

OPEN

# Label fusion method combining pixel greyscale probability for brain MR segmentation

Monan Wang\* &amp; Pengcheng Li

Multi-atlas-based segmentation (MAS) methods have demonstrated superior performance in the field of automatic image segmentation, and label fusion is an important part of MAS methods. In this paper, we propose a label fusion method that incorporates pixel greyscale probability information. The proposed method combines the advantages of label fusion methods based on sparse representation (SRLF) and weighted voting methods using patch similarity weights (PSWV) and introduces pixel greyscale probability information to improve the segmentation accuracy. We apply the proposed method to the segmentation of deep brain tissues in challenging 3D brain MR images from publicly available IBSR datasets, including images of the thalamus, hippocampus, caudate, putamen, pallidum and amygdala. The experimental results show that the proposed method has higher segmentation accuracy and robustness than the related methods. Compared with the state-of-the-art methods, the proposed method obtains the best putamen, pallidum and amygdala segmentation results and hippocampus and caudate segmentation results that are similar to those of the comparison methods.

Magnetic resonance (MR) imaging technology plays an important role in neuroscience research and clinical applications. Doctors often need to identify regions of interest (ROIs) or diseased tissues from multiple brain images and make judgements or develop treatment plans. At present, a large number of brain images are generated every day in hospitals, and it is a burden for doctors to accurately segment the brain structures by manual segmentation. Therefore, an automatic segmentation technique is needed to deal with the routine analysis of clinical brain MR images. Many automatic image segmentation algorithms have been proposed<sup>1,2</sup>, but due to the complexity of deep brain structures, it is still a challenging task to develop a fast and accurate automatic segmentation method. FreeSurfer<sup>3</sup> and FIRST<sup>4</sup> are widely used automatic segmentation methods. FreeSurfer uses non-linear registration and an atlas-based segmentation approach. FIRST takes the deformable-model-based active appearance model and puts it in a Bayesian framework. These two methods alleviate the burden of manual segmentation, but their accuracy still needs to be improved.

In the past decade, atlas-based methods have shown superior performance over other automatic brain image segmentation algorithms<sup>5</sup>. The main idea of the atlas-based method is to use the prior knowledge of the atlas to classify the target pixels. Each atlas contains an intensity image and a label image labelled by a doctor. This method registers the target image with the intensity images and then uses the obtained deformation field to propagate the label information to the target image space to achieve the classification of the target pixel. A disadvantage of atlas-based methods is that they are sensitive to noise and differences between images. This problem can be solved by MAS<sup>6</sup> and atlas selection<sup>7</sup>.

MAS methods mainly include two steps: image registration and label fusion. At present, many registration algorithms have been proposed<sup>8</sup>. The registration performance of an MAS method has a great influence on the segmentation results. Therefore, most MAS methods use a highly accurate non-linear registration method. Alvé<sup>et al.</sup><sup>9</sup> proposed an Überatlas registration method for MAS, which effectively shortens the registration time. Alchatzidis *et al.*<sup>10</sup> presented a method that integrates registration and segmentation fusion in a pairwise MRF framework.

Label fusion is a key step in MAS methods and can reduce the errors caused by registration errors or morphological differences between the target image and the atlas images. Weighted voting (WV) is a widely used label fusion method. This method can effectively improve the segmentation results. The weight value is mainly defined

School of Mechanical & Power Engineering, Harbin University of Science and Technology, Xue Fu Road No. 52, Nangang District, Harbin City, Heilongjiang Province, 150080, People's Republic of China. \*email: [mnwang@hrbust.edu.cn](mailto:mnwang@hrbust.edu.cn)

by the similarity between the atlas images and the target image; examples include the absolute value of the intensity difference<sup>11</sup>, the Gaussian function of the intensity difference<sup>12</sup>, and the local mutual information<sup>13</sup>. CoupeK *et al.*<sup>14</sup> proposed a hippocampal and ventricular segmentation method based on non-local patches. Wang *et al.*<sup>15</sup> proposed a multi-atlas active contour segmentation method and used a template optimization strategy to reduce registration error. Lin *et al.*<sup>16</sup> proposed a label fusion method based on registration error and intensity similarity. Tang *et al.*<sup>17</sup> proposed a low-rank based image recovery method combined with the sparse representation model, which successfully achieved the segmentation of brain tumours. Roy *et al.*<sup>18</sup> proposed a patch-based MR image tissue classification method. This method uses the sparse dictionary learning method and combines the statistical atlas prior to the sparse dictionary learning. The entire segmentation process does not require deformation registration of the atlas images. Tong *et al.*<sup>19</sup> combined discriminative dictionary learning and sparse coding techniques to design a fixed dictionary for hippocampus segmentation. Lee *et al.*<sup>20</sup> proposed a brain tumour image segmentation method using kernel dictionary learning. This method implicitly maps intensity features to high-dimensional features to make classification more accurate. However, this method is time consuming and cannot be widely applied. To address this problem, a linearized-kernel sparse representative classifier has been proposed<sup>21</sup>.

In recent years, learning-based methods have been widely used in MR image segmentation fields, such as the support vector machine (SVM)<sup>22</sup>, random forest (RF)<sup>23</sup> and convolutional neural network (CNN)<sup>24</sup> methods. These methods are suitable for the image analysis of specific anatomical structures due to their advantages in terms of efficiency in pairwise registration and feature description. Huo *et al.*<sup>25</sup> proposed a multi-atlas segmentation method combined with SVM classifiers and super-voxels. This method is not sensitive to pair registration and reduces registration time by using affine transformation. Xu *et al.*<sup>26</sup> proposed an atlas forest automatic labelling method for brain MR images. They use random forest techniques to encode atlas images into the atlas forest to reduce registration time. Kaiser *et al.*<sup>27</sup> proposed a convolutional neural network model combining convolution and prior spatial features for sub-cortical brain structure segmentation and trained the network using a restricted sample selection to increase segmentation accuracy.

Brain tissue boundary regions are difficult to segment for both the MAS and learning-based segmentation methods. It is mainly because the pixel values of the tissue boundary regions are very similar and it is difficult to identify whether the pixel points of these regions belong to the target organization. In order to overcome this shortcoming and make full use of the prior information of the atlas, we introduce pixel greyscale probability information into the label fusion method.

In this paper, we propose a label fusion method combining pixel greyscale probability (GP-LF) for brain MR image segmentation. The proposed method is mainly aimed at improving the segmentation results in the tissue boundary region and improving the segmentation accuracy. We use the SRLF method and the PSWV method to obtain the fusion results of target tissues and then introduce pixel greyscale probability information to fuse the fusion results obtained by the above two methods. The pixel greyscale probability is obtained through a large amount of segmentation training.

## Methods

We propose our method based on the patch-based label fusion method of MAS. In this section, we introduce the patch-based label fusion method and describe the proposed method in detail.

**Dataset.** We tested the proposed method using the IBSR dataset. The IBSR dataset is a publicly available MR brain image dataset that is provided by the <https://www.nitrc.org/projects/ibsr> website. It contains data from 18 T1-weighted MR image data and the corresponding label images, which are defined by the Center for Morphometric Analysis at Massachusetts General Hospital. The 18 subject images were obtained using two different MRI scanners: GE (1.5T) and SIEMENS (1.5T). The images are from 14 males and 4 females and have dimension  $256 \times 256 \times 128$  and three different resolutions:  $0.84 \times 0.84 \times 1.5 \text{ mm}^3$ ,  $0.94 \times 0.94 \times 1.5 \text{ mm}^3$  and  $1 \times 1 \times 1.5 \text{ mm}^3$ . Additionally, this dataset is part of the Child and Adolescent NeuroDevelopment Initiative<sup>28</sup> (CANDI) and was provided under the Creative Commons: Attribute license<sup>29</sup>.

**Patch-based label fusion method.** The patch-based label fusion method maps the label information to the target image space by registering the target image and the atlas images and then identifies the target tissue pixel by a label fusion algorithm. The detailed process is as follows:

- (1) Image registration: The image to be segmented is taken as the target image T, which is registered with the intensity images in the atlas. Saving the warped atlas  $A_1(I_1, L_1)$ ,  $A_2(I_2, L_2)$ ,  $\dots$ ,  $A_n(I_n, L_n)$ , where  $I_i$  is the warped intensity image of atlas  $A_i$ ,  $L_i$  is the warped label image of atlas  $A_i$ , and  $n$  is the number of atlases, is saved.
- (2) Patch extraction: The patch  $Tp_{x_j}$  centring on the pixel  $x_j$  of the target image is extracted, and then patches  $Ip_i$  and  $Lp_i$  are extracted from the  $x_j$  position of images  $I_i$  and  $L_i$ , respectively.
- (3) Label fusion: A label fusion method is used to fuse the extracted patches.
- (4) Label assignment: The fusion result is binarized to assign a label (background or target organization) to the pixel  $x_j$ .

The WV method is a widely used label fusion method; its equation is as follows:

$$Fv(x_j) = \frac{\sum_{i=0}^n w_i I_i}{\sum_{i=0}^n w_i} \quad (1)$$

$$L_T(x_j) = \begin{cases} 1, & \text{if } Fv(x) \geq \text{Binarization threshold} \\ 0 & \text{if } Fv(x) < \text{Binarization threshold} \end{cases} \quad j = 1, 2 \dots m \tag{2}$$

where  $w_i$  is the weight coefficient,  $Fv(x_j)$  is the fusion result, and  $m$  is the number of target image pixels. If  $L_T(x_j) = 0$ , the pixel  $x_j$  is marked as the background, and if  $L_T(x_j) = 1$ , the pixel  $x_j$  is marked as a target organization pixel.

The calculation of  $w_i$  is the core of the WV method. There are two common methods for calculating label weight: one is to calculate the similarity of  $Tp_{x_j}$  and  $Ip_i$ , and take that as the weight  $w_i$ , and the other is to calculate the weight  $w_i$  by using the sparse representation method.

The SRLF method assumes that  $Tp_{x_j}$ ,  $Ip_i$  and  $Lp_i$  are one-dimensional column vector signals and uses all the images  $Ip_i$  to build an over-complete dictionary  $D_i$ ; the  $Tp_{x_j}$  approximately lie in the subspace spanned by the training patches in the library  $D_i$ .

In this paper, we use  $PT[p_{t_1}, p_{t_2} \dots p_{t_{len}}]^T$  to represent  $Tp_{x_j}$ , where  $p_{t_i}$  corresponds to each pixel in the patch  $Tp_{x_j}$ ,  $len$  is equal to the number of pixels in the  $Tp_{x_j}$ . We use the column vector  $PI_i[p_{i_1}, p_{i_2} \dots p_{i_{len}}]^T$ ,  $i = 1, 2, \dots, m$  to represent  $Ip_i$ , and the over-complete dictionary  $D_i$  is represented as  $D_i[PI_1, PI_2 \dots PI_m]$ . The sparse coefficient  $\alpha = [\alpha_1, \alpha_2 \dots \alpha_m]^T \in R^n$ . Each element of the  $PT$  can be represented by the following formula:

$$p_{t_j} = p_{i_1} \alpha_1 + p_{i_2} \alpha_2 + \dots + p_{i_m} \alpha_m \tag{3}$$

$$PT = D_i \alpha \tag{4}$$

The idea of sparse representation is to represent the signal  $PT$  by solving the most sparse solution  $\alpha$ . This solution for  $\alpha$  can be obtained by solving the following equation:

$$\hat{\alpha} = \arg \min \|\alpha\|_0 \text{ subject to } \|PT_i - D\alpha\|_2^2 \leq \varepsilon \tag{5}$$

The solution for  $\alpha$  is used as the weight  $w_i$  of WV, and then the label fusion is performed by using Eq. (1).

**Proposed method.** Equation (5) is an underdetermined equation, such that the number of non-zero coefficients in the sparse solution is equal to the number of elements of the vector  $Tp_{x_j}$ . Therefore, in the SRLF method, only a small number of patches participate in the fusion process, and most of the patches are discarded, which will lead to segmentation errors. The PSWV method combines all the patches, but is not as good at edge pixel recognition as the SRLF method.

In order to overcome the shortcomings of the above methods, we propose the GP-LF method. The overall framework of the proposed method is shown in Fig. 1. In order to make better use of the prior information of the atlas, we introduce the image grey probability information in the label fusion process and obtain the pixel grey-scale probability value  $P(x_j)$  of the target image through segmentation training.

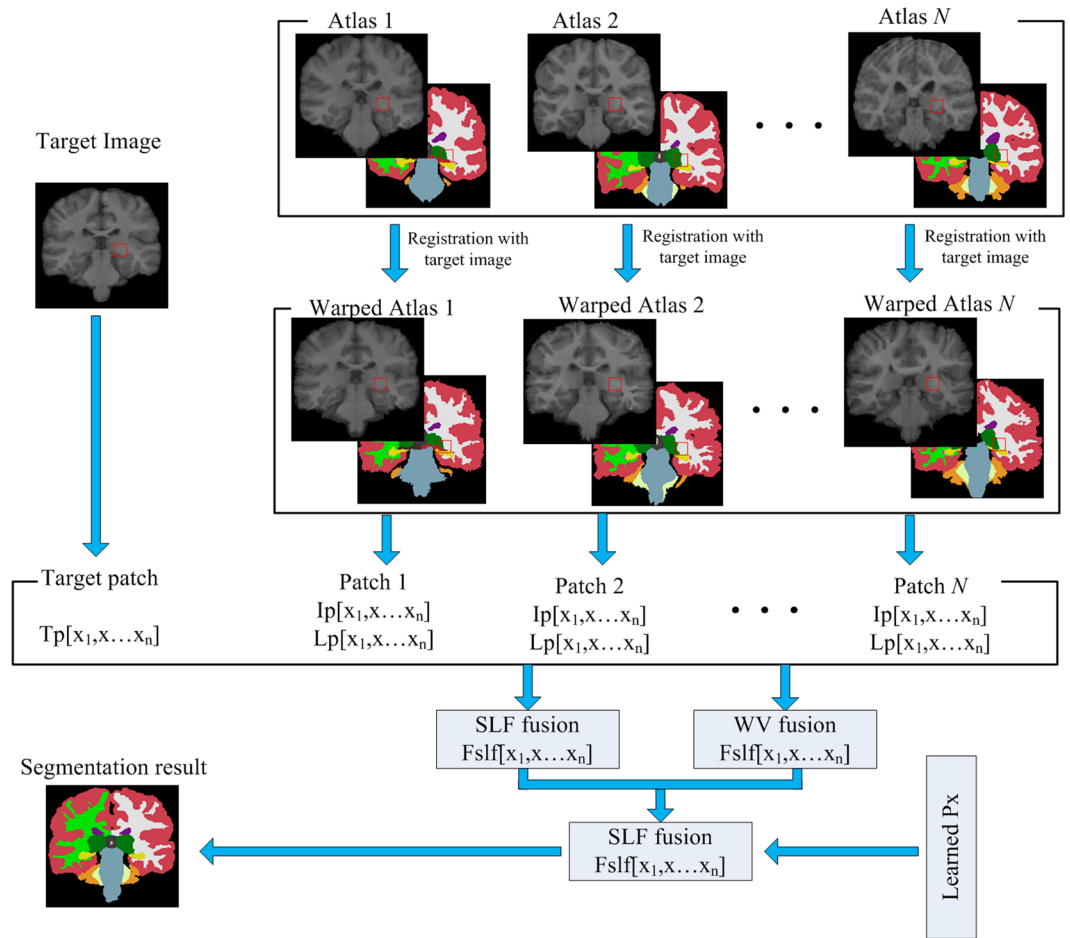
- (1) Multi-atlas registration: The current non-linear registration method has been able to achieve good registration results, so we used an existing non-linear registration method for multi-atlas registration. We selected the SuperElastix registration tool and used the affine and B-spline deformation methods to register the target image and the atlas intensity images. The SuperElastix registration tool is available at <https://github.com/SuperElastix/>.
- (2) Initial label fusion: In order to avoid the influence of non-standard images on the segmentation results, we perform pixel value normalization on the target image and on the atlas intensity images. After multi-atlas registration, the obtained warped atlas is fused using the SRLF method, and the fusion result is represented by  $Fv_{SRLF}(x_j)$ . We use the normalized correlation coefficient (NCC)<sup>30</sup> as a similar measure function to perform PSWV fusion on the warped atlas images, and the fusion result is represented by  $Fv_{PSWV}(x_j)$ .
- (3) Establishment of the GP-LF model: The pixel greyscale probability information is introduced to fuse the results  $Fv_{SRLF}(x_j)$  and  $Fv_{PSWV}(x_j)$  obtained in the above steps. The fusion equation is as follows:

$$L_T(x_j) = \begin{cases} 1 & \text{if } Fv_{SRLF}(x_j) > 0.9 \text{ or } Fv_{PSWV}(x_j) > 0.9 \\ Lf_1(x_j) & \text{if } 0.4 < Fv_{SRLF}(x_j) \leq 0.9 \text{ and } 0.4 < Fv_{PSWV}(x_j) < 0.9 \\ Lf_2(x_j) & \text{if } Fv_{SRLF}(x_j) \leq 0.4 \\ Lf_3(x_j) & \text{if } Fv_{PSWV}(x_j) \leq 0.4 \end{cases} \tag{6}$$

where  $L_T(x_j)$  is a piecewise function. When  $Fv_{SRLF}(x_j)$  and  $Fv_{PSWV}(x_j)$  are greater than 0.9, almost all the identified pixels are target organization pixels. In order to reduce calculation time and not affect segmentation precision, we mark these pixels directly as target organization pixels. When  $Fv_{SRLF}(x_j)$  and  $Fv_{PSWV}(x_j)$  are between 0.4 and 0.9, the fusion is performed by using the formula for  $Lf_1(x_j)$ , which is as follows:

$$Lf_1(x_j) = \begin{cases} 1 & \text{fusion1} > 0.5 \\ 0 & \text{else} \end{cases} \quad \text{fusion1} = \beta_1 \cdot Fv_{SRLF}(x_j) \cdot Fv_{PSWV}(x_j) \cdot P(x_j) \tag{7}$$

when  $Fv_{SRLF}(x_j)$  or  $Fv_{PSWV}(x_j)$  is less than 0.4, the fusion is performed by using the formula for  $Lf_2(x_j)$  or  $Lf_3(x_j)$ , which are as follows:



**Figure 1.** Schematic diagram of the proposed method.

$$Lf_2(x_j) = \begin{cases} 1 & fusion2 > 0.5 \\ 0 & else \end{cases} \quad fusion2 = \beta_2 \cdot Fv_{SLRLF}(x_j) \cdot P(x_j) \tag{8}$$

$$Lf_3(x_j) = \begin{cases} 1 & fusion3 > 0.5 \\ 0 & else \end{cases} \quad fusion3 = \beta_3 \cdot Fv_{PSWV}(x_j) \cdot P(x_j) \tag{9}$$

where  $\beta_1, \beta_2, \beta_3$  are the fusion weight coefficients.

(4)  $P(x_j)$  coefficient training:

Target tissue greyscale range setting consists of the following steps:

- a. After the images are normalized, the pixel value ranges from 0 to 1. Divide the pixel value range 0–1 into  $N_i$  intervals and then extract the greyscale distribution map. Judge the target tissue distribution range based on the peak value of the greyscale map. If the number of peaks is 2 or 3, the upper boundary value of the target organization is set to  $ISN_{max} + Ov$ , where  $ISN_{max}$  is the maximum interval serial number corresponding to the peak and  $Ov$  is the offset value.
- b. If the number of peaks is 1, then divide the pixel value range 0–1 into  $2 \times N_i$  intervals and repeat step 1.
- c. If the number of peaks is greater than 3, then divide the pixel value range 0–1 into  $N_i/2$  intervals and repeat step 1.

Establishment of the  $P(x_j)$  coefficient training model.

The  $P(x_j)$  coefficient training model is defined as follows:

$$Dsc[dsc_1, dsc_2, \dots, dsc_{sn}] \tag{10}$$

$$\hat{P}(x) = \text{are max} \|Dsc\|, \quad 0 \leq P(x_j) \leq 1 \tag{11}$$

where  $dsc$  denote the dice similarity coefficient (Dsc) of the intensity image segmentation result and the label image in an atlas, and  $sn$  denote the number of the atlas used for training. The  $P(x_j)$  coefficient can be obtained by solving equation (11).

$P(x_j)$  coefficient training:

According to the spatial position of the target image, select the atlas images to participate in the  $P(x_j)$  coefficient training. The  $P(x_j)$  of the target tissue will be obtained through training and then Eq. (6) can be used to classify the target organization.

## Experiments and Results

In this section, we use the leave-one-out cross-validation method to verify the segmentation performance and robustness of the proposed method. One atlas image from the IBSR dataset serves as the target image to be segmented, and the others in the dataset are used as the atlas. The above steps are repeated until each atlas image of the IBSR dataset is segmented. In the experiments, we mainly focused on the extraction of subcortical structures. We selected six subcortical structures for segmentation, including the thalamus, hippocampus, caudate, putamen, pallidum and amygdala. We also evaluated the segmentation results of our method, the SRLF method and the PSWV method.

**Segmentation evaluation index.** In order to verify the segmentation performance and robustness of the proposed method, we used the leave-one-out cross-validation method to test the proposed method. We applied four widely used segmentation metrics, Dsc, Recall, Precision and Hausdorff distance (HD) to measure the segmentation result of the proposed method. The Dsc, Recall and Precision segmentation metrics mainly measure the volume overlap ratio between the segmentation results and the manual labels, and the HD measures the surface distance between segmentation results and manual labels. These metrics are defined as follows:

$$Dsc(A, B) = \frac{2V(T \cap F)}{V(T) + V(F)} \quad (12)$$

$$Recall(T, F) = \frac{V(T \cap F)}{V(T)} \quad (13)$$

$$Precision(T, F) = \frac{V(T \cap F)}{V(F)} \quad (14)$$

$$HD(T, F) = \max(H_1(T, F), H_2(F, T)) \quad (15)$$

$$H_1(T, F) = \max_{p_i \in T} \left( \min_{p_j \in F} d(p_i, p_j) \right) \quad (16)$$

$$H_2(F, T) = \max_{p_j \in F} \left( \min_{p_i \in T} d(p_j, p_i) \right) \quad (17)$$

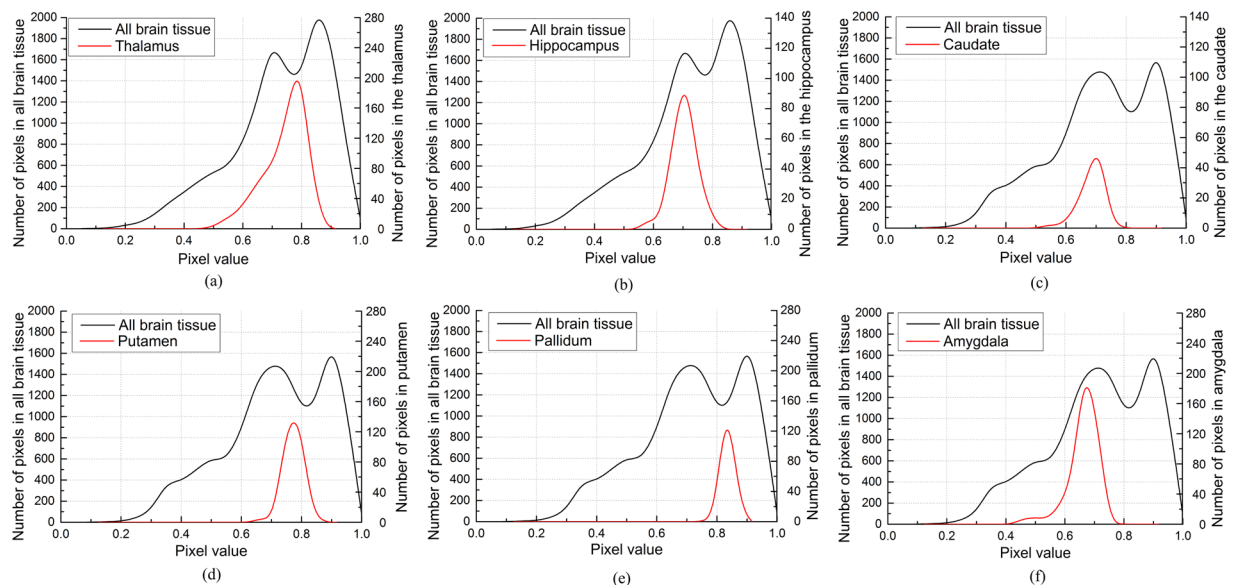
where  $T$  is the target tissue pixel set in the target label image,  $F$  is the segmented target tissue pixel set, and  $d(p_i, p_j)$  is the distance between pixels  $p_i$  and  $p_j$ .

All experiments and training in this study were performed on a single desktop PC (Intel(R) Core(TM) 3.70 GHz CPU).

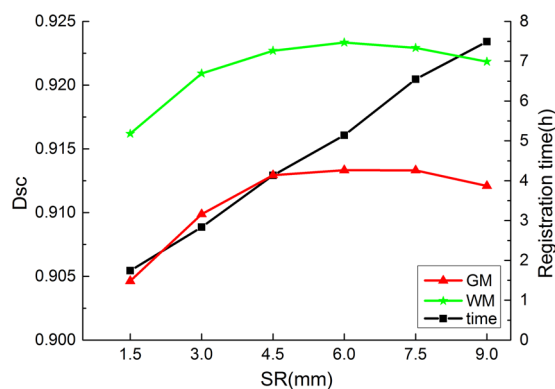
**Target organization pixel greyscale range setting.** We divide the pixel greyscale value range 0–1 into 20 intervals and determine the  $P(x_j)$  of the pixels in each interval by training. In order to reduce the training time, we extracted the greyscale range of the target tissue from the greyscale distribution map. Due to the non-standard intensity between MR images, the same target tissue can have different grey distribution ranges in different images, so we determined the target tissue grey distribution range according to the grey distribution map.

As shown in Fig. 2, the pixel grey value span of deep brain tissue is between 0.3 and 0.6. We used the target tissue greyscale range-setting method mentioned above to obtain the  $ISN_{max}$  of the target image greyscale distribution map and then set the grey value range of the deep brain tissues according to the greyscale distribution map. The thalamus has a pixel value span of 0.6 and its maximum pixel value is  $ISN_{max} + 0.15$ ; the hippocampus has a pixel value span of 0.5 and its maximum pixel value is  $ISN_{max}$ ; the caudate has a pixel value span of 0.5 and its maximum pixel value is  $ISN_{max}$ ; the putamen has a pixel value span of 0.5 and its maximum pixel value is  $ISN_{max} + 0.1$ ; the pallidum has a pixel value span of 0.5 and its maximum pixel value is  $ISN_{max} + 0.15$ ; and the amygdala has a pixel value span of 0.5 and its maximum pixel value is  $ISN_{max} - 0.05$ .

**Influence of the number of patches and atlas selection.** It has been proven that selecting the appropriate atlas images in the MAS method can effectively improve the segmentation result<sup>31</sup>. The atlas images with low similarity to the target image will be discarded by atlas selection. There are large differences among slices in different positions of the human brain, so we selected atlas images to participate in target tissue segmentation based on the spatial location of the target image. Suppose the spatial position of the target image is  $TSP$ ,  $SR$  is the



**Figure 2.** Deep brain tissue greyscale distribution map, (a) Thalamus, (b) Hippocampus, (c) Caudate, (d) Putamen, (e) Pallidum, (f) Amygdala.



**Figure 3.** Segmentation results of different SRs. The green line is WM segmentation result, the red line is GM segmentation result, and the black line is the registration time.

search radius, and the images located in the space from  $TSP - SR$  to  $TSP + SR$  are selected as the images to be fused. The deep brain tissues belong to grey matter (GM) or white matter (WM). Therefore, we used GM and WM as the target tissues to test the influence of atlas selection on the segmentation results.

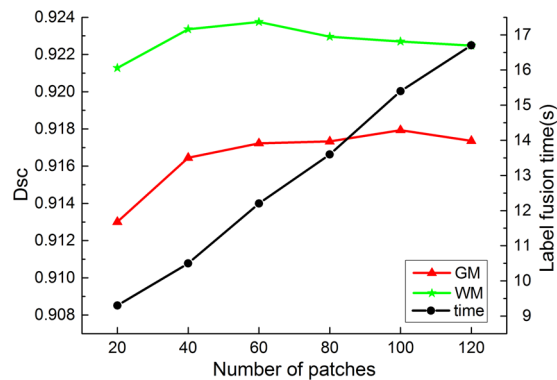
As shown in Fig. 3, the best WM and GM Dscs were obtained when  $SR = 4$ , and the Dsc was similar to  $SR = 4$  when  $SR = 3$  or  $SR = 5$ . However, for every 1.5 mm increase in SR, the registration time increased by approximately 1.5 hours. Therefore, we chose  $SR = 3$  to consider the registration time and segmentation accuracy.

Fusing the patches with lower similarity to  $Tp_{xy}$  will reduce the segmentation accuracy. In this study, we used the NCC method to measure the similarity between each  $I_{p_i}$  and  $Tp_{xy}$  and selected patches with higher similarity values for label fusion. We also tested the effect of varying the number of selected patches on the segmentation results.

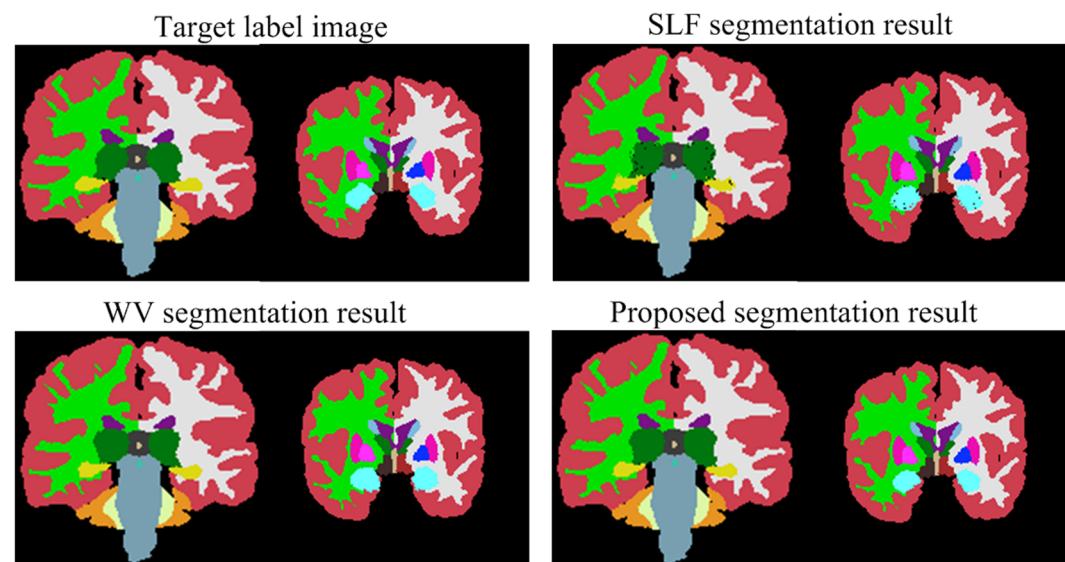
It can be seen from Fig. 4 that the WM and GM segmentation results are best when the number of selected patches is 60, and as the number of selected patches increases, the label fusion time only increases by a few seconds. Therefore, we select 60 patches with high similarity values for label fusion.

**$P(x_j)$  coefficient training and final segmentation results.** In this section, we performed segmentation tests on the thalamus, hippocampus, caudate, putamen, pallidum and amygdala. Firstly, setting the weighting factor  $\beta$  for each target tissue. The  $\beta_1$  is set to 3.13 for all the deep brain tissues mentioned earlier; the  $\beta_2 = 1.25$  and  $\beta_3 = 0.67$  for thalamus; the  $\beta_2 = 1.67$  and  $\beta_3 = 2.5$  for hippocampus; the  $\beta_2 = 0.625$  and  $\beta_3 = 0.83$  for caudate; the  $\beta_2 = 0.72$  and  $\beta_3 = 1$  for putamen; the  $\beta_2 = 0.83$  and  $\beta_3 = 1.25$  for pallidum; the  $\beta_2 = 1$  and  $\beta_3 = 0.25$  for amygdala. Secondly, selecting the atlas images located in  $TSP - 3\text{ mm}$  to  $TSP + 3\text{ mm}$  as the training data. Thirdly,





**Figure 4.** Segmentation results of numbers of selected patches. The green line is WM segmentation result, the red line is GM segmentation result, and the black line is the label fusion time.



**Figure 5.** Sample segmentation results of the proposed method, SRLF method and PSWV method.

performing segmentation training on the target organization to obtain  $P(x_i)$  coefficients. Finally, using Eq. (6) to identify the target tissues.

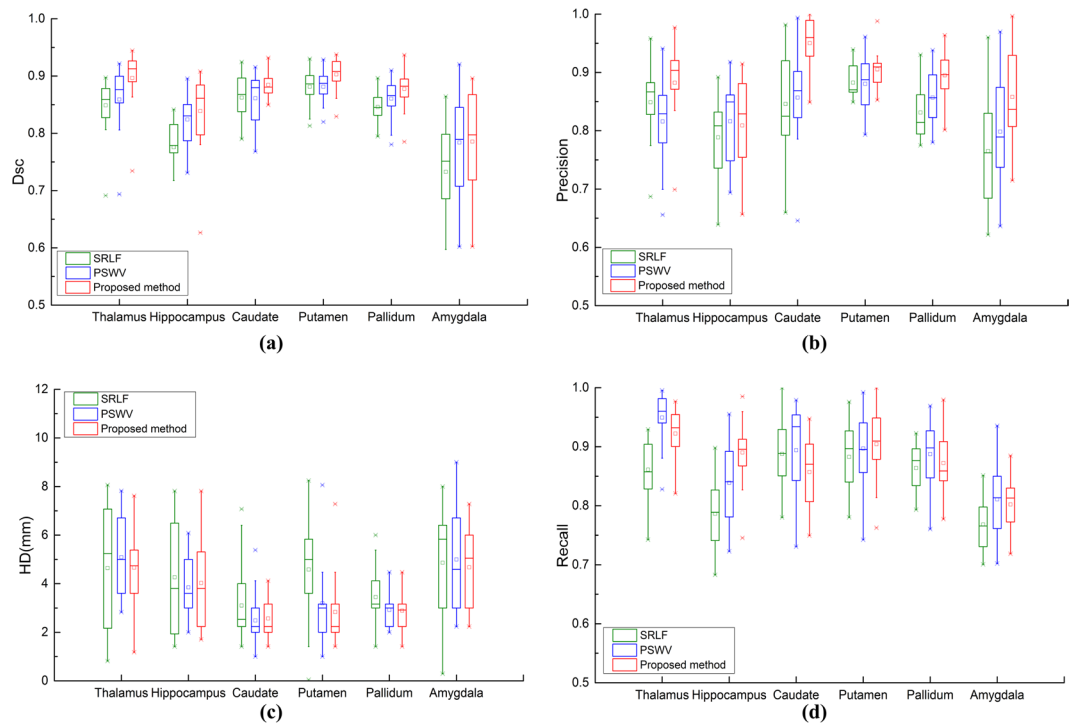
As shown in Fig. 5, there are some empty points (pixels that are not accurately classified) within the target tissues that are classified by the SRLF method, and the WV method does not work well for pixel classification of target tissue boundaries. The proposed method effectively overcomes the shortcomings of the above two methods and obtains a segmentation result similar to the target label image.

We use four metrics to verify the segmentation performance of the proposed method, the SRLF method and the PSWV method; the segmentation results are shown in Fig. 6. It can be seen from Fig. 6(a,b) that the proposed method achieves the best Dsc and Precision results. This shows that our method achieves the best segmentation accuracy and has fewer pixels for false recognition. As shown in Fig. 6(c), for hippocampus, putamen and pallidum segmentation, the proposed method obtains the best recall results. For thalamus, caudate and amygdala segmentation, the proposed method obtains Recall results similar to those of the PSWV method. As shown in Fig. 6(d), for thalamus, caudate, putamen and pallidum segmentation, the proposed method achieves the smallest HD. For hippocampus and amygdala segmentation, the proposed method achieves an HD similar to those of the SRLF and PSWV methods.

## Discussion and Conclusion

In this section, we compare the performance of the proposed method to the related methods and state-of-the-art methods in deep brain tissue segmentation. We also introduce the limitations and future directions of the proposed method.

One disadvantage of the PSWV approach is the fusion of all extracted patches. Although the labels are assigned weights  $w_i$  based on patch similarity, the fusion results are still affected by the label frequency. Due to the influence of registration error, there is a large difference between the organizational boundaries of warped atlas labels and those of the target image label, which will cause the PSWV method to be unsuitable for the



**Figure 6.** Deep brain tissue segmentation results of the proposed method, SRLF method and PSWV method.

Target tissues	FIRST <sup>4</sup>	FreeSurfer <sup>3</sup>	Lin <i>et al.</i> <sup>16</sup> method	Kaisar <i>et al.</i> <sup>27</sup> method	Xu <i>et al.</i> <sup>26</sup> method	Liu <i>et al.</i> <sup>21</sup> method	Proposed method
L Thalamus	0.889 ± 0.018	0.830 ± 0.018	0.87 ± 0.03	0.910 ± 0.014	<b>0.921 ± 0.017</b>	0.904	0.914 ± 0.022
R Thalamus	0.890 ± 0.017	0.849 ± 0.021	0.86 ± 0.03	0.914 ± 0.016	<b>0.920 ± 0.018</b>		0.902 ± 0.024
L Hippocampus	0.809 ± 0.022	0.784 ± 0.054	0.77 ± 0.02	<b>0.851 ± 0.024</b>	0.819 ± 0.050	0.829	<b>0.857 ± 0.045</b>
R Hippocampus	0.810 ± 0.140	0.794 ± 0.025	0.76 ± 0.02	<b>0.851 ± 0.024</b>	0.842 ± 0.039		0.849 ± 0.057
L Caudate	0.797 ± 0.046	0.808 ± 0.079	0.78 ± 0.04	0.896 ± 0.018	0.824 ± 0.079	0.862	0.885 ± 0.021
R Caudate	0.837 ± 0.117	0.801 ± 0.042	0.81 ± 0.03	0.896 ± 0.020	0.854 ± 0.044		<b>0.898 ± 0.027</b>
L Putamen	0.860 ± 0.060	0.771 ± 0.039	0.84 ± 0.02	0.900 ± 0.014	0.900 ± 0.024	0.897	<b>0.903 ± 0.037</b>
R Putamen	0.876 ± 0.080	0.799 ± 0.026	0.84 ± 0.02	0.904 ± 0.012	0.905 ± 0.026		<b>0.905 ± 0.024</b>
L Pallidum	0.815 ± 0.088	0.693 ± 0.189	0.71 ± 0.07	0.825 ± 0.050	0.752 ± 0.080	0.814	<b>0.878 ± 0.029</b>
R Pallidum	0.799 ± 0.060	0.792 ± 0.085	0.74 ± 0.05	0.829 ± 0.046	0.812 ± 0.039		<b>0.876 ± 0.031</b>
L Amygdala	0.721 ± 0.053	0.585 ± 0.064	0.70 ± 0.06	0.763 ± 0.052	0.723 ± 0.076	0.753	<b>0.801 ± 0.055</b>
R Amygdala	0.707 ± 0.054	0.576 ± 0.076	0.67 ± 0.07	0.768 ± 0.058	0.729 ± 0.091		<b>0.795 ± 0.079</b>
Average	0.818	0.757	0.780	0.859	0.833	0.843	<b>0.872</b>

**Table 1.** Comparison of the proposed method with the state-of-the-art methods on IBSR dataset in terms of Dsc and standard deviation.

identification of boundary pixels. The SRLF method selects the least element from the over-complete dictionary  $D_i$  and represents  $PT$  by a linear combination of the selected elements. Therefore, most of the patches in the SRLF method are discarded, which will result in empty points in the classified target organization and reduce the segmentation accuracy.

In order to overcome the shortcomings of the above methods, We combine pixel greyscale probability information to fuse the fusion results obtained by the above two methods. Most of the segmentation errors in the multi-atlas patch-based label fusion method are concentrated on the tissue boundary. This is because the pixel gradient at the tissue boundary is not obvious, but there are still some differences in pixel grey values of different tissues. Therefore, the segmentation accuracy can be effectively improved by assigning a probability of belonging to the target tissue to different pixel values. As shown in Fig. 6, compared to the PSWV method and the SRLF method, the proposed method obtained the best Dsc and Precision results and obtained better or similar Recall and HD results.

We compared the proposed method with the state-of-the-art methods on the IBSR dataset in terms of Dsc and standard deviation. As shown in Table 1, the proposed method performed better than both FIRST and FreeSurfer for the six deep brain tissues. The overall Dsc mean for our method was significantly higher than for both of these



other methods, with mean Dscs of 0.818, 0.757 and 0.872 for FISRT, FreeSurfer and the proposed method, respectively. In the pallidum and amygdala segmentation, the proposed method obtained the highest Dsc values of 0.878, 0.876, 0.801 and 0.795 for left pallidum, right pallidum, left amygdala and right amygdala, respectively. In the putamen segmentation, our method obtained the best left putamen segmentation results with a Dsc of 0.903 and obtained the same right putamen segmentation results as Xue's method (Dsc = 0.905). The proposed method obtained similar caudate and hippocampus segmentation results to those obtained using Kaiser's method. In the caudate segmentation, our method obtained Dscs of 0.885 and 0.898 for the left caudate and right caudate, respectively, and Kaiser's method obtained Dscs of 0.896 and 0.896 for the left caudate and right caudate, respectively. In the hippocampus segmentation, our method obtained Dscs of 0.857 and 0.849 for the left hippocampus and right hippocampus, respectively, and Kaiser's method obtained Dscs of 0.851 and 0.851 for the left hippocampus and right hippocampus, respectively. The proposed method obtained the third highest hippocampus segmentation accuracy among all the evaluated methods. This is because of the influence of the non-standard image intensity; the results of the thalamic segmentation of different target images using the same parameters are very different. Therefore, it is difficult to find a parameter that can fit all the images. Among the deep brain tissue segmentations mentioned in Table 1, the amygdala segmentation achieved the lowest segmentation accuracy. This is because the amygdala is relatively small in volume compared to other brain tissues and its registration effect is poor.

In the segmentation test of the deep brain tissues of the IBSR dataset, the multi-atlas registration time of each image of the proposed method was 2.83 hours. The training time of the pixel greyscale probability coefficient is 2.2–2.6 hours. The label fusion time for each image is around 15 seconds.

The limitations and future directions of the proposed method are as follows.

- (1) The parameter setting in the GP-LF model directly affects the quality of the segmentation result. In this study, the same target organization set the same parameter  $\beta$ . In fact, due to the influence of differences between images and registration errors, the fusion results of PSWV and SRLF for the same target tissue segmentation of different images are quite different. Some images have a high fusion value while some images have a low fusion value, so it is difficult to find a suitable parameter  $\beta$  to accommodate all subjects. A good method is to set different parameters  $\beta$  for different target images through a majority of segmentation training and then match the optimal parameters according to the target image information features.
- (2) The optimal  $P(x_j)$  values differ for the same target organization of different images. In this paper, we selected the atlas images located in  $TSP - 3\text{ mm}$  to  $TSP + 3\text{ mm}$  for training to obtain adaptive  $P(x_j)$ , although this  $P(x_j)$  is not the optimal  $P(x_j)$  for each image segmentation.
- (3) The selection of atlas images and patches has a great impact on the segmentation results of the patch-based label fusion method. In the test, not all selected atlas images and patches were suitable; there were still atlas and patches that were not similar to the target image or to  $Tp_{xp}$ , which affected the segmentation accuracy. Consequently, establishing accurate atlas and patch selection models will not only reduce segmentation time but also improve segmentation accuracy.
- (4) The pixels with the same grey value at the tissue boundary may belong to different tissues, which makes it difficult to classify these pixels accurately. The supervoxel-based multi-atlas segmentation method<sup>25</sup> is suitable for addressing this problem and can improve the segmentation results.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 16 July 2019; Accepted: 13 November 2019;

Published online: 29 November 2019

## References

1. Ivana, D., Bart, G. & Wilfried, P. Mri segmentation of the human brain: challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*. 1–23(2015).
2. Sandra, G. *et al.* A review on brain structures segmentation in magnetic resonance imaging. *Artificial Intelligence in Medicine*. **73**, 45–69 (2016).
3. Fischl, B. Freesurfer. *Neuroimage*. **62**, 774–781 (2012).
4. Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*. **56**, 907–922 (2011).
5. Babalola, K. *et al.* An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage*. **47**, 1435–1447 (2009).
6. Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D. & Hammers, A. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *Neuroimage*. **33**, 115–126 (2006).
7. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J. & Rueckert, D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage*. **46**, 726–738 (2009).
8. Wang, M. & Li, P. A Review of Deformation Models in Medical Image Registration. *Journal of Medical and Biological Engineering*. **39**, 1–17 (2019).
9. Alven, J., Norlen, A., Enqvist, O. & Kahl, F. Uberatlas: fast and robust registration for multi-atlas segmentation. *Pattern Recognition Letters*. **80**, 249–255 (2016).
10. Alchatzidis, S., Sotiras, A., Zacharaki, E. I. & Paragios, N. A discrete MRF framework for integrated multi-atlas registration and segmentation. *International Journal of Computer Vision*. **121**, 169–181 (2017).
11. Isgum, I. *et al.* Multi-atlas-based segmentation with local decision fusion – application to cardiac and aortic segmentation in CT scans. *IEEE Trans Med Imaging*. **28**, 1000–1010 (2009).
12. Sabuncu, M. R., Yeo, B. T., Leemput, K. V., Fischl, B. & Golland, P. A generative model for image segmentation based on label fusion. *IEEE Trans Med Imaging*. **29**, 1714–1729 (2010).

13. Nie, J. & Shen, D. Automated segmentation of mouse brain images using multi-atlas multi-ROI deformation and label fusion. *Neuroinformatics* **11**, 35–45 (2013).
14. Coupe, P. *et al.* Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* **54**, 940–954 (2011).
15. Wang, M., Li, P. & Liu, F. Multi-atlas active contour segmentation method using template optimization algorithm. *BMC medical imaging*. **19** (2019).
16. Lin, X. B., Li, X. X. & Guo, D. M. Registration Error and Intensity Similarity Based Label Fusion for Segmentation. *IRBM* **40**, 78–85 (2019).
17. Tang, Z., Sahar, A., Yap, P. T. & Shen, D. Multi-atlas segmentation of MR tumor brain images using low-rank based image recovery. *IEEE Trans. Med. Imag.* **37**, 2224–2235 (2018).
18. Roy, S. *et al.* Subject-specific sparse dictionary learning for atlas-based brain MRI segmentation. *IEEE J. Biomed. Health Inform.* **19**, 1598–1609 (2015).
19. Tong, T., Wolz, R., Coupe, P., Hajnal, J. V. & Rueckert, D. Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling. *NeuroImage*. **76**, 11–23 (2013).
20. Lee, J., Kim, S. J., Chen, R. & Herskovits, E. H. Brain tumor image segmentation using kernel dictionary learning. In *Proc. 37th Annu. Int. Conf. IEEE EMBC, Milan, Italy, Aug. 658–661*(2015).
21. Liu, Y., Wei, Y. & Wang, C. Subcortical Brain Segmentation Based on Atlas Registration and Linearized Kernel Sparse Representative Classifier. *IEEE ACCESS*. **7**, 31547–31557 (2019).
22. Bai, W., Shi, W., Ledig, C. & Rueckert, D. Multi-atlas segmentation with augmented features for cardiac mr images. *Medical image analysis*. **19**, 98–109 (2015).
23. Zikic, D., Glocker, B. & Criminisi, A. Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. *Medical image analysis*. **18**, 1262–1273 (2014).
24. Moeskops, P. *et al.* Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*. **35**, 1252–1261 (2016).
25. Huo, J. *et al.* Supervoxel based method for multi-atlas segmentation of brain MR images. *NeuroImage*. **175**, 201–214 (2018).
26. Xu, L. *et al.* Automatic labeling of mr brain images through extensible learning and atlas forests. *Medical Physics*. **44**, 6329–6340 (2017).
27. Kaiser, K. *et al.* Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. *Medical Image Analysis*. **48**, 177–186 (2018).
28. Kennedy, D. N. *et al.* CANDIShare: a resource for pediatric neuroimaging data. *Neuroinformatics* **10**, 319–22 (2012).
29. About Te Creative Commons Licenses, <http://creativecommons.org/about/licenses> (2019).
30. Lemieux, L., Jagoe, R., Fish, D. R., Kitchen, N. D. & Thomas, D. G. T. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Medical Physics*. **21**, 1749–1760 (1994).
31. Zaffino, P., Ciardo, D., Raudaschl, P., Fritscher, K. & Spadea, M. F. Multi atlas based segmentation: should we prefer the best atlas group over the group of best atlases? *Physics in Medicine and Biology*. **63**, 12NT01 (2018).

## Acknowledgements

We thank to the Center for Morphometric Analysis Internet Brain Segmentation Repository for providing image data. This research was supported by the program for National Natural Science Foundation of China (No. 61572159), NCET (NCET-13-0756) and Distinguished Young Scientists Funds of Heilongjiang Province (JC201302). The funding body didn't have any influence on the collection, analysis, and interpretation of data or on writing the manuscript.

## Author contributions

M.W. conceived the research. P.L. performed the experiments and analyses and wrote the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019