



TURN UNCERTAINTY INTO PEACE OF MIND

Get ready for Beam Commissioning 4.0

Be the first to know about the all-new solution
that revolutionizes your Beam Commissioning.

Quality. Performance. Peace of Mind.

Soon to be revealed – at iba-innovations.com

PROTECT +
ENHANCE +
SAVE LIVES

Automatic labeling of MR brain images through extensible learning and atlas forests

Lijun Xu, Hong Liu,^{a)} Enmin Song, Meng Yan, and Renchao Jin
School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
Key Laboratory of Education Ministry for Image Processing and Intelligent Control, Wuhan, Hubei 430074, China

Chih-Cheng Hung

Center for Machine Vision and Security Research, Kennesaw State University, Marietta, GA 30144, USA

(Received 8 December 2016; revised 31 July 2017; accepted for publication 8 September 2017; published 24 October 2017)

Purpose: Multiatlas-based method is extensively used in MR brain images segmentation because of its simplicity and robustness. This method provides excellent accuracy although it is time consuming and limited in terms of obtaining information about new atlases. In this study, an automatic labeling of MR brain images through extensible learning and atlas forest is presented to address these limitations.

Methods: We propose an extensible learning model which allows the multiatlas-based framework capable of managing the datasets with numerous atlases or dynamic atlas datasets and simultaneously ensure the accuracy of automatic labeling. Two new strategies are used to reduce the time and space complexity and improve the efficiency of the automatic labeling of brain MR images. First, atlases are encoded to atlas forests through random forest technology to reduce the time consumed for cross-registration between atlases and target image, and a scatter spatial vector is designed to eliminate errors caused by inaccurate registration. Second, an atlas selection method based on the extensible learning model is used to select atlases for target image without traversing the entire dataset and then obtain the accurate labeling.

Results: The labeling results of the proposed method were evaluated in three public datasets, namely, IBSR, LONI LPBA40, and ADNI. With the proposed method, the dice coefficient metric values on the three datasets were $84.17 \pm 4.61\%$, $83.25 \pm 4.29\%$, and $81.88 \pm 4.53\%$ which were 5% higher than those of the conventional method, respectively. The efficiency of the extensible learning model was evaluated by state-of-the-art methods for labeling of MR brain images. Experimental results showed that the proposed method could achieve accurate labeling for MR brain images without traversing the entire datasets.

Conclusion: In the proposed multiatlas-based method, extensible learning and atlas forests were applied to control the automatic labeling of brain anatomies on large atlas datasets or dynamic atlas datasets and obtain accurate results. © 2017 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.12591>]

Key words: atlas selection, brain MR images, image segmentation, learning, random forest

1. INTRODUCTION

Accurate brain anatomy labeling is a crucial prerequisite for numerous clinical and research applications. However, manual labeling is a time-consuming task because labeling a set of MR brain image requires a specialist to work for 2 or 3 days.¹ With the development and recognition of public atlas datasets, the automatic labeling of MR brain images has been extensively investigated. Many institutions provide atlases with the advancements in medical imaging and clinical applications, which provide basis for a future research. However, the studies have yet to determine how numerous atlases can be used rapidly and efficiently label MR brain images.

In this study, the proposed automatic labeling method through the extensible learning and atlas forests aimed to resolve previously encountered problems. The proposed method mainly includes three steps: training the atlas dataset,

establishment of the extensible learning model and selection atlases for target image, and labeling the target image.

The detailed work of each step is described below:

- (1) In the training stage, random forest strategy is utilized to encode atlases and generate atlas forests by using each voxel of the atlases as a sample. In most currently multiatlas-based methods, all atlases in datasets should be registered to target image in each labeling. However, multiatlas-based methods become time-consuming because of this requirement. In the proposed method, atlases are required to register only once in the training stage to obtain labels; this is due to the use of spatial features in the atlas encoding.
- (2) We construct an extensible learning model and select atlases for the target image without traversing the whole dataset through the model. The

extensible learning model provides a pre-estimation of atlases in the dataset and the model will be dynamically updated during the selection of atlases for a new target.

- (3) The labeled mean atlas which is aligned to the target image will serve as an augmenting channel to provide spatial information as spatial features, and then the labels of the target image are predicted by trained atlas forests.

This paper is organized as follows. Section 2 analyzes the related studies on multiatlas-based methods of MR image labeling. Section 3 discusses the proposed method and its implementation steps. Section 4 shows the experimental results and analysis. Section 5 presents our conclusion.

2. RELATED WORKS

2.A. Multiatlas-based labeling

The multiatlas-based method is efficient and attractive in the field of MR brain image labeling. This method can use the prior knowledge provided by the atlases to obtain an accurate and robust labeling of the target image. In general, the multiatlas-based method is mainly divided into two steps: (a) Register the atlas to the target image to accurately map the information of the atlas to the target image. (b) Use the label fusion strategy to evaluate the labels provided by the atlases to obtain the labeling of the target image.

Registration performance possibly influences labeling accuracy.² Some studies have been performed by nonlinear registration.² These methods can achieve high-accuracy labeling, but several parameters closely related to labeling accuracy are required. A registration scheme without any parameters has been proposed to solve this problem and achieves accuracy at a pixel level.^{3,4} In other studies, all atlases considered rather than a single one in the dataset to improve the registration performance.⁵

Some studies have focused on label fusion. The most common method involved in label fusion is weighted voting. In this method, similarities between the target image and reference atlases are determined to locate the reliability of label propagation. Weighted-based evaluation methods can be local, semilocal, or global.⁶⁻⁹ The patch-based strategy also plays a significant role in multiatlas-based methods. In patch-based methods, the similar image patches should belong to the same area and have the same label.¹⁰ There are various patch-based methods, including sparsity¹¹ and k-NN (k-nearest neighbor) search structure¹² have been developed.

With the development of machine learning technology, machine learning has been introduced to multiatlas-based labeling. Bai et al.¹³ used the augmented features to extract intensity, gradient, and texture information from the atlas and applied SVM (support vector machine) to the label fusion. Hao et al.¹⁴ showed a local label learning strategy to evaluate labels. This function combines SVM and k-NN to find

similar image patches from the entire available atlases and obtain the local information around the target area.

Atlas selection is an efficient mechanism to reduce the time required in multiatlas-based methods. The mechanism uses the similarity among the atlases and target image to select those that provide more valuable information and thus improve labeling accuracy.¹⁵⁻¹⁷ Asman et al.¹⁸ showed that atlas selection improved the performance of majority voting. Langerak et al.¹⁷ filtered a few atlases before registration to save time.

2.B. Application of random forest in medical image segmentation

Random forest is a fast classifier with robustness. It performs two random operations based on the decision tree.¹⁹ Owing to the random strategy, the random forest is more robust than other classifications and is less prone to overfitting.¹⁹⁻²¹ With regard to the preceding advantages mentioned, the random forest is recognized as an efficient method in the field of medical image analyses.²²⁻²⁷

Zikic et al.²⁸ applied random forest to high-grade segment gliomas of multichannel MR images and achieved success. Also, they proposed encoding atlases by randomized classification forests to atlas forests for multiatlas label propagation.^{29,30} For this way, each forest contains intensity, texture, and location information of the corresponding atlas. Therefore, the method is independent on registration accuracy.

The strategy of atlas forest reduces the time of multiatlas-based method for registration. However, the labeling accuracy of the target image does not improve much because the prior spatial information is too simple. Zhang et al.³¹ showed a hierarchical learning based atlas selection method to gather more information from atlases through a hierarchical structure. Their study indicated more excellent results than Zikic et al.³⁰. Meanwhile, the method requires more computational resources than the last one and is nonextensible. Retraining is necessary for the hierarchical model when a new atlas is added to the dataset.

In our proposed method, we apply extensible learning strategy to solve the retraining problem. We used diffuse spatial feature vector to enhance the information of voxel location to improve the labeling accuracy of brain MR images.

3. MATERIALS AND METHODS

This section describes the proposed method in detail. The method is divided into three parts: the first part is the training of atlas forests; the second part is the establishment of self-learning model based on extensible learning strategy and atlases selection through the extensible learning model; and the third part is the labeling of brain MR images. Figure 1 illustrates the overview of our method. According to the framework, this section is arranged as follows. Section 3.A introduces the training method for the atlas forests.

Section 3.B describes the method of establishing a self-learning model. Section 3.C describes how to obtain the labels of the target image by trained atlas forests. Section 3.D shows the feature extraction method of brain MR images. Section 3.E summarizes the framework of the proposed method.

3.A. Encoding atlas by using random forest

Encoding atlas is a prerequisite step of the proposed extensible learning, which allows each atlas not to be affected by other atlases in the dataset during the propagation of labels. In this paper, we use the random forest for atlas encoding strategy^{29,30} to meet the requirements above.

The random forest consists of decision trees with the advantage of bagging. Each tree in the forest chooses samples with replacement and selects features randomly. The random forests generated by the atlas are called the atlas forests.^{29,30} In this paper, we consider a voxel on each atlas image as a sample and extract the features through intensity, texture, and location. The feature extraction is discussed in detail in Section 3.D.

We use the ID3 strategy²¹ to build the decision tree in which the split function is optimized to maximize the information gain of splitting the training set. The decision

tree grows until the subsets of training samples divided by the split function are sufficiently small. Then, each leaf node in the decision tree has a corresponding category label. We select the training samples and features randomly and repeat the procedure above to build decision trees. The decision trees generate through atlases constitute an atlas forest.

Each atlas forest contains the information of the corresponding atlas and it can propagate the labels to the target image independently of other atlases in the dataset. The label fusion method of the trained atlas forests is described in detail in Section 3.C.

3.B. Extensible learning model

Unlike the current multiatlas-based methods in labeling a new target, the proposed approach constructs an extensible learning model which remembers the evaluation of atlases in the dataset while labeling the target so that it is not necessary to traverse the entire dataset for atlases selection.

The extensible learning model is divided into two parts: one is the establishment of the extensible learning model and the other is atlas selection based on the model which will be updated simultaneously.

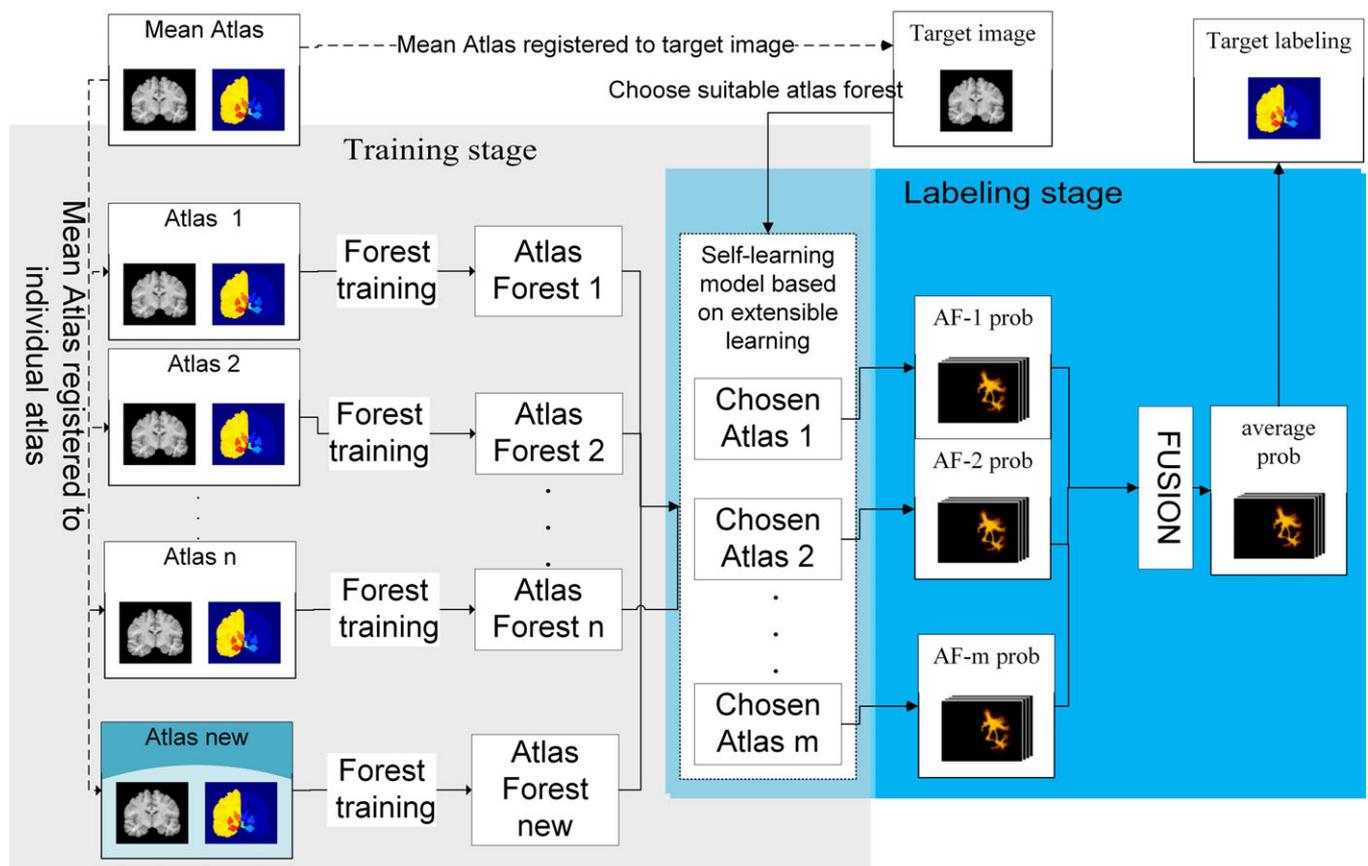


FIG. 1. Overview of the proposed method. In this method, the labeled mean atlas as an augmented channel provides spatial information to the atlases and target image. Each atlas is encoded to the atlas forest in the training stage. An atlas weight table is established in the extensible learning model. The weight of the atlases is changed according to the number of times selected and the timestamp in which the atlas added in the processing of labeling. The extensible learning strategy selects appropriate atlases for target image and target image is then labeled in the labeling stage. [Color figure can be viewed at wileyonlinelibrary.com]

3.B.1. Establishment of the extensible learning model

To construct the extensible learning model, we build a global atlas information list to remember the evaluation of the atlases for a dataset. The atlas information list includes two parameters which determine the priority of the atlas: the selected times of each atlas q_i and the time when the atlas is added to the dataset t_i . The more times the atlas is selected, the higher the priority is. The recently added atlases will have the high priority also. The priority is calculated by Eq. (1) with parameter $\alpha = 1$. The initial values of q_i and t_i are zero, and it can also be set to different values. The extensible learning model automatically updates the atlas information list during atlas selection procession.

$$w_i = q_i - \alpha t_i \quad (\text{when } w_i < 0, w_i = 0) \quad (1)$$

3.B.2 Atlas selection and update of the extensible learning model

The main idea of atlas selection through the extensible learning model is to select a group of priority atlases for test and keep the atlases with high similarity to the target for label fusion. We update the atlas information list to renew the extensible learning model. The selection of priority atlases will be repeated until there is no new atlas added for label fusion. The proposed method uses a selective and iterative method³² for the estimation of similarity between target image and atlases.

The mathematical description of the process is as follows. Assuming that A_i represents the i th atlas in the dataset, S^j represents the set of atlases that have not been evaluated at the j th selection. We define $S^0 = \{A_1, A_2, \dots, A_n\}$ (n is the number of atlases in the dataset). Let S_{choose}^j denote the atlases in the test set at the j th selection and S_{select}^j denote the selected atlases set at the j th selection. Also, let S_{fusion}^k denote the set of atlases to be fused at the k th iteration, L'_i the labels of the target image predicted by atlas forest corresponding to the i th atlas of S_{fusion}^k , and L_{est}^k the labels of target image that is fused by the atlases in the S_{fusion}^k . We use the notation $\phi_i = DSC(L'_i, L_{est}^k)$ which is obtained using the Dice similarity ratio (DSC) between L'_i and L_{est}^k to represent a similarity score of i th atlas in S_{fusion}^k . The symbol $\{\phi\}$ is used to denote the set of ϕ_i . The details are given as follows:

When $S_{select}^j \neq S_{select}^{j-1}$ repeat steps 1–4

Step 1: Select a subset of unevaluated atlases from S^j to compose S_{choose}^j , let $S_{fusion}^0 = S_{choose}^j$ and update $S^{j+1} = S^{j+1} - S_{choose}^j$.

Step 2: Predict L_{est}^k with the corresponding atlases in S_{fusion}^k and compute the dice similarity ratio between L'_i and L_{est}^k by $\phi_i = DSC(L'_i, L_{est}^k)$ in S_{fusion}^k . Remove the atlases with the low similarity to the target from S_{fusion}^k and then generate S_{fusion}^{k+1} for next iteration based on the Eq. (2).

$$\phi_i < \text{mean}\{\phi\} - \alpha \cdot \text{std}\{\phi\} \quad (2)$$

Step 3: Repeat Step 2 until no atlas forests need to be removed, which means $S_{fusion}^k = S_{fusion}^{k-1}$, and then let $S_{select}^j = S_{fusion}^k$.

Step 4: If $S_{select}^j = S_{select}^{j-1}$, then S_{select}^j is the final selected atlases sets for the target image. Update the atlas information table according to Eq. (3) and refresh the weight of atlases according to Eq. (1)

$$q_i = q_i + 1(A_i \in S_{select}) \text{ and } t_i = t_i + 1(A_i \in S^0) \quad (3)$$

We use the extensible learning model to provide the pre-evaluation for the atlases and then choose a fixed-size set of atlases with the highest priority to compose S_{choose}^j to reduce the evaluation times. However, the similar atlases will be clustered in the dataset. When the number of elements in the atlas test set is less than those in a cluster, the test set may select all of the elements from the same cluster. In such a situation, the suitable atlases cannot be properly selected because we cannot find the poorest atlas in the same cluster as each element has the same low similarity to the target image. To avoid this, we add the uncertainty to S_{choose}^j : some elements are from the high-priority atlases and the others are selected randomly in the unevaluated set.

3.C. Label fusion by atlas forests

In the labeling stage, the labels in mean atlas align to the target image as an augmenting channel to provide spatial information for spatial feature, and then the label of the target image is predicted by the trained atlas forests. Assuming that f_{xyz} indicates feature vector of point (x, y, z) of target image I , c is the label of brain tissue, then the probability map of each tree in atlas forest is $p_i(c|f_{xyz})$. An atlas forest F , which includes n_t decision trees, predicts the probability of label c at this point as Eq. (4):

$$p_F(c|f_{xyz}) = 1/n_t \sum_{i=1}^{n_t} p_{t_i}(c|f_{xyz}) \quad (4)$$

Averaging those probability maps predicted from n_a selected atlas forests for fusion as Eq. (5), then the maximum probability of the label is the final result as Eq. (6).

$$p(c|f_{xyz}) = 1/n_a \sum_{i=1}^{n_a} p_{F_i}(c|f_{xyz}) \quad (5)$$

$$\hat{c} = \arg \max_c p(c|f_{xyz}) \quad (6)$$

3.D. Feature extraction

To describe the characteristics of the samples, we extract two types of features: one is based on the images that are used to describe the intensity and local texture, and the other is based on the spatial location as shown in Fig. 2. The spatial feature is used to describe the location of the sample tissue. Given that $p(x, y, z)$ is the point at the site (x, y, z) , $N_s(p)$ is cubic centered at p with the side length s , $I(p)$ is the intensity in p , and μ is the average operation. The notation

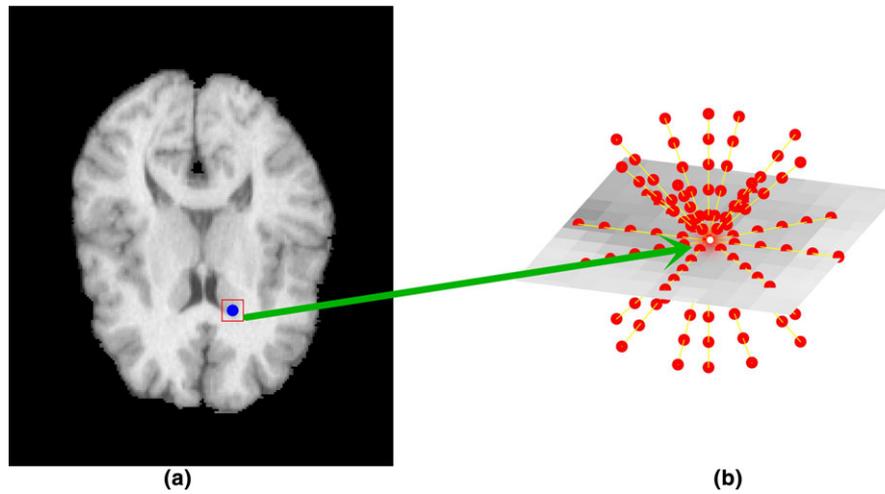


FIG. 2. The detail of spatial features. The blue point in (a) is the sample point. The red points in (b) are diffused from the sample point shown in the white point located in the center. [Color figure can be viewed at wileyonlinelibrary.com]

$p + u$ is the point which has the distance u from $p(x_p, y_p, z_p)$ that obtained by Eq. (7), where $\alpha_x, \alpha_y, \alpha_z = 1$ or 0 and $\alpha_x + \alpha_y + \alpha_z > 0$.

$$p + u = p(x_p \pm \alpha_x u, y_p \pm \alpha_y u, z_p \pm \alpha_z u) \quad (7)$$

A. Local intensity feature is represented in Eq. (8).

$$F_s^{local1}(I, p) = \mu(I(N_s(p))) \quad (8)$$

B. Local texture feature is denoted in Eq. (9).

$$F_s^{local2}(I, p) = I(p) - \mu(I(N_s(p + u))) \quad (9)$$

The labeled mean atlas is registered to the atlas which will be encoded to improve the spatial location information for the spatial feature. Notation L_{space} is the label image of the mean atlas and $L_{space}(p)$ is the label of L_{space} in a location p . The notation $R_r(p)$ represents diffused points centered at p with radius r as in Eq. (10)

$$R_r(p) = p(x_p + r \cos \theta \sin \varphi, y_p + r \sin \theta \sin \varphi, z_p + r \cos \varphi) \\ \theta, \varphi \in \{0, \pm\pi/4, \pm\pi/2, \pm3\pi/4, \pm\pi\} \quad (10)$$

C. Spatial feature is denoted in Eq. (11)

$$F_r^{space} = L_{space}(R_r(p)) \quad (11)$$

3.E. Summary of the extensible learning framework

This section summarizes the proposed method which includes the following three steps:

- (1) In the training stage, each atlas in the dataset is encoded into an atlas forest by following the strategy described in Section 3.A. For each atlas, the local intensity, texture information, and the spatial information are extracted as the features (described in Section 3.D). And then the encoded atlases propaganda label to target image.

- (2) Construct a global atlas information list for establishing the extensible learning model as described in Section 3.C.
- (3) In the testing stage with a given new target image, we extract the features and select the suitable atlases through the extensible learning method and update the learning model as described in Section 3.C. The labels of the target image will be obtained according to the method described in Section 3.B.

For the initialization, the parameters for local intensity and texture features are $s = 11$, $u = 1, 2, \dots, 5$ in Eqs. (7), (8), and (9), and the radius of the spatial feature is $r = 5$ in Eqs. (10) and (11). In the extensible learning model, the atlas information list is set to empty so that each atlas has the same priority, the size of the test atlas set is 10, the random rate of test atlas set is 0.8 (80% elements of the test atlas set come from the high-priority group and 20% elements are selected randomly from the unevaluated atlases in the dataset), and the maximum number of iterations for the procedure is 50.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section first introduces the preprocessing of test data and then verifies the proposed method by leave-one-out cross-validation on the datasets. Finally, we compare the experimental results of the proposed method with the results of other algorithms.

Before the computation, we perform the standard preprocessing in the following order:

- (1) Align prior label to the atlas. Aligning the label of the mean atlases to the images is necessary to provide the spatial information to eliminate the error make by the imprecise registration. We randomly select an atlas as mean atlas and apply it in the experiment to save time. We used elastix 4.7³³ as a registration tool.

- (2) Histogram matching. We used the ITK-based histogram-matching program to the atlas images for overall intensity normalization, and the intensities are resealed to the interval 0 to 255 after normalization.
- (3) Find the ROI (region of interest). We locate the ROI wherein the intensities of the image are greater than zero.

The proposed method computed on a standard PC (CPU i5-4570, memory 16 GB) which was similar to the computer performance used in the comparative experiment.

4.A. Experimental results based on IBRS dataset

The IBRS data contain 18 labeled T1 MR Images, each with 32 manual labels. We use the set of 18 primarily subcortical brain tissues for labeling accuracy assessment. The parameter setting used in our approach reaches a mean dice score of $84.17 \pm 4.61\%$. Figure SA1 shows the labeling of the target image produced by the proposed method.

We used the method proposed in the work of Zhang et al.³¹ as a comparative experiment to evaluate the efficiency of the proposed method. Table I presents the results of the DSC measures for the selected subcortical region in the IBRS, indicating approximately 5% improvement compared with the method proposed by Zhang et al.³¹ after adopting the extensible learning method and spatial feature.

In the proposed method, the encoding time per atlas requires 7 min, in which extracting the features takes

TABLE I. Quantitative comparison of DSC values obtained by the Zhang et al. method³¹ and the proposed method for the 18 labeled primarily subcortical regions in the IBRS dataset.

Brain regions	Zhang et al. method ³¹ (%)	Proposed method (%)
L.lateral ventricle ^{a,b}	85.86 ± 7.51	91.50 ± 6.45
L.thalamus ^{a,b}	87.80 ± 2.73	92.13 ± 1.66
L.caudate	82.83 ± 4.23	82.42 ± 7.93
L.putamen ^{a,b}	81.56 ± 6.53	90.07 ± 2.44
L.pallidum	73.30 ± 7.80	75.22 ± 8.05
3rd Ventricle	76.62 ± 10.41	82.13 ± 4.30
4th ventricle ^{a,b}	75.81 ± 8.35	82.37 ± 4.03
L.hippocampus ^{a,b}	75.90 ± 5.39	81.93 ± 4.96
L.amygdala ^a	69.78 ± 9.89	72.33 ± 7.57
L.VentralDC ^{a,b}	82.47 ± 4.35	83.01 ± 2.72
R.lateral ventricle ^{a,b}	82.18 ± 6.08	91.84 ± 4.90
R.thalamus ^{a,b}	88.21 ± 3.59	92.00 ± 1.83
R.caudate ^{a,b}	84.61 ± 7.93	85.43 ± 4.38
R.putamen	84.67 ± 5.09	90.57 ± 2.57
R.pallidum ^{a,b}	76.00 ± 5.58	81.21 ± 3.86
R.hippocampus ^{a,b}	76.00 ± 5.59	84.18 ± 3.87
R.amygdala ^{a,b}	67.27 ± 8.43	72.94 ± 9.16
R.VentralDC	81.09 ± 3.46	83.76 ± 2.24
Overall	78.64 ± 6.42	84.17 ± 4.61

^aThe label index indicates $P < 0.05$ with the two-tailed paired t -test.

^bThe label index indicates $P < 0.05$ with the Wilcoxon rank-based t -test.

4–5 min and generating the atlas forests takes 2–3 min. This part is offline. In the labeling stage, it takes around 17 s to evaluate whether an atlas is suitable for target image and about 16 atlases are tested before labeling the target image. It takes about 4 min 20 s for atlas selection and labeling a target image.

4.B. Experimental results based on LONI LPBA40 dataset

In the second experiment, we evaluate our proposed method on the LONI LPBA40 dataset. This database contains 40 brain atlases from healthy people; each atlas contains 54 manual labels and a brain MR image. Our strategy obtained the average DSC of 83.34%, indicating approximately 5% of overall improvement for the method of Zhang et al.³¹ Figure SA2 shows the labeling of the target image produced by the proposed method. Tables II and III illustrate the DSC measure of the obtained by the method³¹ and the proposed method.

TABLE II. Quantitative comparison of DSC values obtained by the Zhang et al. method³¹ and the proposed method for the left-hemisphere ROIs in the LONI LPBA40 dataset.

Brain regions	Zhang et al. method ³¹ (%)	Proposed method (%)
Superior frontal gyrus ^{a,b}	85.44 ± 2.91	93.03 ± 1.64
Middle frontal gyrus ^{a,b}	84.42 ± 2.80	90.27 ± 2.32
Inferior frontal gyrus ^{a,b}	79.05 ± 4.36	87.97 ± 4.29
Precentral gyrus ^{a,b}	80.73 ± 4.05	86.06 ± 3.31
Middle orbitofrontal gyrus ^{a,b}	76.19 ± 6.06	84.56 ± 3.69
Lateral orbitofrontal gyrus ^{a,b}	65.23 ± 9.36	73.18 ± 8.36
Gyrus rectus ^{a,b}	76.89 ± 3.82	73.58 ± 5.79
Postcentral gyrus ^{a,b}	77.00 ± 4.92	85.50 ± 4.78
Superior parietal gyrus ^{a,b}	80.25 ± 3.81	88.36 ± 3.28
Supramarginal gyrus ^{a,b}	72.97 ± 6.23	84.52 ± 4.36
Angular gyrus ^{a,b}	75.00 ± 4.80	79.84 ± 6.82
Precuneus ^{a,b}	77.13 ± 4.79	79.48 ± 3.42
Superior occipital gyrus ^{a,b}	70.28 ± 7.61	80.61 ± 6.95
Middle occipital gyrus ^{a,b}	78.38 ± 4.70	83.99 ± 4.41
Inferior occipital gyrus ^{a,b}	74.86 ± 5.30	85.67 ± 4.12
Cuneus ^{a,b}	74.67 ± 6.88	78.03 ± 5.46
Superior temporal gyrus ^{a,b}	82.91 ± 2.71	89.46 ± 2.49
Middle temporal gyrus ^{a,b}	77.46 ± 3.83	81.02 ± 4.01
Inferior temporal gyrus ^{a,b}	77.80 ± 5.71	82.66 ± 4.28
Parahippocampal gyrus ^{a,b}	78.02 ± 4.12	80.33 ± 2.38
Ligular gyrus ^{a,b}	80.01 ± 5.54	79.84 ± 5.05
Fusiform gyrus ^{a,b}	79.82 ± 5.54	79.71 ± 5.31
Insular cortex ^{a,b}	83.33 ± 2.40	87.23 ± 1.02
Cingulate gyrus ^{a,b}	77.10 ± 5.32	79.83 ± 4.71
Caudate ^{a,b}	80.60 ± 4.16	84.55 ± 4.48
Putamen ^{a,b}	81.72 ± 2.54	85.04 ± 2.67
Hippocampus ^{a,b}	80.39 ± 2.46	81.61 ± 3.23
Overall	78.06 ± 4.69	83.25 ± 4.29

^aThe label index indicates $P < 0.05$ with the two-tailed paired t -test.

^bThe label index indicates $P < 0.05$ with the Wilcoxon rank-based t -test.

TABLE III. Quantitative compaction of DSC values obtained by the Zhang et al. method³¹ and the proposed method for the right-hemisphere ROIs in the LONI LPBA40 database.

Brain regions	Zhang et al. method ³¹ (%)	Proposed method (%)
Superior frontal gyrus ^{a,b}	86.61 ± 1.97	93.03 ± 1.64
Middle frontal gyrus ^{a,b}	84.94 ± 3.04	90.27 ± 2.32
Inferior frontal gyrus ^{a,b}	80.08 ± 3.75	87.97 ± 4.29
Precentral gyrus	82.36 ± 4.19	86.06 ± 3.31
Middle orbitofrontal gyrus ^{a,b}	75.33 ± 6.30	84.56 ± 3.69
Lateral orbitofrontal gyrus	69.72 ± 6.89	73.18 ± 8.36
Gyrus rectus	74.82 ± 5.34	73.58 ± 5.79
Postcentral gyrus ^{a,b}	78.00 ± 5.36	85.50 ± 4.78
Superior parietal gyrus ^{a,b}	81.17 ± 2.58	88.36 ± 3.28
Supramarginal gyrus ^{a,b}	75.70 ± 6.89	84.52 ± 4.36
Angular gyrus ^{a,b}	73.76 ± 7.96	79.84 ± 6.82
Precuneus ^{a,b}	77.05 ± 4.56	79.48 ± 3.42
Superior occipital gyrus ^{a,b}	69.78 ± 7.31	80.61 ± 6.95
Middle occipital gyrus ^{a,b}	77.45 ± 6.71	83.99 ± 4.41
Inferior occipital gyrus ^{a,b}	77.89 ± 5.35	85.67 ± 4.12
Cuneus ^{a,b}	74.21 ± 7.14	78.03 ± 5.46
Superior temporal gyrus ^{a,b}	83.61 ± 4.25	89.46 ± 2.49
Middle temporal gyrus ^{a,b}	77.47 ± 4.70	81.02 ± 4.01
Inferior temporal gyrus ^{a,b}	74.80 ± 4.94	82.66 ± 4.28
Parahippocampal gyrus	80.14 ± 3.35	80.33 ± 2.38
Ligual gyrus ^{a,b}	77.77 ± 5.12	79.84 ± 5.05
Fusiform gyrus ^{a,b}	81.68 ± 4.01	79.71 ± 5.31
Insular cortex ^{a,b}	85.27 ± 2.05	87.23 ± 1.02
Cingulate gyrus	78.82 ± 3.27	79.83 ± 4.71
Caudate ^{a,b}	78.81 ± 6.77	84.55 ± 4.48
Putamen ^{a,b}	81.52 ± 2.65	85.04 ± 2.67
Hippocampus ^{a,b}	80.67 ± 2.66	81.61 ± 3.23
Overall	78.50 ± 4.78	83.25 ± 4.29

^aThe label index indicates $P < 0.05$ with the two-tailed paired t -test.

^bThe label index indicates $P < 0.05$ with the Wilcoxon rank-based t -test.

In the proposed method, the encoding time per atlas requires 8 min, in which extracting the features takes 5–6 min with four threads in parallel and generating the atlas

TABLE IV. Quantitative compaction of DSC values obtained by the Zhang et al. method³¹ and the proposed method for the labeled left and right hippocampus in the ADNI dataset.

Brain regions	Zhang et al. method ³¹ (%)	Proposed method (%)
Left hippocampus ^{a,b}	75.72 ± 5.01	81.44 ± 4.32
Right hippocampus ^{a,b}	76.27 ± 5.85	82.32 ± 4.74
Overall	76.00 ± 5.43	81.88 ± 4.53

^aThe label index indicates $P < 0.05$ with the two-tailed paired t -test.

^bThe label index indicates $P < 0.05$ with the Wilcoxon rank-based t -test.

forests with interlaced sampling takes 2–3 min. This part is offline. In the labeling stage, it takes an average of 16 s with eight threads in parallel to evaluate whether an atlas is suitable for target image and about 24 atlases are tested before labeling the target image. It takes about 6 min 24 s for labeling a target image.

4.C. Experimental results based on ADNI datasets

In the third experiment, we evaluate our proposed method on the ADNI dataset. This dataset provides an extensive set of adult brain MR images acquired from 1.5T MR scanners. In the experiment, we randomly selected 100 images with the manual labels of the hippocampus in the ADNI dataset, with 34 from the Normal Control (NC) subjects, 33 from Mild Cognitive Impairment (MCI) subjects, and 33 from Alzheimer’s disease (AD) subjects. The atlases of the dataset come from the difference subject so that we can evaluate the robustness of proposed algorithm and the performance of the extensible learning model for processing the dataset with scores of atlases.

To ensure the robustness of the algorithm, we randomly select each of atlases from the MCI, NC, and AD subjects, respectively, and combines them into an mean atlas to provide the position information. Our method obtained the average DSC of 81.88%, indicating approximately 5% of overall improvement for the method of Zhang et al.³¹ as shown in Table IV. Figure SA3 shows the labeling of the target image by the proposed method.

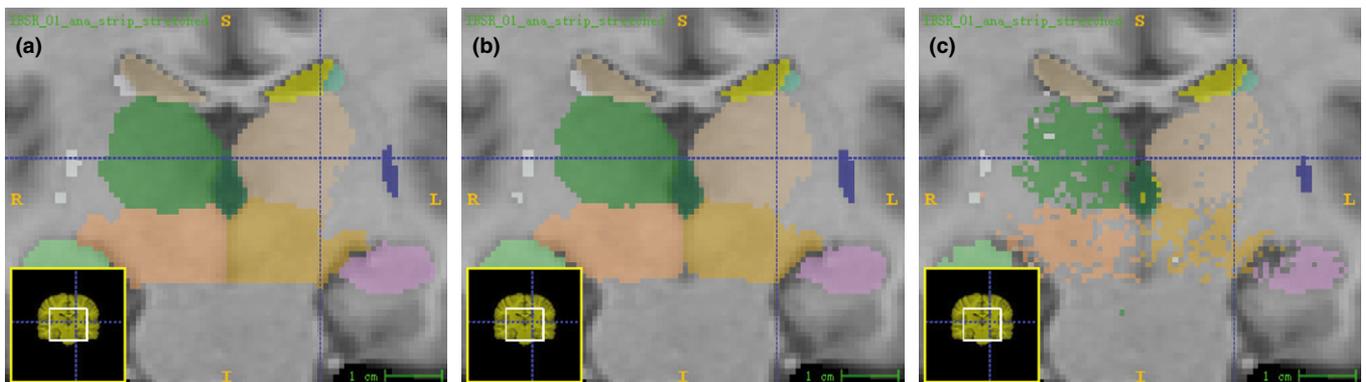


FIG. 3. (a) The manual labeling, (b) the labeling is obtained with the spatial feature, (c) the labeling is obtained without the spatial feature. [Color figure can be viewed at wileyonlinelibrary.com]

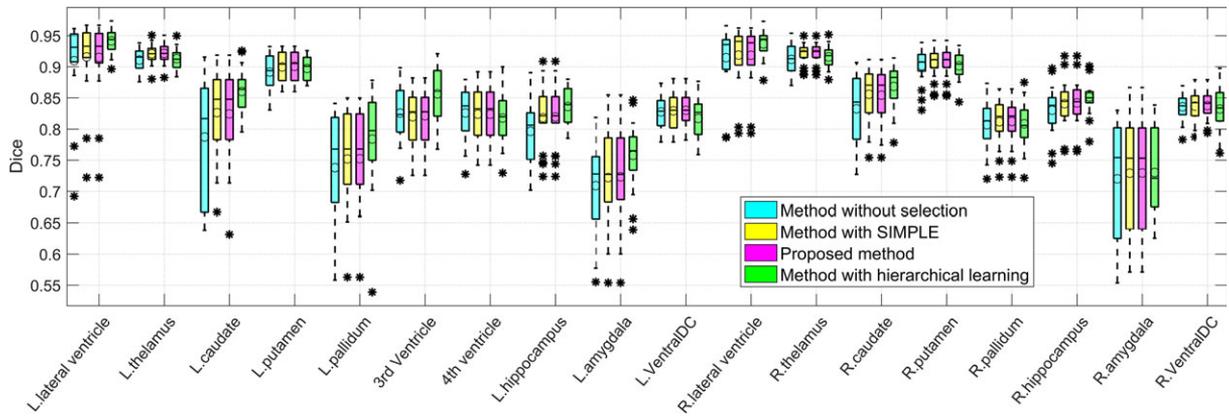


FIG. 4. Comparison of DSC measures obtained by the method without atlas selection proposed by Zikic et al.³⁰ the method with atlas selection by SIMPLE³² (yellow), the proposed method (magenta), and hierarchical learning method proposed by Zhang et al.³¹ (green) in the IBRS dataset. [Color figure can be viewed at wileyonlinelibrary.com]

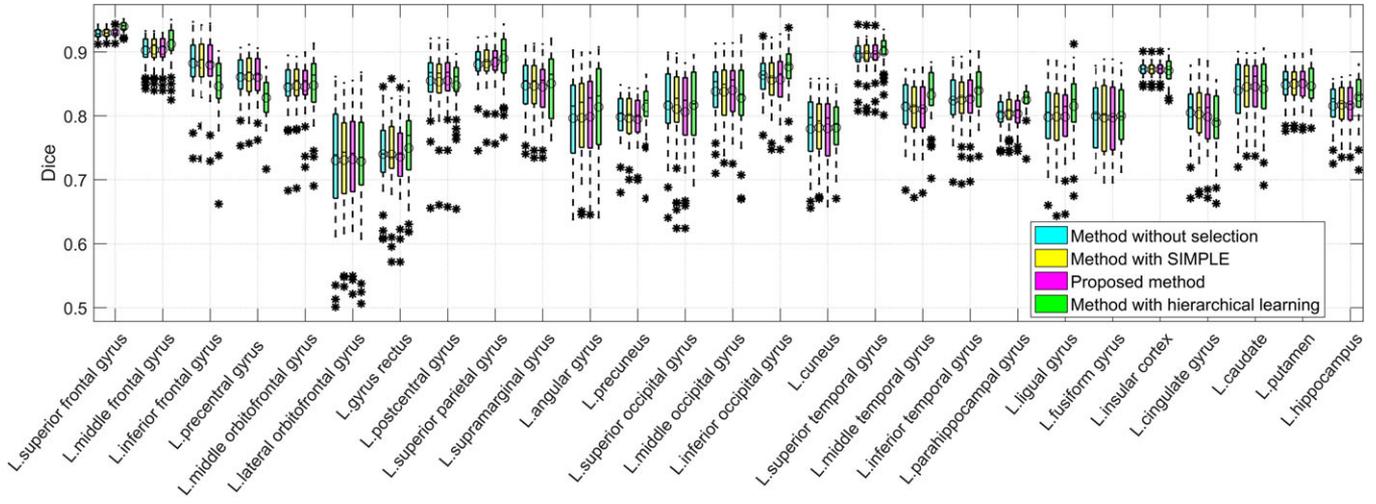


FIG. 5. Comparison of DSC measures obtained by the method without atlas selection proposed by Zikic et al.³⁰ method (blue), the method with atlas selection by SIMPLE³² (yellow), the proposed method (magenta), and hierarchical learning method proposed by Zhang et al.³¹ (green) for the left-hemisphere ROIs in the LONI LPBA40 dataset. [Color figure can be viewed at wileyonlinelibrary.com]

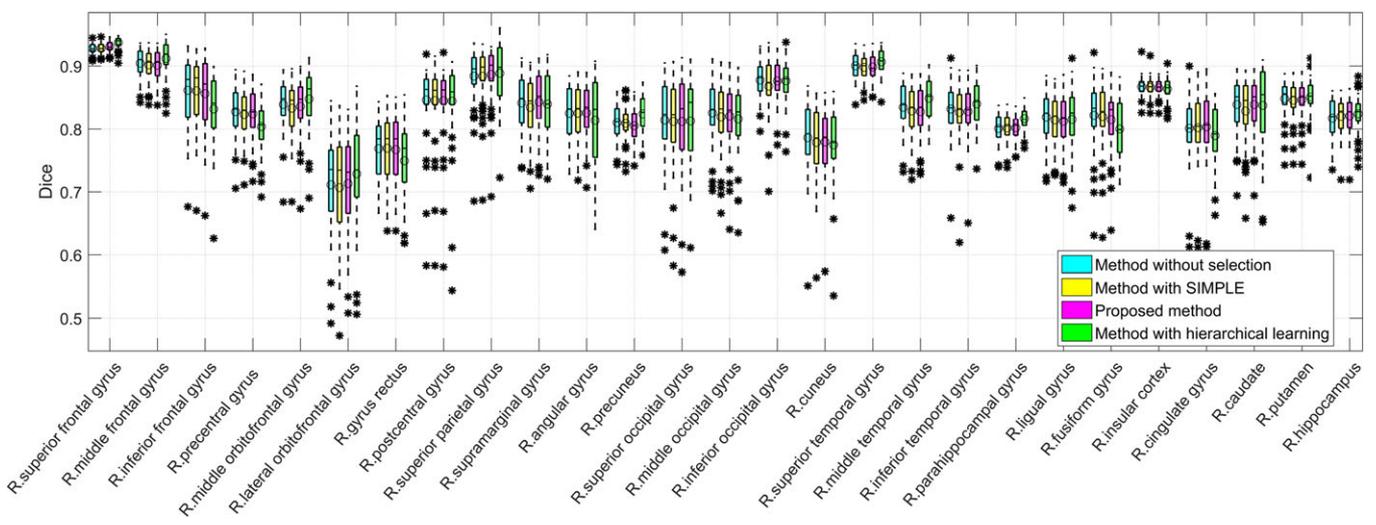


FIG. 6. Comparison of DSC measures obtained by the method without atlas selection proposed by Zikic et al.³⁰ method (blue), the method with atlas selection by SIMPLE³² (yellow), the proposed method (magenta), and hierarchical learning method proposed by Zhang et al.³¹ (green) for the right-hemisphere ROIs in the LONI LPBA40 dataset. [Color figure can be viewed at wileyonlinelibrary.com]

In the proposed method, the encoding time per atlas requires 2 min, in which extracting the features takes 1–2 min and generating the atlas forests takes around 30 s. This part is offline. In the labeling stage, it takes around 8 s to evaluate whether an atlas is suitable for target image and about 40 atlases are tested before labeling the target image. It takes about 5 min 20 s for labeling a target image. This experiment shows that the proposed method is also competent with atlases from different subjects, and the extensible learning model selected suitable atlases for the target image without traversing all the datasets as expected.

4.D. Influence of spatial feature

The spatial feature we proposed has noticeable effect on labeling performance. Figure 3 shows the comparison of labeling with a spatial feature and without, indicating under segmentation will happen in the tissue border and the place that the texture is not clear.

The larger radius would give more spatial information with the increased processing time. When the radius is

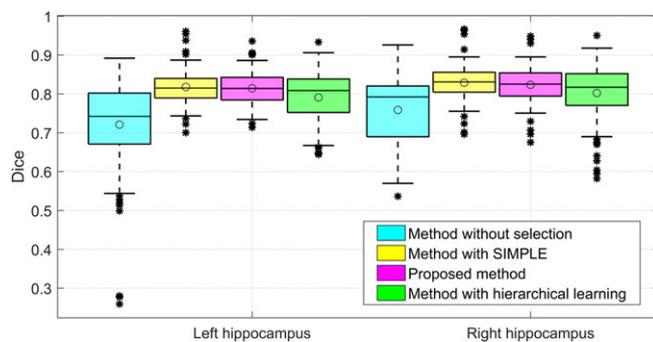


FIG. 7. Comparison of DSC measures obtained by the method without atlas selection proposed by Zikic et al.³⁰ (blue), the method with atlas selection by SIMPLE³² (yellow), the proposed method (magenta), and hierarchical learning method proposed by Zhang et al.³¹ (green) in ADNI dataset. [Color figure can be viewed at wileyonlinelibrary.com]

sufficiently large, the efficiency will not improve with any further increase in radius. When the radius is zero, the method simply relies on a probabilistic atlas and simple location information as Zikic et al.³⁰ described. In such a situation, the DSC measure value is only about 60%. We set the radius of spatial feature $r = 5$ in our experiments to take into account the accuracy and efficiency of the algorithm. Figure SA4 illustrates the effect of the radius for the DSC measures, and Figure SA5 shows the effect of the radius for the computation time.

4.E. Evaluation of extensible learning

We compared our strategy with the method without atlas selection proposed by Zikic et al.³⁰, method using the selective and iterative approach³² and hierarchical learning method proposed by Zhang et al.³¹ to evaluate the performance of extensible learning strategy in detail. All of these four methods utilized spatial features. Figures 4–7 illustrate the DSC measure of four methods in a box plot and Tables SA1 to SA4 show the detail. Tables SA5 and SA6 show the analysis of the time complexity and the extensibility of four methods. Figure 9 shows that the extensible learning model only needs to compare 16, 24, and 40 times in the dataset of IBRS, LONI LPBA40, and ANDI, respectively. For the traditional method, such as SIMPLE method,³² it requires 17, 39, and 99 times. If we assume that a comparison takes 20 s, the proposed method will reduce the computational time with 20 s, 5 min and 30 s, 13 min, respectively, in the three datasets for the completion of the atlas selection.

Experimental results indicate that the accuracy of labeling will be improved after atlas selection and more accurate results in some small tissue can be obtained with the hierarchical learning method. However, the hierarchical learning method requires more computational resources and need to retrain the model when training set changed. The extensible learning process selected atlases for target and achieved excellent performance as the selection strategy that requires testing of all atlases in the dataset.

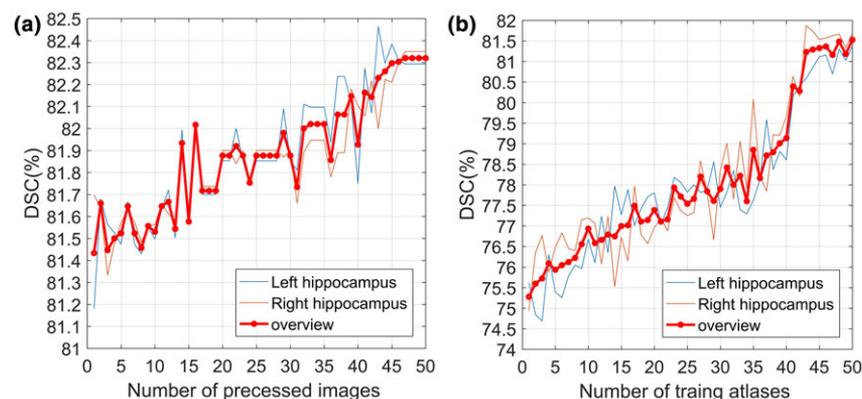


FIG. 8. The dynamic behavior evaluation of the extensible learning model: (a) shows the tendency of DSC values to the number of processed images and (b) shows the tendency of DSC values to the number of training atlases in the training set. [Color figure can be viewed at wileyonlinelibrary.com]

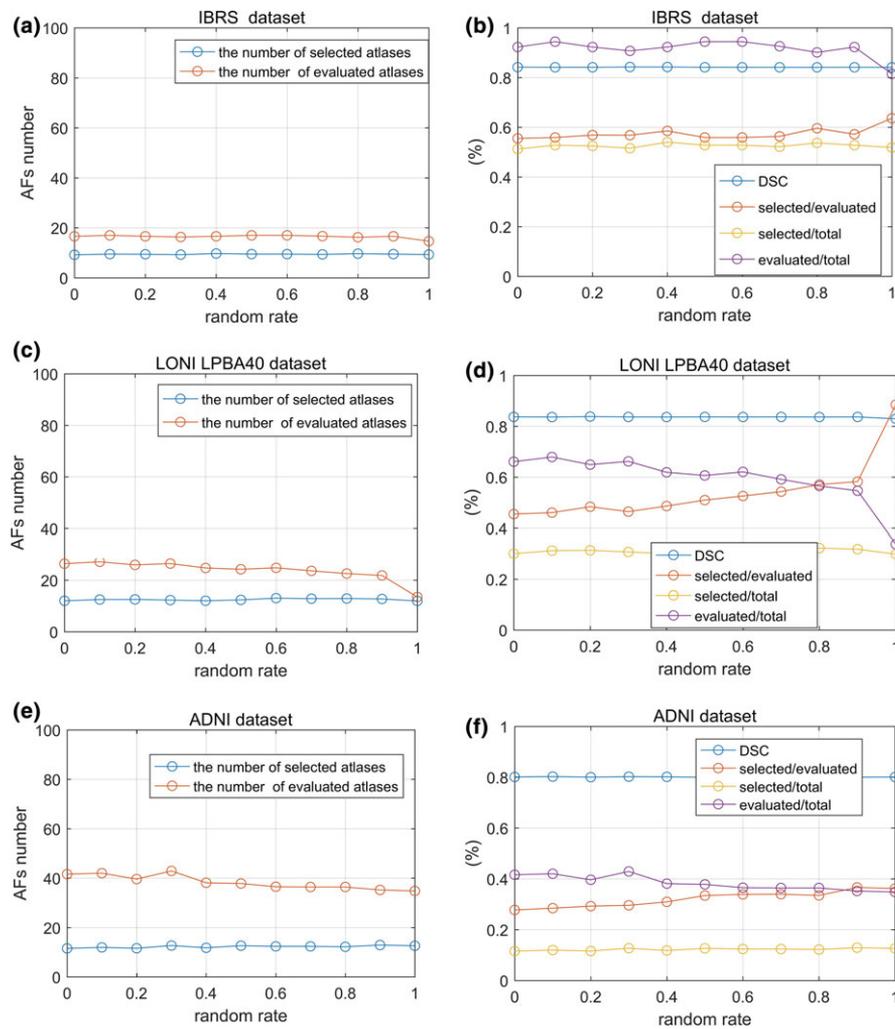


FIG. 9. The effect of random rate. (a), (c), and (e) represent the trend effects of the random rate on the number of selected and test atlases. (b), (d), and (f) represent DSC measure of the proposed method, the ratio of the average number of selected atlases to the total number of the atlas in the dataset, the ratio of the number of evaluated atlases to the total number of the atlas in the dataset, and the ratio of the number of selected atlases to the number of evaluated atlases. [Color figure can be viewed at wileyonlinelibrary.com]

In addition, we evaluate the dynamic behavior of the extensible learning model. The experiment is divided into two parts and performed on ADNI dataset which has 100 atlases. We randomly selected 50 atlases from the dataset as the training set and the rest was used as the test set. In the test set, an image is randomly selected as the observation object, and the DSC value is counted after the model processed the remaining images sequentially. The above process is repeated 10 times and the labeling results of the observation objects are averaged and shown in Fig. 8(a) to illustrate the dynamic behavior of the extensible learning model with the number of processed images. The experimental results show that labeling efficiency of the model is improved with the number of processed images and will become stable after an enough number of images is processed. Then, we randomly selected 50 atlases from the dataset as the training set. In the rest dataset, we randomly select one image as the test object and add another 49 atlases to the training set sequentially. The DSC value of the test image is cumulated at each atlas addition.

The above process is repeated 10 times and the labeling results of the test images are averaged and shown in Fig. 8(b) to demonstrate the dynamic behavior of the extensible model with the number of training atlases. The experimental results that showed in Fig. 8(b) illustrate that the extensible learning model is able to obtain the useful information in the new training atlases so that to improve the efficiency the image labeling.

4.F. Influence of random rate in extensible learning

In this section, we conducted further experiments to evaluate the efficiency of extensible learning and analyze the influence of the random rate. We designed the following experiments. According to different random rates, the statistics is calculated for (a) the average number of atlases selected for each target, (b) the average number of atlases evaluated by the extensible learning model, and (c) the labeling accuracy of the proposed method for brain MR images.

The experimental results in the IBRS, LONI LPBA40, and ADNI datasets are shown in Fig. 9.

The experimental results show that the random rate slightly influences the labeling precision. In the process of labeling, the number of atlases evaluated by the extensible learning model is slightly decreased and the number of the selected target atlases is slightly increased with the increase in the random rate. However, when the random rate is one, the number of the atlas to be estimated by the extensible model is significantly reduced and the number of the selected target atlases is significantly increased. This finding is attributed to the absence of the randomness. When the size of an atlas test set is less than the scale of a cluster, the test set always selects the atlases in the cluster as elements and then the elements will be clustered by themselves. For this reason, the extensible learning model will converge quickly to result in a noticeable reduction in the number of atlas to be evaluated. Simultaneously, if the selected atlases clustered by themselves, we cannot remove the atlases with low similarity to the target image, thus apparently the number of selected atlases will be increased. We add the random rate to avoid the clustering so that the robustness of the proposed method can be improved.

5. CONCLUSION

This paper proposed an extensible learning method of atlas forest to label brain MR images. We imported spatial features into atlas encoding by random forest to obtain excellent labeling without cross-registration. We also designed a self-learning model based on an extensible learning strategy. With this model, the proposed method selects atlases without traversing the entire dataset. Thus, the proposed method can be applied to large datasets.

The proposed method was evaluated in three public datasets: IBRS, LONI LPBA40, and ADNI. We then tested the proposed method in the IBRS and LONI LPBA40 datasets and observed that the proposed method performed well in complex tissues. We subsequently evaluated the proposed method in the LONI LPBA40 and ADNI datasets which contain numerous atlases and found that the proposed method could be employed to select atlases for target image without traversing the entire dataset. Our experiments on these three datasets revealed that the proposed method could manage large-scale or varying atlas datasets under the condition of ensuring the labeling accuracy of MR brain images.

In future studies, our proposed method will be evaluated in other datasets, and the more comprehensive strategy will be utilized to assess our proposed method. The proposed method will also be extended to label other no brain structures, including abdominal organs.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (NSFC) Grants (61370179, and 81671768).

^{a)}Author to whom correspondence should be addressed. Electronic mail: hongliu@hust.edu.cn.

REFERENCES

1. Klein A, Tourville J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci.* 2012;6:171.
2. Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage.* 2006;33:115–126.
3. Beg MF, Miller MI, Trounev A, Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int J Comput Vision.* 2005;61:139–157.
4. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage.* 2009;45:S61–S72.
5. Jia H, Yap PT, Shen D. Iterative multi-atlas-based multi-image segmentation with tree-based registration. *NeuroImage.* 2012;59:422–430.
6. Artaechevarria X, Munoz-Barrutia A, Ortiz-De-Solorzano C. Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans Med Imaging.* 2009;28:1266–1277.
7. Isgum I, Staring M, Rutten A, Prokop M, Viergever MA, van Ginneken B. Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans. *IEEE Trans Med Imaging.* 2009;28:1000–1010.
8. Rohlfing T, Brandt R, Menzel R, Maurer CR Jr. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage.* 2004;21:1428–1442.
9. Sabuncu MR, Yeo BT, Van Leemput K, Fischl B, Golland P. A generative model for image segmentation based on label fusion. *IEEE Trans Med Imaging.* 2010;29:1714–1729.
10. Rousseau F, Habas PA, Studholme C. A supervised patch-based approach for human brain labeling. *IEEE Trans Med Imaging.* 2011;30:1852–1862.
11. Wu G, Wang Q, Zhang D, Nie F, Huang H, Shen D. A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Med Image Anal.* 2014;18:881–890.
12. Wang Z, Wolz R, Tong T, Rueckert D. Spatially aware patch-based segmentation (SAPS): an alternative patch-based segmentation framework presented at the international conference on medical computer vision: recognition techniques and applications in medical imaging; 2012 (unpublished).
13. Bai W, Shi W, Ledig C, Rueckert D. Multi-atlas segmentation with augmented features for cardiac MR images. *Med Image Anal.* 2015;19:98–109.
14. Hao Y, Wang T, Zhang X, et al., Alzheimer's Disease Neuroimaging. Local label learning (LLL) for subcortical structure segmentation: application to hippocampus segmentation. *Hum Brain Mapp.* 2014;35:2674–2697.
15. Cao Y, Yuan Y, Li X, Turkbey B, Choyke PL, Yan P. Segmenting images by combining selected atlases on manifold presented at the international conference on medical image computing & computer-assisted intervention; 2011 (unpublished).
16. Hoang Duc AK, Modat M, Leung KK, et al., Alzheimer's Disease Neuroimaging. Using manifold learning for atlas selection in multi-atlas segmentation. *PLoS ONE.* 2013;8:e70059.
17. Langerak TR, Berendsen FF, Ua VDH, Kotte AN, Pluim JP. Multiatlas-based segmentation with preregistration atlas selection. *Med Phys.* 2013;40:091701–091701.
18. Asman AJ, Landman BA. Non-local STAPLE: an intensity-driven multi-atlas rater model presented at the medical image computing & computer-assisted intervention: MICCAI international conference on medical image computing & computer-assisted intervention; 2012 (unpublished).
19. Breiman L. Random forests. *Mach Learn.* 2001;45:5–32.
20. Shepherd BA. An appraisal of a decision tree approach to image classification presented at the eighth international joint conference on artificial intelligence; 1983 (unpublished).
21. Quinlan JR. Induction of decision trees. *Mach Learn.* 1986;1:81–106.

22. Zhang L, Wang Q, Gao Y, Li H, Wu G, Shen D. Concatenated spatially-localized random forests for hippocampus labeling in adult and infant MR brain images. *Neurocomputing*. 2016;229:3–12.
23. Mesejo P, Ugolotti R, Cagnoni S, Cunto FD. Automatic segmentation of hippocampus in histological images of mouse brains using deformable models and random forest. *Proc IEEE Symp Comput Based Med Syst*. 2012;1–4.
24. Maiora J, Ayerdi B, Graña M. Random forest active learning for AAA thrombus segmentation in computed tomography angiography images. *Neurocomputing*. 2014;126:71–77.
25. Maglietta R, Amoroso N, Bruno S, et al., Random forest classification for hippocampal segmentation in 3D MR images presented at the international conference on machine learning and applications; 2013 (unpublished).
26. Criminisi A, Shotton J, Konukoglu E. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Found Trends Comput Graph Vis*. 2011;7:81–227.
27. Chatelain P, Pauly O, Peter L, et al. *Learning from Multiple Experts with Random Forests: Application to the Segmentation of the Midbrain in 3D Ultrasound*. Berlin, Heidelberg: Springer; 2013.
28. Zikic D, Glocker B, Konukoglu E, et al. *Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-Channel MR*. Berlin, Heidelberg: Springer; 2012.
29. Zikic D, Glocker B, Criminisi A, Atlas encoding by randomized forests for efficient label propagation presented at the medical image computing & computer-assisted intervention: MICCAI international conference on medical image computing & computer-assisted intervention; 2013 (unpublished).
30. Zikic D, Glocker B, Criminisi A. Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. *Med Image Anal*. 2014;18:1262–1273.
31. Zhang L, Wang Q, Gao Y, Wu G, Shen D. Automatic labeling of MR brain images by hierarchical learning of atlas forests. *Med Phys*. 2016;43:1175–1186.
32. Langerak TR, van der Heide UA, Kotte AN, Viergever MA, van Vulpen M, Pluim JP. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Trans Med Imaging*. 2010;29:2000–2008.
33. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29:196–205.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Figure SA1. (a–d) the manual labeling in IBRS dataset, (e–h) the labeling is obtained by the proposed method in the IBRS dataset.

Figure SA2. (a–e) The manual labeling in LONI LPBA40 dataset, (f–j) The labeling is obtained by the proposed method in the LONI LPBA40 dataset.

Figure SA3. (a) The manual labeling for NC subjects, (b) the labeling of the proposed method for NC subjects, (c) the manual labeling for MCI subjects, (d) the labeling of the proposed method for MCI subjects, (e) the manual labeling for AD subjects, and (f) the labeling of the proposed method for AD subjects.

Figure SA4. The effect of the radius of the spatial feature for DSC measure in IBRS dataset.

Figure SA5. The effect of the spatial feature radius for the computation time in IBRS dataset.

Table SA1. Quantitative compaction of DSC values obtained by the method without atlas selection proposed by Zikic et al.¹, the method with atlas selection by SIMPLE,² the hierarchical learning method proposed by Zhang et al.³, and the proposed method for the 18-labeled primarily subcortical regions in the IBRS dataset.

Table SA2. Quantitative compaction of DSC values obtained by the method without atlas selection proposed by Zikic et al.¹, the method with atlas selection by SIMPLE,² the hierarchical learning method proposed by Zhang et al.³, and the proposed method for the left-hemisphere ROIs in the LONI LPBA40.

Table SA3. Quantitative compaction of DSC values obtained by the method without atlas selection proposed by Zikic et al.¹, the method with atlas selection by SIMPLE,² the hierarchical learning method proposed by Zhang et al.³, and the proposed method for the right-hemisphere ROIs in the LONI LPBA40.

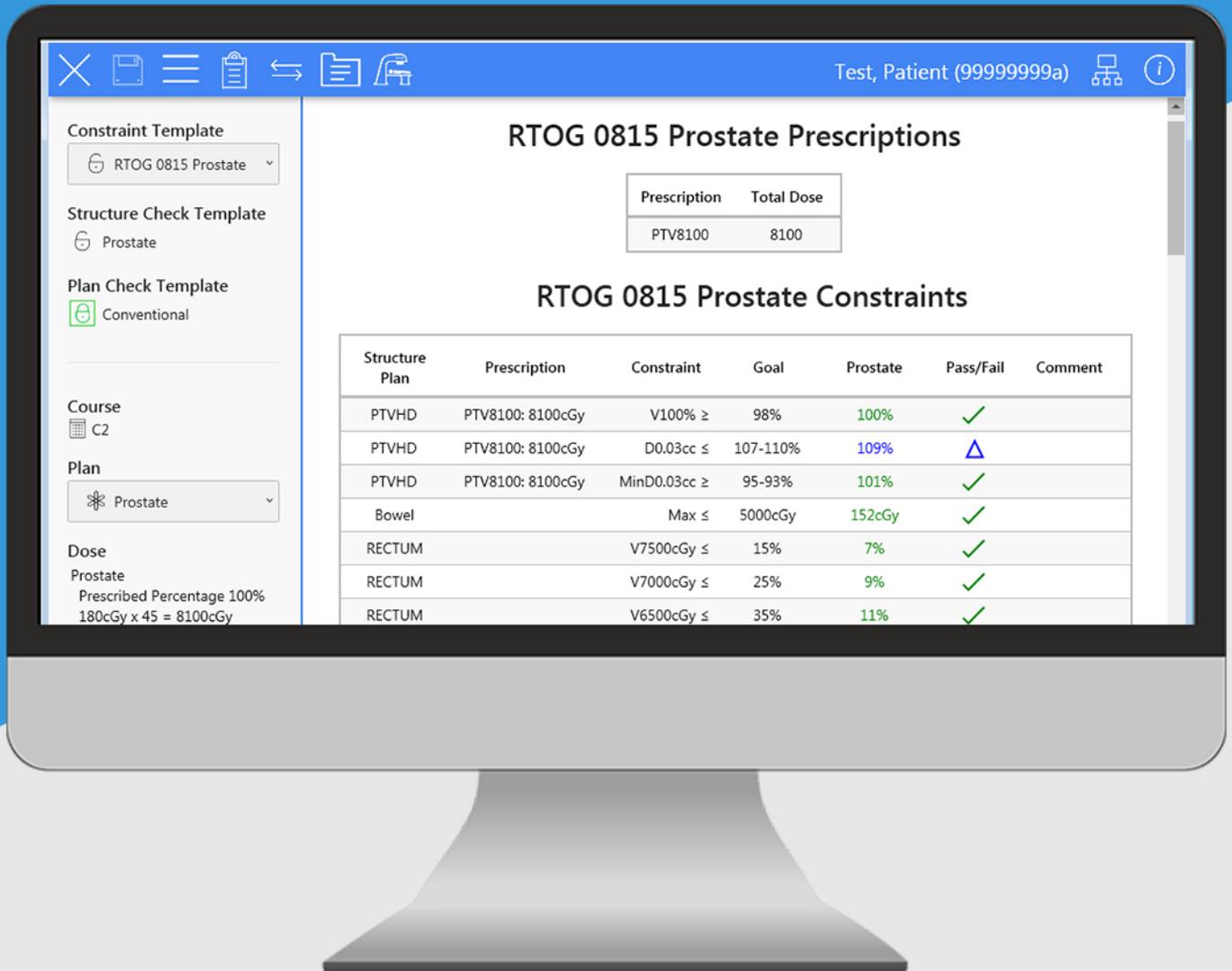
Table SA4. Quantitative compaction of DSC values obtained by the method without atlas selection proposed by Zikic et al.¹, the method with atlas selection by SIMPLE,² and the hierarchical learning method proposed by Zhang et al.³, and the proposed method for the labeled left and right hippocampus in the ADNI dataset.

Table SA5. Comparison of Time Complexity of the method without atlas selection proposed by Zikic et al.¹, the method with atlas selection by SIMPLE,² the method with hierarchical learning proposed by Zhang et al.³ method, and the proposed method.

Table SA6. The extensibility of the method without atlas selection proposed by Zikic et al.¹, the method with atlas selection by SIMPLE,² the method with hierarchical learning proposed by Zhang et al.³ method, and the proposed method.

ClearCheck

One-click plan evaluation



Dose Constraints • Plan Checks
Structure Checks • Collision Checks