J Med Syst (2019) 43: 241
https://doi.org/10.1007/s10916-019-1385-3

IMAGE & SIGNAL PROCESSING

Automatic Labeling of MR Brain Images Through the Hashing Retrieval Based Atlas Forest

Hong Liu[1] · Lijun Xu[1] · Enmin Song[1] · Renchao Jin[1] · Chih-Cheng Hung[2]

Received: 10 March 2019 / Accepted: 10 June 2019 / Published online: 21 June 2019

Abstract

The multi-atlas method is one of the efficient and common automatic labeling method, which uses the prior information provided by expert-labeled images to guide the labeling of the target. However, most multi-atlas-based methods depend on the registration that may not give the correct information during the label propagation. To address the issue, we designed a new automatic labeling method through the hashing retrieval based atlas forest. The proposed method propagates labels without registration to reduce the errors, and constructs a target-oriented learning model to integrate information among the atlases. This method innovates a coarse classification strategy to preprocess the dataset, which retains the integrity of dataset and reduces computing time. Furthermore, the method considers each voxel in the atlas as a sample and encodes these samples with hashing for the fast sample retrieval. In the stage of labeling, the method selects suitable samples through hashing learning and trains atlas forests by integrating the information from the dataset. Then, the trained model is used to predict the labels of the target. Experimental results on two datasets illustrated that the proposed method is promising in the automatic labeling of MR brain images.

Keywords MR brain · Multi-atlas-based method · Label fusion · Locality sensitive hashing · Random forest

Introduction

Labeling of MR brain image plays a crucial role in the research of complex relationships between structure and function of human brain. It also provides useful information for treatment planning [1, 2], clinic assessment [3], and diagnosis of brain disorders [4, 5]. Expert-labeled images are always considered as the gold standards, but the manual labeling is inter-rater variable and inefficient [6].

Some fully automatic methods based on intensity images [7, 8] can generate accuracy results. However, these methods are inadequate at robustness due to the low image quality and tissue morphological variation. The atlas-based methods may solve these problems, which uses the prior information provided by expert-labeled images to guide the labeling of target [9, 10]. The multi-atlas-based methods are advanced to avoid random errors caused by the MR scanners or the manual labeling. Multi-atlas-based methods usually consist two major steps: register atlases to the target image space thus the labels propagate to target directly [11, 12], and fuse propagated labels through various label fusion strategies to acquire the results [13–17].

Label fusion can reduce the errors caused by registration and low quality of the images. The most typical label fusion methods are the majority voting and its variant forms such as locally [18, 19], globally [20, 21] and semi-local voting [22]. Another way to calculate the weights in label fusion is machine learning, such as the label fusion method based on sparse representation [23], the label fusion method based on k-nearest neighbor classification [24]. Some other methods fused the propagated labels directly using classification such as local label learning (LLL) method [25], the method with support vector machine(SVM) [26, 27], and

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Enmin Song
enmin_song@126.com

[1] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

[2] Kennesaw State University, Marietta, GA, USA

hierarchical learning based atlas forest(HF) [28]. Although these methods evaluated and applied the propagated labels carefully, they depend on the accuracy of registration which is time-consuming.

Some researchers use the prior information without the registration. Zikic et al. [12] encoded the atlas using random forest and considered each atlas as a classifier to predict the target labels. This method produced results independent of the registration, however, it is poor in the accuracy. The methods based on the deep learning take advantage of atlases directly by replacing the registration and label fusion with the neural network [29, 30]. However, the deep learning based methods always encounter the insufficient of samples that may cause overfitting of the trained model, and training a network is very resources demand and time consuming. In addition, Huo et al. [31] proposed a supervoxel based method that groups pixels with the same attribute into supervoxel, then label the image using these supervoxels. This method addresses the problem encountered with deep learning methods, however, it ignores the relationship between the atlases and 3-D morphological features in the medical images.

To solve these problems, we proposed an automatic labeling method through hashing retrieval based hybrid atlas forest, which is abbreviated as the Hash-Atlas Forest(Hash-AF). This new method can avoid the errors produced by registration and integrate the information among atlases in dataset. The main innovations are as follows.

(1) A strategy of coarse classification is proposed to preprocess the dataset, which classifies the data based on the spatial of tissues, to preserves the correlation among the data and reduces the computing time. Furthermore, this strategy provides a way to deal with the labeling in ambiguous area, thus improving the accuracy when labeling the multiple tissues.
(2) A sample retrieval method based on locally sensitive hashing (LSH) is proposed, which retrieves the samples quickly and eliminates irrelevant ones in the training set according to the similarity between samples and target.
(3) The proposed method encodes voxels in the atlases into Hash-AF that integrates the appearance information of images, the correlations between atlases and the priori spatial information of brain tissues. This strategy allows the atlases propagates their labels directly rather than register into target image space and fuse the labels, thus avoiding the errors caused by label propagation and reducing the computing time spent on cross-registration.

Method

The proposed method considers each voxel in the atlas as an independent sample, and hashes the samples for fast data retrieval. Then, this method trains the retrieved samples into the Hash-AF and predicts results. The overview of our proposed method is shown in Fig. 1, and there are four main steps described below.

Coarse classification

A strategy of coarse classification is proposed to pre-process the dataset, which separates images into several parts based on the location of target tissues and discards the irrelevant parts. In initiation, an atlas,which is selected arbitrary as the reference, is registered to the atlas image space using the non-rigid registration to provide the location information of the brain tissues. Then, the images are classified into several coarse classifications according to the mapped location information. Each coarse classification includes a target tissue and the areas around the tissue as illustrated in Fig. 2.

Feature extraction

The features are consisted of two parts to describe the samples by its appearance and spatial information. Let the notation $I(x)$ indicates the intensity at a point x, $N_s(x)$ represents the points in the cubic with side s centered on x. The symbol v is a vector, and we defined $x + v = (x + v, y + v, z + v)$ in the 3-D space. The notation μ is the operation to average. Then, the features of local intensity and texture in the point x is calculated by Eq. 1, where s, u are the parameters to control the size of image patch centered on x. The symbol λ is a flag, when $\lambda = 0$, the $F^{local}_{s,u}(x)$ represents the local intensity features and when $\lambda = 1$ the $F^{local}_{s,u}(x)$ represents the local texture features.

$$F^{local}_{s,u}(x) = \mu(I(N_s(x + \lambda \cdot u))) - \lambda \cdot I(x), \lambda \in \{0, 1\} \quad (1)$$

Given that the notation $L'_{ref}(x)$ is the label provided by the label image mapped from the reference atlas to the image space of x and the symbol $R_r(x)$ indicates the points in the sphere centered on x with radius r. The features of spatial information are calculated by Eq. 2, where the radius $r = 5$ according to the experiments.

$$F^{spatial}_{r}(x) = L'_{ref}(R_r(x)) \quad (2)$$

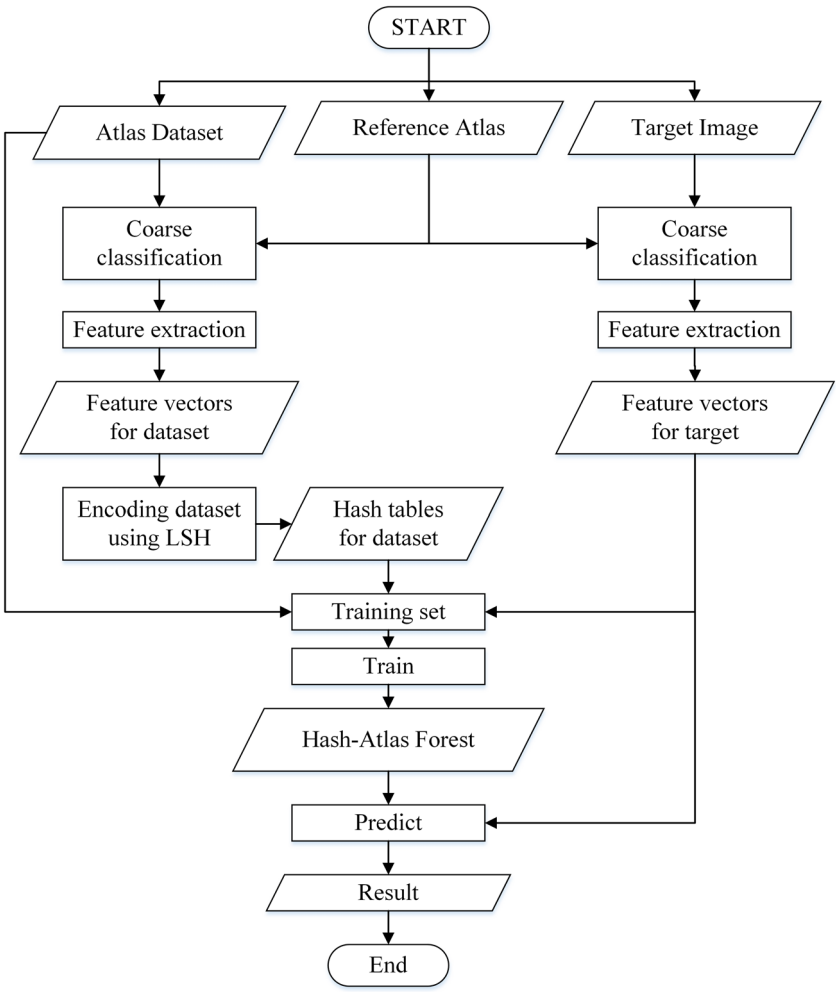

Fig. 1 The overview of the proposed method. 1) Register the reference atlas to the image space of the target using non-rigid registration for providing the spatial information, and generate the coarse classifications for each target tissue. 2) Construct the hash tables by LSH. 3) Retrieve training samples and train Hash-Hybrid forest. 4) Predict probability maps for each tissues and obtain the labeling of target

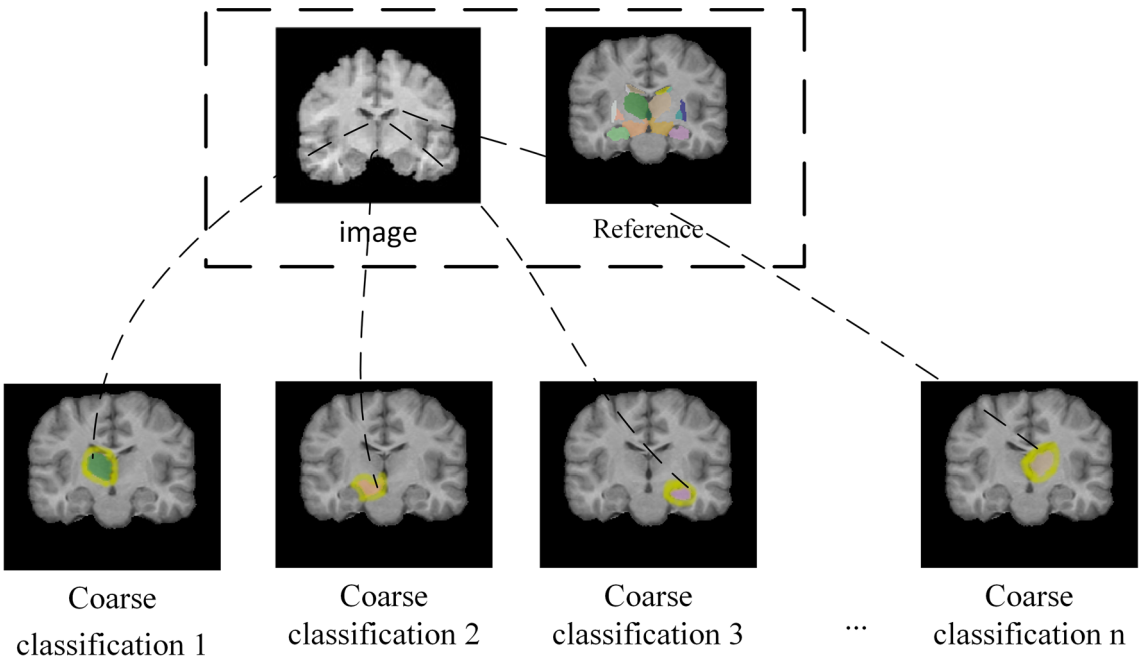

Fig. 2 The illustration of coarse classification. The region where has the same color in the reference label image is the target tissue, and the region in yellow is the surrounding areas of the tissue whose Euclidean distance from the boundary of target tissues is less than r

Data retrieval

The theory of LSH

Given D is a p-stable distribution, which means if there exists $p \geq 0$, for arbitrarily n real numbers $v_1, v_2, ..., v_n$ and n variables $X_1, X_2, ..., X_n$ with identical distribution D, the random variable $\sum_i v_i X_i$ and $(\sum_i |v_i|^p)^{1/p} X$ has the same distribution. According to the property of p-stable distribution, it can estimate the length of a vector v in the European space with the norm p, namely $\|v\|_p$. The LSH method takes the advantage of this property and designs its hash function family as Eq. 3, where $V = (v_1, v_2, ..., v_n)$ is a vector of n real numbers and $a = (X_1, X_2, ..., X_n)$ is a vector that consists of random variables with p-stable distribution. The symbol $b \in (0, r)$ is a random variable, r is the length of a line segment.

$$h_{a,b}(V) = \left\lfloor \frac{a \cdot V + b}{r} \right\rfloor \tag{3}$$

When $a \cdot V_1 + b$ and $a \cdot V_2 + b$ project to the same line segment means two samples V_1, V_2 project to the same hash code or collision. The probability of V_1, V_2 projecting to the same line segment is calculated by Eq. 4, where $c = \|V_1 - V_2\|_2$. We apply 2-stable distribution when encoding the dataset so that f is the probability density function as Eq. 5. For a given r, when c is smaller, the probability of collision is high. Otherwise, the probability of collision is low. The hash family function is the more sensitive to local differences with smaller r, but it increases the search time when r is too small.

$$\begin{aligned} p(c) &= \Pr\left[h_{a,b}(V_1) = h_{a,v}(V_2)\right] \\ &= \int_0^r \frac{1}{c} f\left(\frac{t}{c}\right)\left(1 - \frac{t}{r}\right) dt \end{aligned} \tag{4}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{5}$$

Encoding the dataset using LSH

The dataset are encoded into hash tables through the theory of LSH for each coarse classification and Fig. 3 shows the structure of hash tables.

Given that the number of target tissue is m, the corresponding coarse classification is $C = \{C_i | i = 1, 2, ..., m\}$. The notation $H = \{H_i | i = 1, 2, ..., m\}$ means the hash tables that the dataset are encoded into, and $G = \{G_i | i = 1, 2, ..., m\}$ means the parameters set to generate hash tables. The detail of the algorithm for encoding the dataset using LSH in Hash-AF is described as Algorithm 1:

Algorithm 1 Encoding the dataset using LSH.

Input: $C = \{C_i | i = 1, 2, ..., m\}$;
Output: $H = \{H_i | i = 1, 2, ..., m\}$, $G = \{G_i | i = 1, 2, ..., m\}$;

```
for (i = 1; i ≤ m; i++) do
    // step1: generate parameter set
    initiate H_i, G_i;
    for (j = 1; j ≤ tablesize; j++) do
        //tablesize is the number of hash function group in a parameter set G_j
        initiate g_j
        for (k = 1; k ≤ K; k++) do
            //K is the number of hash function in hash function group g_j
            select h_k randomly from the LSH hash family;
        end for
        g_j = {h_1, h_2, ..., h_K};
    end for
    G_i = {g_1, g_2, ..., g_tablesize};
    //step2: generate hash table
    j = 1
    for each x ∈ C_i do
        code_j = [g_1(f_x), g_2(f_x), ..., g_tablesize(f_x)];
        j = j + 1;
    end for
    H_i = {code_1, code_2, ..., code_j};
end for
return G = {G_1, G_2, ..., G_m}, H = {H_1, H_2, ..., H_m}
```

Sample retrieval through the hash tables

Each voxel in the target image selects the similarity samples from the dataset through the constructed hash table H, then the selected samples are the training set for the learning model. Given the symbol $t \in T$ means a voxel in the target image T with the feature vector f_t,which hash codes are calculated by parameters set G. The similar samples consist the voxels from dataset whose hash code is identical to the hash code of voxel from target. Supposed that $SIM(code_t, code_x)$ is the similarity between the target t and sample x, the symbol $\oplus$ means bitwise XOR and the $Negate()$ flips the value by bit. The notation $count()$ is a function to count the number of digits larger than 1, the similarity between the target and sample is calculated by Eq. 6. If $SIM(code_t, code_x) > 0$, the voxel x is a similarity sample to the target.

$$\begin{aligned} &SIM(code_t, code_x) \\ &= \sum_{i=1}^{n} count(Negate(code_t(i) \oplus code_x(i)) \end{aligned} \tag{6}$$

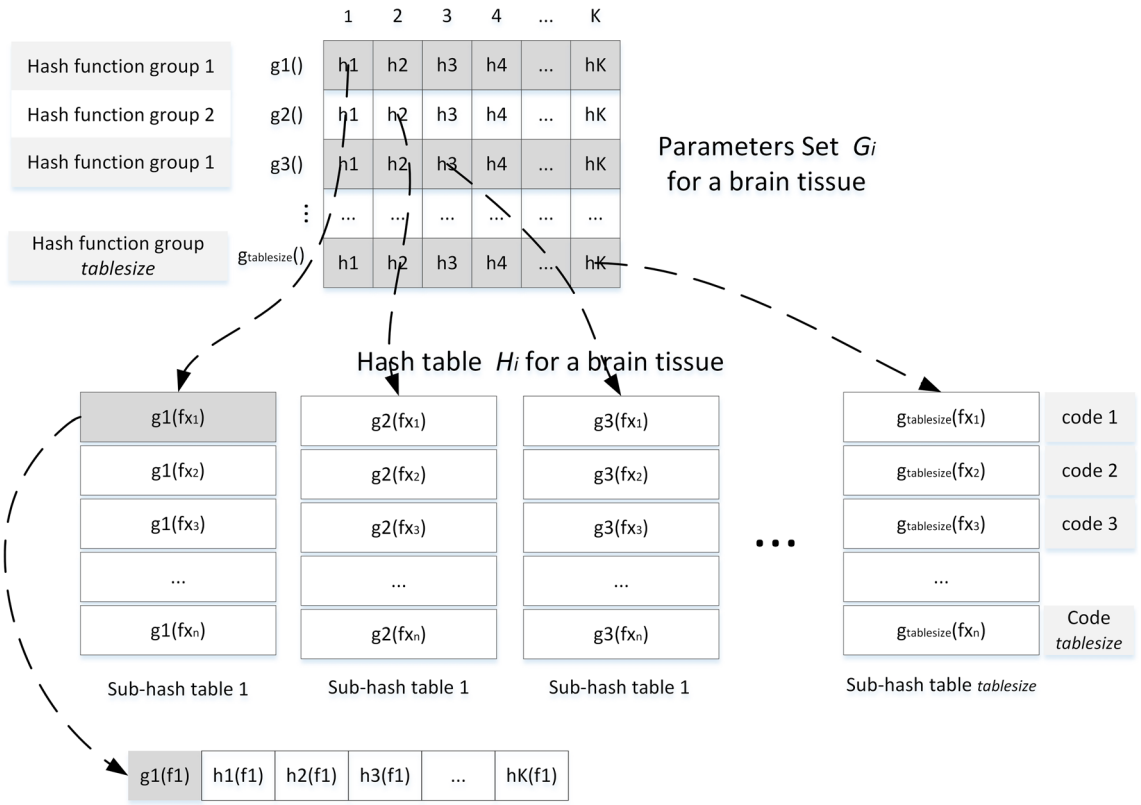

Fig. 3 The illustration of the dataset encoding with LSH for a coarse classification

To reduce the redundant, we sort the similarity by the value of SIM, and choose the samples whose value of SIM are the top k as the similar samples. The detail of the algorithm for sample retrieval through the hash table is shown in Algorithm 2.

Algorithm 2 Sample retrieval through the hash table.

Input: T, $H = \{H_1, H_2, ..., H_m\}$, $G = \{G_1, G_2, ..., G_m\}$;
Output: $S = \{S_1, S_2, ..., S_m\}$;

segment T into several coarse classifications;
c_t is the coarse classification of voxel $t \in T$;
initiate S;
for each $t \in T$ **do**
  $code_t = G_{c_t}(f_t)$;
  // f_t is the feature vector of t
  $i = 1$;
  for each $x \in C_{c_t}$ **do**
    $code_i = H_{c_t}(i)$;
    $similar = SIM(code_t, code_x)$;
    $i = i + 1$;
  end for
  sort $similar$ descend;
  st is a samples set whose $similar$ in top k;
  $S_{t_i} = S_{t_i} + st$;
end for

Hash-Atlas forest

A Hash-Atlas Forest is trained by random forest on a training set that is searched via hashing method proposed in "The theory of LSH". Each coarse classification has its corresponding Hash-Atlas Forest to integrate the intensity and prior information of the dataset.

We use the ID3 strategy [32] to build the decision trees in forest, and the termination condition for the growth of a tree is that the leaf nodes are less than eight. For a voxel x with the feature f_x, the probability predicted by a decision tree is $p_i(c|f_x)$, where c is the potential label of the sample x. Supposed that there are n_t trees in a random forest F, and the Eq. 7 calculates the predicted label of the voxel.

$$P_F(c|f_x) = 1/n_t \sum_{i=1}^{n} p_i(c|f_x) \tag{7}$$

The final prediction of the voxcel x is the label which has the max probability as Eq. 8

$$\hat{c} = \arg\max_c P_F(c|f_x) \tag{8}$$

Figure 4 shows an example of how to choose the Hash-Atlas Forests to predict labels. Given m is the number of target tissues, the symbols $c_1, c_2, ..., c_m$ are the label of coarse classifications. When the label is predicted by

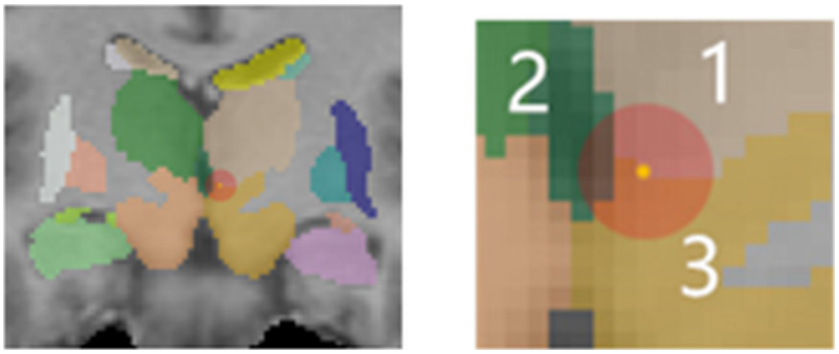

Fig. 4 An illustration when a target voxel belongs to multiple coarse classifications. The label of target voxel (the yellow point) is predicted by the multiple Hash-Atlas Forest whose coarse classifications are the label of region(1,2,3) where the surrounding area covered(the red area)

multiple Hash-Atlas Forests $F_1, F_2, ..., F_m$, the result are depended on the max probability calculated as Eq. 9.

$$\hat{c} = \begin{cases} \arg\max_{c_i} p_{F_i}(c_i|f_x) & if \quad max\, p_{F_i}(c_i|f_x) \geq 0.5 \\ 0 & else \end{cases} \tag{9}$$

Experiments results

The proposed method are evaluated on two public datasets (IBSR and ADNI), and then the influence of primary parameters are discussed.

In prepossessing, we rescaled the intensity of image between 0 and 255 using an ITK-based[1] histogram-matching program. Then a reference atlas is selected to map its label image to the image space of atlases and target in order to provides spatial information.

The initialization parameters are set as follows and all experiments use them.

The parameter of coarse classification $r = 5$. The parameters of feature extraction $s, u, r \in \{\pm1, \pm2, \pm3, \pm4, \pm5\}$. The number of hash function group in a hash table is $tablesize = 20$, and the chosen top $k = 21$. The number of decision tree in an atlas forest is $n_t = 50$.

We use the leave-one-out cross-validation approach to evaluated the propose method. The Dice similarity coefficient (DSC) value and the Maximum symmetric surface distance (MSD) are the metrics, which are calculated in Eqs. 10–11, where v_A is the result of proposed method, v_B is the manual segmentation and notation $B(\cdot)$ means the boundary of segmentation.

$$Dice(v_A, v_B) = 2 \times \frac{|v_A \bigcap v_B|}{|v_A| + |v_B|} \tag{10}$$

$$MSD(v_A, v_B) = \max \left\{ \max_{a \in B(v_A)} \min_{b \in B(v_B)} \|a-b\|, \max_{b \in B(v_B)} \min_{a \in B(v_A))} \|a-b\| \right\} \tag{11}$$

Evaluation on IBSR dataset

IBSR dataset[2] has 18 atlases, each atlas consists of a T1 MR Brain image and corresponding manually guided expert segmentation results of 32 tissues. The Experiments labeled 18 primary subcortical brain tissues, and Fig. 5 shows the manual labels and the results produced by proposed method.

The proposed *Hash-AF* is compared with four existing automatic labeling methods including:

1) *MV*(the majority voting method).
2) *NPBL*(the Nonlocal patch-based label fusion method) [33] which is a derivative of the MV. It calculates the weights of propagated labels using the Euclidian distance between the atlas image and target in a fixed size patches.
3) *AF*(the random forest encoded atlas method) [12] which fuses the labels using atlas forest without cross-registration.
4) *Hierarchical-AF*(the hierarchical atlas forest method) [28] which integrates the information by learning the atlas forest hierarchy.

   Two comparison methods are designed to evaluate the efficiency of each component of the proposed.
5) *Hash-MV*, which retrieves samples for target through the proposed LSH-based data retrieval algorithm and fuses the propagated labels of these chosen samples as MV do.
6) *Coarse-AF*, which trains the atlas forest for each coarse classification without data retrieval.

In the experiments, all the atlas-forest-based methods including AF, Hierarchical-AF, Coarse-AF and our proposed Hash-AF which use the same feature vectors for training and each of their forests has 50 decision trees. The size of patches used in the algorithms of MV, NPBL and Hash-MV is $5 \times 5 \times 5$. MV and AF are two categories in labeling the brain tissues. They are chosen in our study as the baseline that represents two different types of

[1] https://itk.org/ITK/resources/software.html

[2] http://www.nitrc.org/projects/ibsr

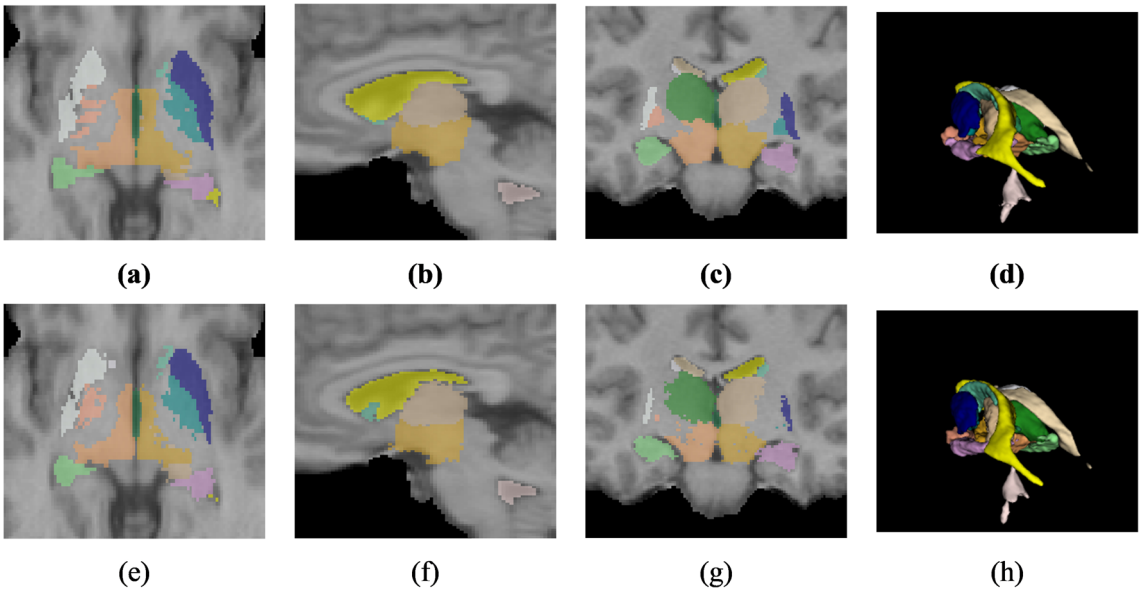

Fig. 5 **a**–**d** the manual labeling of 18 primarily subcortical regions (as shown in Table 1) in the IBSR dataset and **e**–**h** the corresponding labeling obtained by our proposed method

algorithms which use the weight calculation and atlas forest. The comparison among the MV, NPBL and Hash-MV intends to show the efficiency of proposed sample retrieval. Algorithms of AF, Hierarchical-AF and Coarse-AF are selected to verify the improvement of our proposed method among the atlas-forest-based methods.

Table 1 A quantitative DSC (%) obtained by six comparative methods and the proposed method for 18-labeled primarily sub-cortical regions in the IBSR dataset

Brain regions	MV	NPBL	AF	Hierarchical-AF	Hash-MV	Coarse-AF	Hash-AF
L.lateral ventricle	88.16 ± 5.10	90.18 ± 4.01	90.97 ± 6.96	**94.14 ± 2.00**	90.94 ± 3.42	93.76 ± 2.99	94.07 ± 3.05
L.thelamus	88.82 ± 2.10	89.60 ± 1.35	91.07 ± 1.97	91.29 ± 1.79	89.56 ± 1.09	**92.37 ± 1.34**	92.12 ± 0.93[a,b]
L.caudate	82.98 ± 4.28	86.61 ± 2.52	78.71 ± 10.11	86.02 ± 3.71	86.61 ± 4.29	86.98 ± 3.52	88.45 ± 3.31[a,b]
L.putamen	88.35 ± 2.55	88.11 ± 2.55	89.14 ± 2.90	89.72 ± 1.84	89.22 ± 1.77	90.38 ± 2.08	**90.63 ± 1.93**[a,b]
L.pallidum	80.60 ± 4.17	79.51 ± 4.66	73.79 ± 9.07	78.38 ± 7.79	80.16 ± 5.24	80.73 ± 4.79	80.73 ± 5.83[a,b]
3rd Ventricle	75.55 ± 7.37	76.42 ± 5.70	82.56 ± 4.68	85.64 ± 4.43	80.68 ± 4.40	86.06 ± 4.44	**86.54 ± 3.78**
4th ventricle	81.47 ± 3.90	80.88 ± 5.11	82.57 ± 4.23	81.75 ± 4.27	84.56 ± 4.12	80.69 ± 4.67	**85.09 ± 3.80**[a,b]
L.hippocampus	79.38 ± 3.70	80.67 ± 2.88	79.63 ± 5.37	83.61 ± 3.12	81.79 ± 2.62	83.23 ± 3.83	**85.23 ± 2.72**[a,b]
L.amygdala	74.88 ± 6.60	74.14 ± 4.34	70.89 ± 7.46	75.77 ± 5.61	75.92 ± 3.28	74.80 ± 6.15	**77.19 ± 4.27**[b]
L.VentralDC	84.03 ± 2.30	82.91 ± 3.02	82.75 ± 2.52	81.69 ± 3.43	82.80 ± 2.52	83.27 ± 2.77	**84.22 ± 2.40**[a,b]
R.lateral ventricle	87.41 ± 4.65	89.14 ± 4.66	91.45 ± 5.13	**93.65 ± 2.36**	89.67 ± 4.55	92.79 ± 3.65	93.31 ± 3.28
R.thelamus	88.70 ± 2.09	89.52 ± 1.70	91.18 ± 2.39	91.57 ± 1.87	89.79 ± 1.08	92.18 ± 1.53	**92.50 ± 1.21**[a,b]
R.caudate	83.25 ± 3.93	86.32 ± 3.41	83.24 ± 5.69	86.83 ± 3.72	85.31 ± 5.12	86.27 ± 3.85	**87.07 ± 4.26**
R.putamen	88.33 ± 2.33	88.23 ± 2.18	90.15 2.96	90.14 ± 2.15	88.92 ± 1.74	90.23 ± 2.42	**90.63 ± 1.89**
R.pallidum	81.45 ± 4.53	79.90 ± 4.08	80.68 ± 4.01	80.83 ± 3.76	78.61 ± 6.23	81.26 ± 3.59	**81.50 ± 3.79**
R.hippocampus	80.00 ± 3.92	81.23 ± 3.04	83.01 ± 3.95	**85.13 ± 2.92**	81.40 ± 2.37	84.33 ± 3.69	**85.04 ± 2.84**
R.amygdala	72.61 ± 8.43	72.11 ± 6.31	71.99 ± 9.19	73.01 ± 7.26	73.90 ± 4.54	72.98 ± 8.21	**74.42 ± 6.20**
R.VentralDC	84.17 ± 2.85	82.80 ± 3.19	83.57 ± 2.14	83.20 ± 3.55	83.17 ± 2.50	83.84 ± 2.50	**84.46 ± 2.58**[a,b]
Overall	82.79 ± 4.11	83.24 ± 3.60	83.19 ± 5.04	84.06 ± 3.87	84.06 ± 3.38	85.43 ± 3.74	**86.29 ± 3.27**[a,b]

[a]The label index indicates $p < 0.05$ with the two-tailed paired t-test between the hierarchical-AF and Hash-AF

[b]The label index indicates $p < 0.05$ with the Wilcoxon rank based t-test between the hierarchical-AF and Hash-AF

Table 2 A quantitative MSD(mm) obtained by six comparation methods and the proposed method for 18-labeled primarily subcortical regions in the IBSR dataset

Brain regions	MV	NPBL	AF	Hierarchical-AF	Hash-MV	Coarse-AF	Hash-AF
L.lateral ventricle	8.94 ± 4.42	5.04 ± 2.77	10.22 ± 6.73	7.31 ± 6.52	4.47 ± 2.99	4.52 ± 3.02	**3.86 ± 0.87**
L.thelamus	3.72 ± 0.78	4.03 ± 0.91	6.16 ± 7.94	6.65 ± 8.33	3.96 ± 0.81	3.91 ± 0.84	**3.18 ± 0.64**
L.caudate	3.54 ± 0.93	3.75 ± 0.96	4.69 ±.99	6.42 ± 6.28	3.72 ± 0.56	3.68 ± 0.67	**3.32 ± 0.49**
L.putamen	3.83 ± 1.67	4.22 ± 1.89	5.26 ± 1.72	6.43 ± 4.98	4.08 ± 1.69	3.92 ± 1.03	**3.46 ± 1.14**
L.pallidum	3.52 ± 1.02	3.89 ± 1.16	4.43 ± 0.95	3.61 ± 0.88	3.61 ± 0.80	3.52 ± 0.86	**3.13 ± 0.70**
3rd Ventricle	4.79 ± 1.84	4.95 ± 1.55	5.15 ± 1.32	5.42 ± 2.87	3.50 ± 1.17	3.39 ± 0.77	**2.67 ± 0.59**
4th ventricle	5.88 ± 2.78	5.91 ± 2.67	6.54 ± 2.48	6.07 ± 2.76	5.60 ± 2.83	5.92 ± 2.62	**3.75 ± 1.46**
L.hippocampus	4.25 ± 0.97	4.45 ± 1.30	4.92 ± 1.07	5.82 ± 4.22	4.05 ± 1.08	4.36 ± 1.31	**3.68 ± 0.63**
L.amygdala	**3.89 ± 1.48**	4.24 ± 1.23	5.07 ± 1.70	4.45 ± 1.41	4.57 ± 1.53	4.13 ± 1.42	3.94 ± 1.70
L.VentralDC	4.37 ± 0.96	4.54 ±.98	5.31 ± 3.73	4.64 ± 2.95	4.16 ± 1.16	4.19 ± 1.14	**3.88 ± 1.19**
R.lateral ventricle	13.69 ± 5.86	7.51 ± 5.92	12.90 ± 9.28	8.78 ± 7.19	4.42 ± 1.17	5.22 ± 3.50	**3.65 ± 1.11**
R.thelamus	3.91 ± 0.87	3.94 ± 0.90	4.78 ± 1.36	5.85 ± 4.76	3.72 ± 0.85	3.80 ± 0.82	**3.13 ± 0.65**
R.caudate	3.32 ± 0.91	3.36 ± 0.75	5.35 ± 3.45	11.30 ± 10.91	4.13 ± 0.86	4.03 ± 0.85	**3.01 ± 0.60**
R.putamen	**3.70 ± 1.21**	4.58 ± 1.49	5.10 ± 1.78	3.92 ± 1.03	4.20 ± 0.97	3.99 ± 0.87	3.72 ± 0.67
R.pallidum	**3.12 ± 0.59**	3.47 ± 0.57	4.58 ± 0.84	3.76 ± 0.60	3.95 ± 0.97	3.94 ± 0.90	3.49 ± 0.78
R.hippocampus	**4.14 ± 0.81**	4.43 ± 1.29	6.16 ± 2.13	5.22 ± 1.25	5.32 ± 1.26	5.33 ± 1.19	4.19 ± 1.21
R.amygdala	4.20 ± 1.63	4.55 ± 1.43	5.20 ± 1.67	4.88 ± 2.23	4.13 ± 1.52	4.41 ± 1.48	**3.40 ± 1.19**
R.VentralDC	4.07 ± 1.04	4.40 ±.94	6.18 ± 3.26	5.60 ± 2.85	3.71 ± 1.10	3.85 ± 1.08	**3.58 ± 0.92**
Overall	4.83 ± 1.65	4.51 ± 1.60	6.00 ± 2.91	5.90 ± 4.00	4.18 ± 1.30	4.23 ± 1.35	**3.50 ± 0.92**

The comparison results are shown in Tables 1 and 2, which indicate that the overall improvement of DSC is more than 5% and the performance of MSD is also better than other compared method. The experiments reveal that AF-based methods have more excellent performance when handling an image with multiple tissues, while the MV-based methods have better result when dealing with the local information. The proposed Hash-AF method integers the advance of MV-based methods and AF-based methods through hash-based data retrieval so that the segmentation has a higher DSC and lower MSD.

Supervoxel based method proposed by Huo et al. [31] can label the multiple brain tissues and obtain average DSC 85.2% ± 1.6% in IBSR dataset. The DSC in our proposed method is 86.29% ± 3.27% when labeling the same dataset, which is better than the Supervoxel based method.

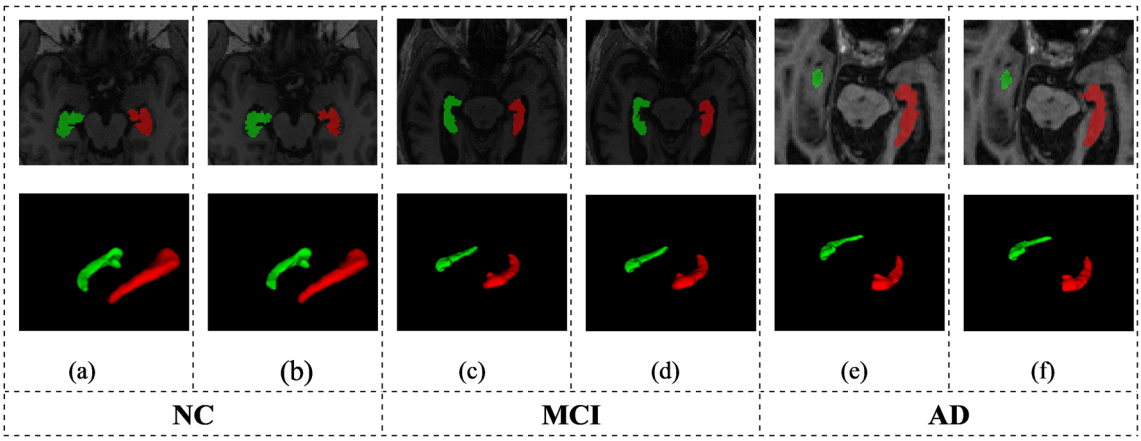

Fig. 6 **a**, **c**, and **e** the manual labeling of left and right hippocampus from NC, MCI and AD subjects in ADNI dataset, **b**, **d**, and **f** the corresponding labeling obtained by our proposed method

Table 3 A quantitative DSC(%) obtained by six comparation methods and the proposed method for the labeled hippocampus in the ADNI dataset

Subject	Brain region	MV	NPBL	AF	Hierarchical-AF	Hash-MV	Coarse-AF	Hash-AF
NC	L.hippocampus	81.35 ± 3.95	86.84 ± 2.41	82.15 ± 3.70	80.14 ± 5.73	84.69 ± 2.54	85.15 ± 3.08	**87.43 ± 2.50**[a,b]
	R.hippocampus	80.87 ± 4.35	87.04 ± 2.81	83.25 ± 3.92	81.72 ± 4.85	84.82 ± 2.73	83.89 ± 3.63	**87.65 ± 2.31**[b]
	Overall	81.11 ± 4.15	86.94 ± 2.61	82.70 ± 3.81	80.93 ± 5.29	84.75 ± 2.64	84.52 ± 3.35	**87.54 ± 2.40**[a,b]
MCI	L.hippocampus	78.81 ± 5.32	84.71 ± 6.95	81.81 ± 4.52	79.96 ± 6.38	83.58 ± 5.60	84.41 ± 3.52	**86.19 ± 4.17**[a,b]
	R.hippocampus	79.28 ± 7.28	82.14 ± 15.68	82.96 ± 5.79	79.57 ± 9.73	80.55 ± 10.74	81.95 ± 6.95	**83.92 ± 7.25**[b]
	Overall	79.04 ± 6.30	83.42 ± 11.32	82.39 ± 5.16	79.77 ± 8.06	82.07 ± 8.17	83.18 ± 5.23	**85.05 ± 5.71**[a,b]
AD	L.hippocampus	77.14 ± 9.00	**86.00 ± 4.67**	81.34 ± 5.37	77.41 ± 6.68	84.98 ± 4.74	83.70 ± 4.14	85.59 ± 5.27
	R.hippocampus	78.11 ± 6.25	85.67 ± 5.42	82.42 ± 4.00	79.60 ± 6.20	83.96 ± 6.07	79.73 ± 5.63	**87.46 ± 4.40**[a,b]
	Overall	77.63 ± 7.62	85.83 ± 5.04	81.88 ± 4.69	78.51 ± 6.44	84.47 ± 5.41	81.72 ± 4.89	**86.53 ± 4.83**[a,b]

[a]The label index indicates $p < 0.05$ with the two-tailed paired t-test between the hierarchical-AF and Hash-AF

[b]The label index indicates $p < 0.05$ with the Wilcoxon rank based t-test between the hierarchical-AF and Hash-AF

Table 4 A quantitative MSD(mm) obtained by six comparation methods and the proposed method for the labeled hippocampus in the ADNI dataset

Subject	Brain region	MV	NPBL	AF	hierarchical-AF	Hash-MV	Coarse-AF	Hash-AF
NC	L.hippocampus	2.92 ± 0.86	2.51 ± 0.72	3.22 ± 0.67	5.29 ± 10.25	3.36 ± 0.71	2.79 ± 0.71	**2.36 ± 0.62**
	R.hippocampus	2.57 ± 0.48	2.70 ± 0.87	3.27 ± 0.89	4.45 ± 7.17	2.91 ± 0.53	2.61 ± 0.63	**2.20 ± 0.67**
	Overall	2.75 ± 0.67	2.61 ± 0.80	3.25 ± 0.78	4.87 ± 8.71	3.14 ± 0.62	2.70 ± 0.67	**2.28 ± 0.65**
MCI	L.hippocampus	3.44 ± 1.49	3.08 ± 1.98	3.70 ± 1.16	4.91 ± 7.23	3.81 ± 1.28	3.24 ± 1.32	**2.75 ± 2.11**
	R.hippocampus	3.12 ± 1.11	3.01 ± 2.05	3.92 ± 1.57	3.06 ± 1.18	3.57 ± 1.12	3.11 ± 1.19	**2.88 ± 2.10**
	Overall	3.28 ± 16.30	3.05 ± 2.02	3.81 ± 1.37	3.99 ± 4.21	3.69 ± 1.20	3.18 ± 1.26	**2.82 ± 2.11**
AD	L.hippocampus	3.12 ± 0.77	2.85 ± 0.99	5.53 ± 7.37	6.00 ± 9.28	3.62 ± 1.17	**2.85 ± 0.66**	2.85 ± 0.99
	R.hippocampus	2.63 ± 0.44	2.53 ± 0.64	3.28 ± 0.55	2.57 ± 0.39	2.93 ± 0.46	2.54 ± 0.42	**2.29 ± 0.68**
	Overall	2.88 ± 0.61	2.69 ±.82	4.41 ± 3.96	4.29 ± 4.84	3.28 ±.82	2.70 ± 0.54	**2.57 ± 0.84**

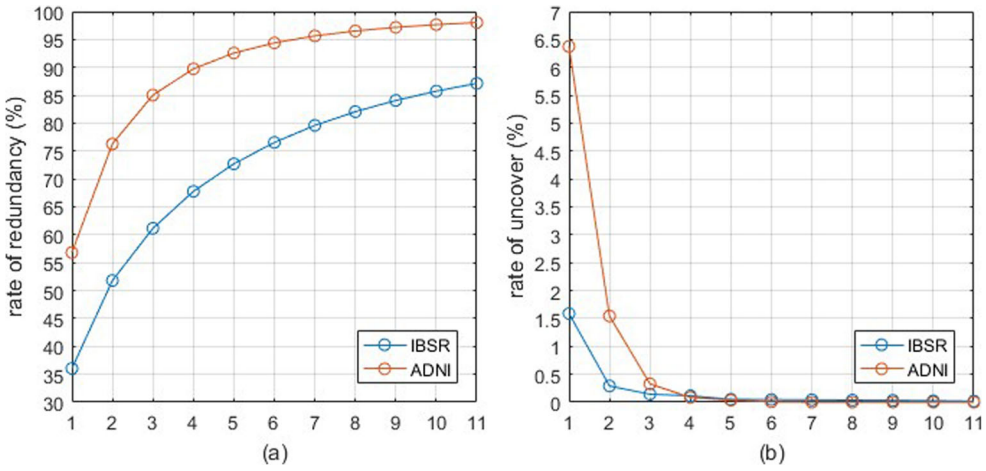

Fig. 7 The influence of parameter r of the coarse classification. **a** shows the tendency of the rate redundancy with r. **b** shows the tendency of the rate of uncover with r

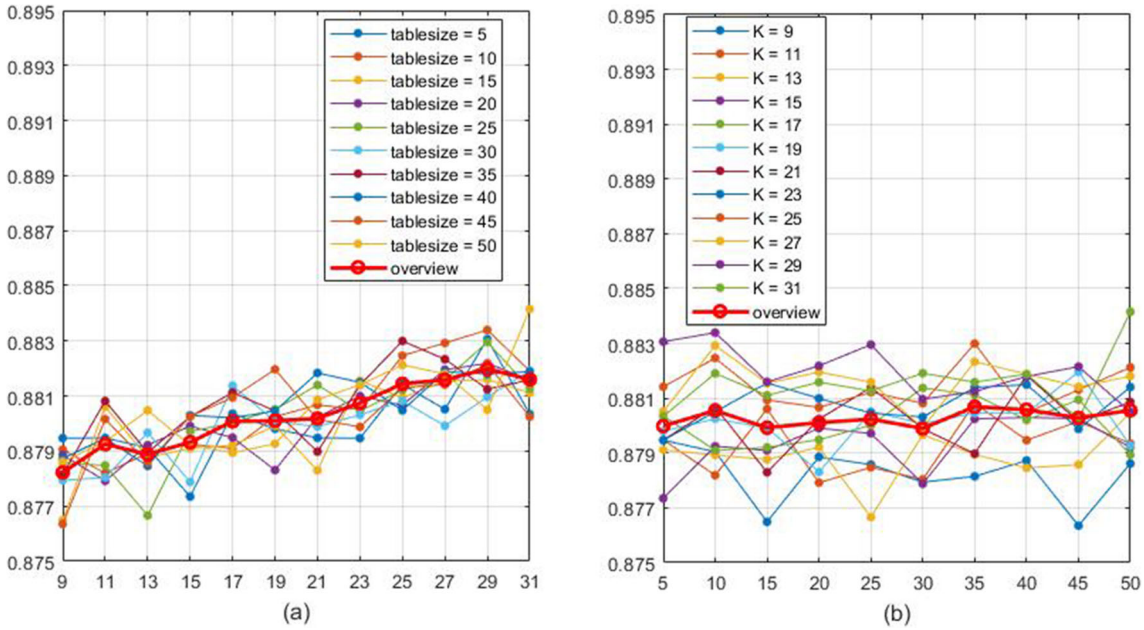

Fig. 8 The influence of parameters in samples retrieval. **a** shows the tend of labeling accuracy with K, **b** shows the tend of labeling accuracy with $tablesize$

Evaluation on ADNI dataset

ADNI[3] dataset provides rich atlases from three different subjects, in which each atlas consists of a brain image obtained by 1.5T MR scanner and a corresponding manual-labeled image with left and right hippocampus. We arbitrary selected 90 cases from the subjects, including 30 from Normal Control (NC) subjects, 30 from Mild Cognitive Impairment (MCI) subjects and 30 from Alzheimer's disease (AD) subjects. The proposed method is evaluated on the ADNI dataset to validate its efficiency when labeling normal and diseased tissue. Figure 6a, c and e show the manual segmentation, which means the histomorphology of tissues in difference subjects are variation. The Fig. 6b, d and e show the results produced by the proposed method.

The proposed method is compared with the six comparative methods described in "Evaluation on IBSR dataset" and the results are shown in Tables 3 and 4, which indicate that the overall improvement of DSC is around 7% and the performance of MSD is also better than other methods for all the three evaluate subjects. The experiments showed the AF-Hash method is capable to segment tissues with disease.

The multi-task neural networks based method for segmenting the hippocampus, Cao et al.(2018) tested 797 subjects in ADNI and obtained the DSC with 89.3%±1.3%. Our DSC result is 86.37% ± 4.31% which was trained on 90 subjects. The deep learning method has a better accuracy, however, the method needs a large number of atlases and a high-performance GPU NVIDIA GTX TITAN 12GB. Moreover, their article did not reveal the result on other dataset and it did not evaluate the performance on segmenting other brain tissues or labeling multiple brain tissues at the same time.

The influence of parameters

Parameter in coarse classification

The parameter r determine the size of the of coarse classification. Figure 7 illustrates the influence of this parameter, indicating that the uncovering region will vanish with the increase of r, while the redundancy among the classifications will increase. An appropriate r can void the uncovering and reduce the redundancy at the same time.For these reasons, we set $r = 5$ in our experiments.

Parameters in sample retrieval

The parameter $tablesize$ means the number of hash function group when encoding samples and the parameter k means the number of similar samples that each target voxel selected. Figure 8 illustrates the influence of those two parameters on the segment accuracy, which indicates that accuracy will increased along with K from 87.8% to 88.2%, while $tablesize$ has little effect on the result with the accuracy fluctuating around 88%. We set $K = 21$ and $tablesize = 20$ considering the robustness and accuracy.

[3] http://adni.loni.ucla.edu

Conclusion

In this study, we proposed an automatic labeling method through the hashing retrieval based atlas forest, which is abbreviated as the Hash-Hybrid Forest(Hash-AF). This method considers the voxels in atlases as independent samples and retrieves them through hashing based method. The proposed method trains a target-orient learning model using the hybrid atlas forest to integrate the information among the atlases, and produces results without registration to avoid the errors generated by label propagation.

We conducted several experiments on the IBSR and ADNI dataset to evaluate the efficiency of proposed method. The results show that the new method is capable to label the targets with multiple tissues and the targets with slight morphological variation tissues. Then, we verified the efficiency of components of the method. Experiments results indicate that the hash-based data retrieval and the hybrid atlas forest are both helpful in improving the accuracy of labeling, and the integration of two strategies can further the results.

In the future work, we will compare the proposed method with other the art-of-the-state methods and extend the study to other medical image segmentation tasks.

Acknowledgments This research was supported by the National Key R&D Program of China(2017YFC0112804).

Funding This study was funded by National Key R&D Program of China(2017YFC0112804).

Compliance with Ethical Standards

Conflict of Interest All authors of this research paper declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Bauer, S., Wiest, R., Nolte, L. P., and Reyes, M., A survey of MRI-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* 58(13):R97, 2013.
2. Burgos, N., Guerreiro, F., McClelland, J., Presles, B., Modat, M., Nill, S., Dearnaley, D., deSouza, N., Oelfke, U., and Knopf, A.-C., Iterative framework for the joint segmentation and ct synthesis of mr images: Application to mri-only radiotherapy treatment planning. *Phys. Med. Biol.* 62(11):4237–4253, 2017.
3. Van Der Lijn, F., en Heijer, T., Breteler, M. M. B., and Niessen, W. J., Hippocampus segmentation in mr images using atlas registration, voxel classification, and graph cuts. *Neuroimage* 43(4):708–720, 2008.
4. García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L., Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Med. Image Anal.* 17(1):1–18, 2013.
5. Iglesias, J. E., Van Leemput, K., Augustinack, J., Insausti, R., Fischl, B., and Reuter, M., Bayesian longitudinal segmentation of hippocampal substructures in brain mri using subject-specific atlases. *Neuroimage* 141:542–555, 2016.
6. Klein, A., and Tourville, J., 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* 6:171, 2012.
7. Išgum, I., Benders, M. J., Avants, B., Cardoso, M. J., Counsell, S. J., Gomez, E. F., Gui, L., Hűppi, P. S., Kersbergen, K. J., and Makropoulos, A., Evaluation of automatic neonatal brain segmentation algorithms: The neobrains12 challenge. *Med. Image Anal.* 20(1):135–151, 2015.
8. Makropoulos, A., Gousias, I. S., Ledig, C., Aljabar, P., Serag, A., Hajnal, J. V., Edwards, A. D., Counsell, S. J., and Rueckert, D., Automatic whole brain MRI segmentation of the developing neonatal brain. *IEEE Trans. Med. Imaging* 33(9):1818–1831, 2014.
9. Rohlfing, T., and Maurer, C. R., Multi-classifier framework for atlas-based image segmentation. *Pattern Recognit. Lett.* 26(13):2070–2079, 2005.
10. Vemuri, B. C., Ye, J., Chen, Y., and Leonard, C. M., Image registration via level-set motion: Applications to atlas-based segmentation. *Med. Image Anal.* 7(1):1–20, 2003.
11. Suh, J. W., Schaap, M., Lee, A., Do, N., Ahiekpor-Dravi, A., Bai, Y., Choi, G., and Moreau-Gobard, R., Automatic multi-atlas segmentation using dual registrations. In: *2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, pp. 1284–1287: IEEE, 2013.
12. Zikic, D., Glocker, B., and Criminisi, A., Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. *Med. Image Anal.* 18(8):1262–1273, 2014.
13. Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D., and Hammers, A., Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* 33(1):115–26, 2006.
14. Romero, J. E., Manjón, J. V., Tohka, J., Coupé, P., and Robles, M., Nabs: Non-local automatic brain hemisphere segmentation. *Magn. Reson. Imaging* 33(4):474–484, 2015.
15. Rousseau, F., Habas, P. A., and Studholme, C., A supervised patch-based approach for human brain labeling. *IEEE Trans. Med. Imaging* 30(10):1852–62, 2011.
16. Wang, Z., Wolz, R., Tong, T., and Rueckert, D., Spatially aware patch-based segmentation (saps): An alternative patch-based segmentation framework. In: *International Conference on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging*, pp. 93–103, 2012.
17. Wu, G., Wang, Q., Zhang, D., Nie, F., Huang, H., and Shen, D., A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Med. Image Anal.* 18(6):881–90, 2014.
18. Bai, W., Shi, W., O'Regan, D. P., Tong, T., Wang, H., Jamilcopley, S, Peters, N. S., and Rueckert, D., A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: Application to cardiac mr images. *IEEE Trans. Med. Imaging* 32(7):1302–15, 2013.
19. Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., and Yushkevich, P. A., Multi-atlas segmentation with joint label fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 35(3):611–623, 2013.
20. Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., and Rueckert, D., Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage* 46(3):726–738, 2009.
21. Wang, H., and Yushkevich, P. A., Groupwise segmentation with multi-atlas joint label fusion. *Medical Image Computing and*

Computer-Assisted Intervention – MICCAI 2013 16(1):711–718, 2013.
22. Sabuncu, M. R., Yeo, B. T., Van Leemput, K., Fischl, B., and Golland, P., A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Imaging* 29(10):1714, 2010.
23. Tong, T., Wolz, R., Hajnal, J. V., and Rueckert, D., Segmentation of brain mr images via sparse patch representation. In: *MICCAI Workshop on Sparsity Techniques in Medical Imaging (STMI)*, 2012.
24. Bai, W., Shi, W., Ledig, C., and Rueckert, D., Multi-atlas segmentation with augmented features for cardiac mr images. *Med. Image Anal.* 19(1):98–109, 2015.
25. Hao, Y., Wang, T., Zhang, X., Duan, Y., Yu, C., Jiang, T., and Fan, Y., Initiative Alzheimer's Disease Neuroimaging. Local label learning (lll) for subcortical structure segmentation: Application to hippocampus segmentation. *Hum Brain Mapp.* 35(6):2674–97, 2014.
26. Akselrod-Ballin, A., Galun, M., Gomori, M. J., Basri, R., and Brandt, A., Atlas guided identification of brain structures by combining 3d segmentation and svm classification, pp. 209–216: Springer, 2006.
27. Kasiri, K., Kazemi, K., Dehghani, M. J., and Helfroush, M. S., Atlas-based segmentation of brain mr images using least square support vector machines. In: *2010 2nd International Conference on Image Processing Theory Tools and Applications (IPTA)*, pp. 306–310: IEEE, 2010.
28. Zhang, L., Wang, Q., Gao, Y., Wu, G., and Shen, D., Automatic labeling of mr brain images by hierarchical learning of atlas forests. *Med. Phys.* 43(3):1175, 2016.
29. Chen, H., Dou, Q., Yu, L., and Heng, P.-A., Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. arXiv preprint arXiv:1608.05895, 2016.
30. Cao, L., Li, L., Zheng, J., Fan, X., Yin, F., Shen, H., and Zhang, J., Multi-task neural networks for joint hippocampus segmentation and clinical score regression. *Multimed. Tools Appl.* 77(22):1–18, 2018.
31. Huo, J., Wu, J., Cao, J., and Wang, G., Supervoxel based method for multi-atlas segmentation of brain mr images. *Neuroimage* 175:201–214, 2018.
32. Quinlan, J. R., Induction of decision trees. *Mach. Learn.* 1(1):81–106, 1986.
33. Fonov, V., Pruessner, J., Robles, M., and Collins, D. L., Nonlocal patch-based label fusion for hippocampus segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 129–136, 2010.