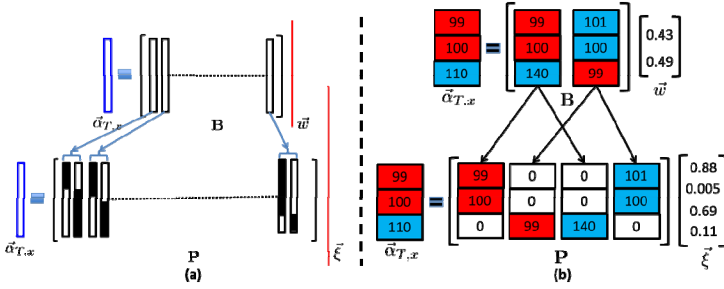


Fig. 1. (a) Construction of label-specific atlas patch set and (b) the advantage in label fusion

2.3 Hierarchical Patch-Based Label Fusion

In the beginning of patch-based label fusion, we often use a large patch size in order to obtain global image information. Since we use the sparsity constraint in solving the weighing vector $\tilde{\xi}$, only a small number of image patches are selected to represent the target image patch $\tilde{\alpha}_{T,x}$, as many weights in $\tilde{\xi}$ are zero or almost zero. After discarding those non-selected atlas patches, we are more confident to reduce the patch size of those selected atlas patches and then repeat the whole label fusion procedure as described in Section 2.1 and 2.2 by using more detailed local features. In this way, our label fusion method can iteratively improve the labeling results in a hierarchical way.

3 Experiments

In the following experiments, we compare our label fusion method (by Eq. 3 and Eq. 4) with the sparse patch-based label fusion method (by Eq. 1 and Eq. 2). To label the target image, we first use FLIRT in FSL package to linearly register all atlas images onto the target image and then use diffeomorphic Demons [7] to compute the remaining local deformations². After optimization, λ is set to 0.1 for both the label fusion methods. The patch size is $5 \times 5 \times 5$ for the sparse patch-based label fusion method. In our method, the patch size is initialized with $11 \times 11 \times 11$ in the first iteration and reduced to $5 \times 5 \times 5$ in the second iteration. Here, we use Dice ratio on each ROI to measure the labeling accuracy.

3.1 Evaluation on Hippocampus Labeling

In this experiment, we randomly select 66 elderly brains from ADNI dataset³, where hippocampus has been manually labeled for each brain. Besides comparing with the

² The main parameters for running diffeomorphic Demons are: 15, 10, and 5 iterations in low, middle, and high resolution, respectively. The smoothing kernel size is 2.0.

³ www.adni-info.org

baseline sparse patch-based label fusion method (*Sparse PBL*), to evaluate the contribution of each component in our label fusion method, we further compare our method with the three degraded versions of our method, (1) Degraded_1: our method using only the multi-scale feature representation (with patch size $11 \times 11 \times 11$), (2) Degraded_2: our method using only the label-specific atlas patches (with patch size $11 \times 11 \times 11$), and (3) Degraded_3: our method using only the hierarchical labeling mechanism.

We evaluate each of the above five label fusion methods with a leave-one-out cross-validation. From all 66 leave-one-out cases, the mean and standard deviation of Dice ratios in hippocampus and the surface distance are calculated and provided in Table 1. It is clear that: (1) Our full method achieves the highest Dice ratio and lowest surface distance over other four comparison methods, where we obtain almost 1.2% improvement over the baseline *Sparse PBL* method; (2) Each component in our label fusion method has contribution in improving the labeling accuracy, as evidenced by 0.6%, 0.9%, and 0.3% Dice ratio increases over the baseline *Sparse PBL* by Degraded_1, Degraded_2, and Degraded_3, respectively. Also, we find all degraded methods have significant improvement over the baseline method in paired *t*-test.

Table 1. The statistics of Dice ratios in hippocampus labeling by 5 different methods

	Sparse PBL	Degraded_1	Degraded_2	Degraded_3	Our method
Dice Ratio	87.3±3.4	87.9±3.0	88.2±2.5	87.6±2.9	88.5±2.2
Surf. Dist	0.38mm	0.35mm	0.34mm	0.35mm	0.33mm

3.2 Evaluation on LPBA40 Dataset

LPBA 40 dataset⁴ consists of 40 MR brain images, each with 54 manually labeled ROIs. We randomly select 20 images as atlases and another 20 as the target images. The statistics of overall Dice ratio across 54 ROIs are given in Table 2, where our full method achieves 1.5% improvement over the baseline *Sparse PBL* method. Apparently, each component in our proposed label fusion method has its contribution in enhancing the label fusion results. Fig. 2 shows the Dice ratio in each left-and-right-combined ROI by *Sparse PBL* (in blue) and our full method (in red), from which we can observe significant improvements in 12 out of 27 ROIs (* denoting the significant improvement confirmed by paired *t*-test ($p < 0.05$)).

Table 2. The statistics of Dice ratios in labeling 54 ROIs on LPBA40 dataset by 5 different methods

	Sparse PBL	Degraded_1	Degraded_2	Degraded_3	Our method
Dice Ratio	80.3±3.2	81.1±2.5	81.5±2.4	80.6±3.0	81.8±2.1

⁴ http://www.loni.usc.edu/atlas/Atlas_Detail.php?atlas_id=12

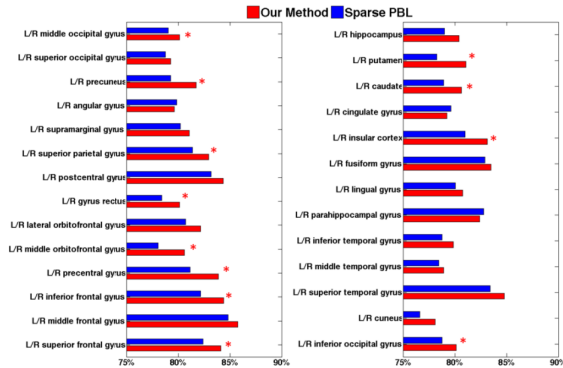


Fig. 2. The Dice ratios of 27 ROIs (left and right combined) in LPBA 40 dataset by *Sparse PBL* (in blue) and our method (in red)

4 Conclusion

In this paper, we explore a new perspective to substantially enhance the discriminative power of the conventional, widely used, image patch in label fusion. Specifically, we assign each atlas patch with multi-scale feature representation, and further develop label-specific atlas patches according to the existing label information in the atlases for making each atlas patch more flexible during label fusion. Moreover, we present a hierarchical label fusion mechanism to iteratively improve the labeling results by gradually reducing the patch size. Promising labeling results have been obtained on ADNI and LPBA40 dataset, by comparing with state-of-the-art methods.

References

1. Coupe, P., Manjoh, J., Fonov, V., Pruessner, J., Robles, M., Collins, L.: 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., Hajnal, J., Rueckert, D.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. *NeuroImage* 76, 11–23 (2013)
4. Li, L., Su, H., Xing, E., Li, F.: Object Bank: A High-Level Image Representation for Scene Classification and Semantic Feature Sparsification. In: *Proceedings of the Neural Information Processing System (NIPS)* (2010)
5. Tong, T., et al.: Segmentation of Brain Images via Sparse Patch Representaion. In: *MICCAI Workshop on Sparsity Techniques in Medical Imaging, Nice, France* (2012)
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. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage* 45(suppl. 1), S61–S72 (2009)