

# SCIENTIFIC REPORTS



OPEN

## Automatic Thalamus Segmentation from Magnetic Resonance Images Using Multiple Atlases Level Set Framework (MALSF)

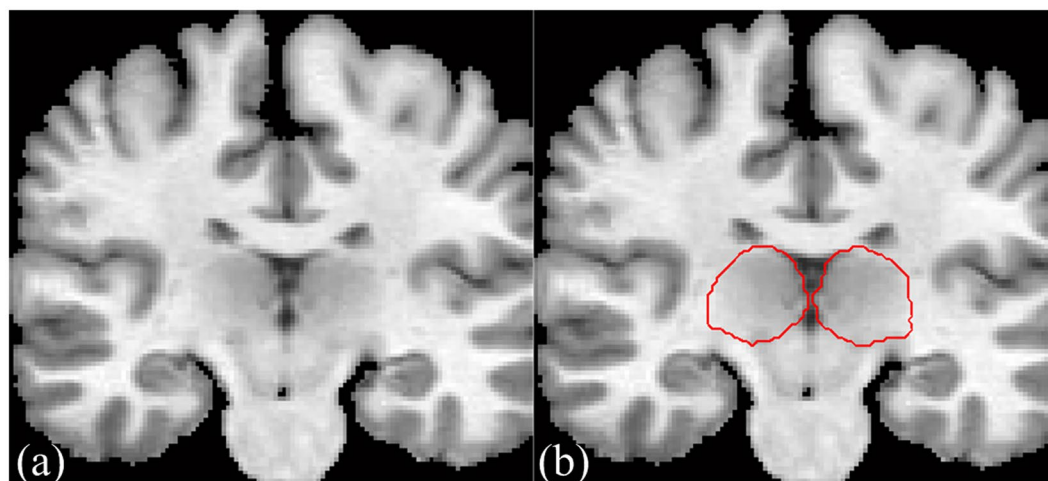
Minghui Zhang, Zhentai Lu, Qianjin Feng & Yu Zhang

In this paper, we present an original multiple atlases level set framework (MALSF) for automatic, accurate and robust thalamus segmentation in magnetic resonance images (MRI). The contributions of the MALSF method are twofold. First, the main technical contribution is a novel label fusion strategy in the level set framework. Label fusion is achieved by seeking an optimal level set function that minimizes energy functional with three terms: label fusion term, image based term, and regularization term. This strategy integrates shape prior, image information and the regularity of the thalamus. Second, we use propagated labels from multiple registration methods with different parameters to take full advantage of the complementary information of different registration methods. Since different registration methods and different atlases can yield complementary information, multiple registration and multiple atlases can be incorporated into the level set framework to improve the segmentation performance. Experiments have shown that the MALSF method can improve the segmentation accuracy for the thalamus. Compared to ground truth segmentation, the mean Dice metrics of our method are 0.9239 and 0.9200 for left and right thalamus.

The thalamus plays a critical role in human brain function. It is the largest, most internal structure of the diencephalon. Segmentation and characterization of the thalamus from brain magnetic resonance images (MRI) are expected to help noninvasive diagnosis and treatment. Thalamus segmentation has become more and more important for a wide range of clinical and research applications. For example, the thalamus changes in terms of volume and intensity are involved in a large number of diseases, such as Parkinson's disease, multiple sclerosis and schizophrenia. Expert manual segmentation of thalamus from MRI data is still the gold standard, but it is labor intensive, and the segmentation results are not reproducible. The thalamus has no clearly contrasted boundaries in MR T1-weighted images, this issue is illustrated by Fig. 1, where no obvious thalamus boundary can be observed. It is a difficult challenge even for an expert radiologist.

Because of low image contrast, noise, and missing or diffuse boundaries, some discrete methods such as thresholding, FCM<sup>1</sup>, and region growing<sup>2</sup> are not reliable because they only use image intensity information. Erhard *et al.* compared the performances of publicly available segmentation tools (volBrain, FSL, FreeSurfer and SPM) and their impact on diffusion quantification, emphasizing the importance of using recently developed segmentation algorithms and imaging techniques. They found that volBrain is superior in thalamus and hippocampus segmentation compared to FSL, FreeSurfer and SPM<sup>3</sup>. Shape is important information for brain structure segmentation. It is preferable to combine the intensity information along with the shape prior information. Active contours model (ACM) is a very popular technique for image segmentation which can combine shape prior information. According to the features of the involved image, ACM can be categorized as the image intensities statistical information (region-based ACM)<sup>4-8</sup> or the image gradients (edge-based ACM)<sup>9-12</sup>. ACM needs precise initialization of the starting contours. It can not converge to the target structures unless the initialization is good enough. Another concern is that most of the boundaries of the thalamus are diffuse or missing, and thus it is difficult to segment the thalamus, even by a radiology expert. The performance of ACM may be adversely affected when these methods are employed. Shiyuan Hu *et al.* combined levelset modeling and appearance

Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, 510515, China. Minghui Zhang and Zhentai Lu contributed equally to this work. Correspondence and requests for materials should be addressed to Z.L. (email: [luzhentai@163.com](mailto:luzhentai@163.com))



**Figure 1.** MR T1-weighted image (a) an image slice obtained from a 3-D thalamus MR image. (b) The manually-delineated thalamus boundary shown by the red contour, superimposed on the original image in (a).

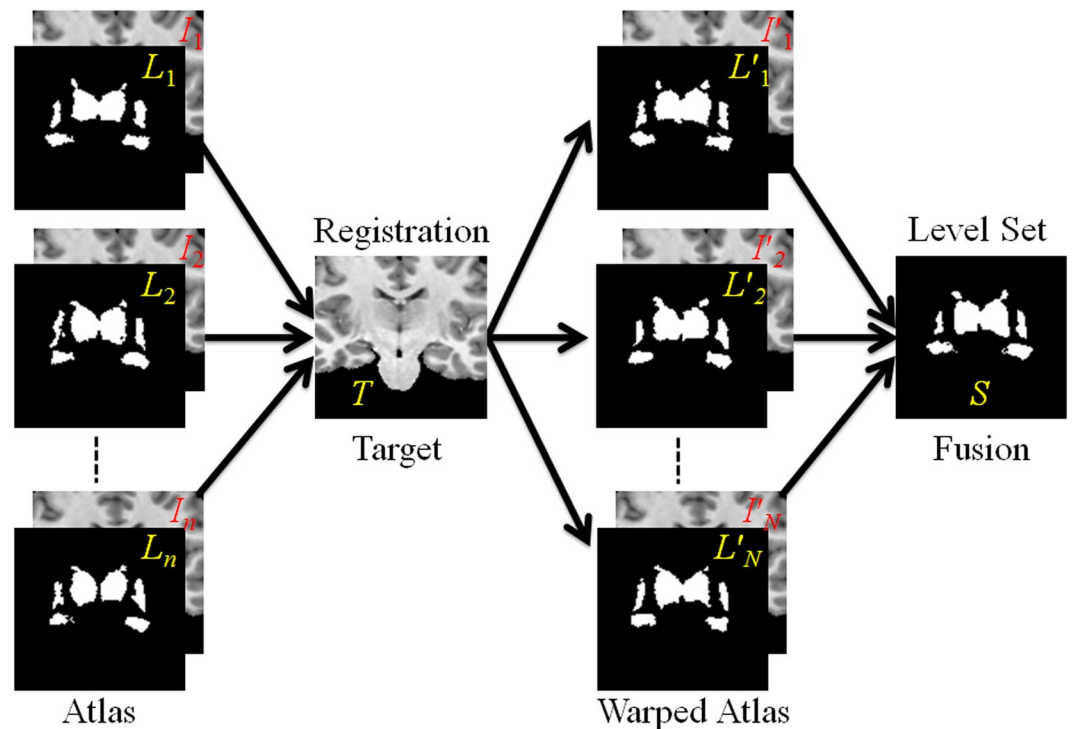
modeling to identify the hippocampus and amygdala using multi-contrast MR imaging. The linear model for the gray-intensity of each contrast MR image can be constructed by applying principle component analysis (PCA) to the pre-processed gray images. PCA-based techniques require that the structures to be segmented are pre-aligned well<sup>13</sup>. Dinggang Shen *et al.* presented a learning-based algorithm by taking advantage of the multi-atlas framework and the auto-context model. Under the multi-atlas framework, auto-context model based classifiers are trained for all atlases to incorporate anatomical variability. It is time-consuming to extract image appearance features, texture features and context features. A sequence classifier is trained in each atlas space by borrowing the training samples from not only the underlying atlas but also all other linearly-aligned atlases, which has high computational complexity<sup>14</sup>.

Another popular methodology is atlas-based methods<sup>15–17</sup>. In these methods, the atlas includes a gray level image and a manually labelled image associated with the training image. The training gray level image is registered to the target image. The manually labelled images are thus propagated to the target image by using the deformation fields determined by registration method. The segmentation is accomplished with the help of image registration. The sophisticated non-rigid registration techniques can be used in a meticulous way, in which the shape prior and spatial information can be incorporated to help segmentation, so the atlas-based segmentation can be highly accurate.

Compared to other generic segmentation techniques, the atlas-based approach has several major advantages. For example, only a registration method and a number of pre-segmented data sets are required. There is no need for complex training procedures. The quality of the atlas-based approach is limited by the accuracy of the registration procedure and the anatomical similarity between the labelled and target image. Rather than depending on a single atlas, an alternative strategy is to register multiple atlas images to the target image separately<sup>15</sup>. These pairwise transformations are then used to propagate the labels to the target image. The final segmentation is achieved by fusing the propagated labels. A number of label fusion strategies have been proposed for multi-atlas based segmentation in the literature. Among them, the majority voting (MV) method is probably the simplest one and has been widely used in medical image segmentation<sup>17</sup>. In the MV method, the weights of candidate segmentations from each atlas are equal. The label with largest agreement from all atlases is assigned as the final label. A natural extension of the MV method is to use adaptive weighted averaging. Another popular approach is simultaneous truth and performance level estimation (STAPLE), which uses the expectation-maximization (EM) algorithm to achieve the best possible final segmentation<sup>18</sup>.

This multi-atlas based segmentation approach reduces the effect of errors associated with individually propagated atlases. A registration error for a particular propagated atlas is less likely to affect the final segmentation when combined with other atlases. Multi-atlas based segmentation methods have been proven to be effective and accurate in comparison with single atlas-based approaches. The quality of the registration algorithm and the selection of the atlas will directly affect the segmentation performance. The main limitations of the multi-atlas segmentation methods are that they often lead to a compromise between the accuracy of the registration and the smoothness of the deformation, and that those methods are time consuming.

To combine the advantages of the methods described above and overcome some of their disadvantages, we propose a new multiple atlas segmentation strategy using different registration methods and a level set fusion framework. The main contributions of this paper include the following. (1) The proposed method employs level set model further to improve the segmentation performance, which is able to provide smooth and closed contours in the final segmentation. (2) We explore the effect of the quality of registration method on segmentation accuracy. Multiple registration methods, i.e., ANTS<sup>19</sup>, DRAMMS<sup>20</sup> and Demons<sup>5</sup> are incorporated into the segmentation procedure, which facilitates combining the advantages of the different registration approaches. (3) The local similarity maps that calculated in the registration procedure are used as local weight of the level set function. This



**Figure 2.** Schematic illustration of our segmentation method.

strategy does not add a substantial computational burden to the algorithm, since the atlases are registered to the target image offline. (4) The algorithm combines local intensity information, atlas prior, and shape regularity, it is very robust and able to accurately segment a new subject.

## Method

The proposed new segmentation method has two stages, multiple atlas registration (blue rectangle) and active contour model label fusion (red rectangle). It is illustrated schematically in Fig. 2. The gray level images  $I_i$  form the atlas is registered to the target image  $T$  by using ANTS, DRAMMS and Demons. The resulting transformations are used to warp the corresponding atlas label  $L_i$ . The propagated label  $L'_i$  are then combined to create the final segmentation  $S$  of the target image by using the level set method.

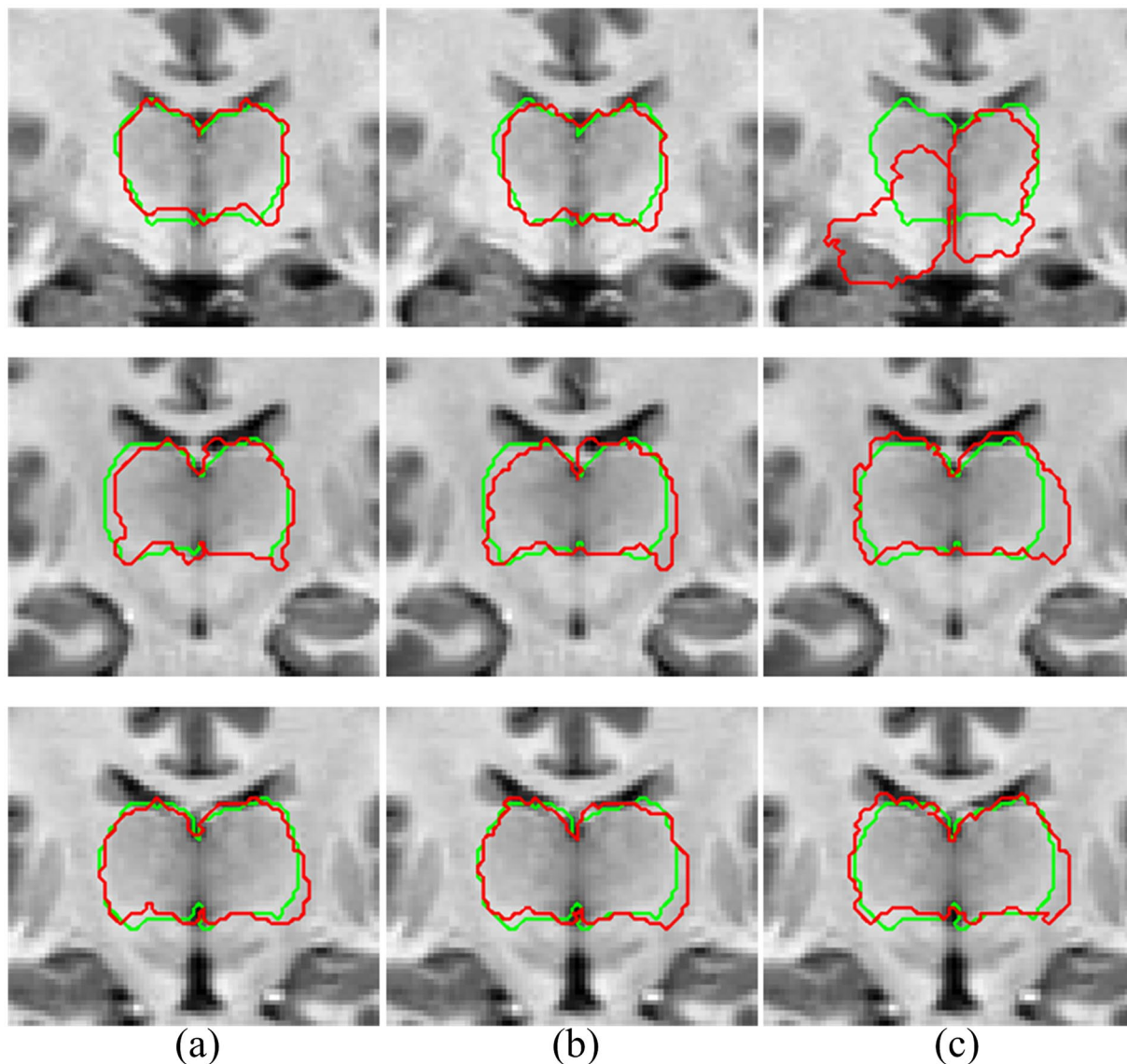
**Ethics Statements.** The original MRI scans were obtained from the Open Access Series of Imaging Studies (OASIS) project web site (<http://www.oasis-brains.org/>). OASIS is a project aimed at making MRI data sets of the brain freely available to the scientific community. Patient's records and information were anonymized and de-identified prior to analysis. A subset of this data was used as the “gold standard” for the MICCAI 2013 Grand Challenge and Workshop on Multi-Atlas Labeling. Our study was approved by the ethics committee of Washington University Alzheimer's Disease Research Center.

**Multiple Atlas Registration.** Sources of error in atlas-based segmentations include registration error and manually delineated errors. The more accurately the registration warps the atlas onto the target image, the more accurate the result of the segmentation. The registration errors produced by using different atlases and different registration methods are not identical, and thus employing multiple atlases and multiple registration methods can effectively reduce the registration errors. To improve the segmentation performance of multiple atlas-based methods, we utilize many of the existing complementary registration strategies, i.e., ANTS<sup>19</sup>, and DRAMMS<sup>20</sup>.

Elastic registration methods such as statistical parametric mapping (SPM)<sup>21</sup>, free-form deformations (FFD)<sup>22</sup>, and Demons operate<sup>23</sup> in the space of vector fields, which do not preserve topology. Not applying ad hoc constraints will result in the brain topology changing in an uncontrolled way. ANTS uses diffeomorphic transformations to warp images, and combines multiple different similarity measures that are optimized in parallel, which provide well-behaved solutions with mathematical guarantees about distance in deformation space and regularity.

DRAMMS bridges the gap between the traditional voxel-wise methods and landmark/feature-based methods. It renders each voxel relatively distinctively identifiable by a rich set of attributes to reduce match ambiguities. Additionally, DRAMMS modulates the registration by assigning higher weights to those voxels which have a higher ability to establish unique correspondences across images, therefore reducing the negative impact of those regions that are less capable of finding correspondences.

As shown in Fig. 3, ANTS and DRAMMS obviously outperform Demons. The performances of ANTS and DRAMMS are very similar, and it is difficult to evaluate which one is better, even by experts. To avoid possible registration error and to take full advantage of the merit of different registration methods, we combine the ANTS and DRAMMS registration results in a level set framework.



**Figure 3.** Registration performance comparison between ANTS, DRAMMS and Demons. (a) ANTS, (b) DRAMMS, (c) Demons. Each row represents different patient. The red contours in each column represent the warped thalamus boundary after ANTS, DRAMMS and Demons registration, respectively. The green contours are the ground truth that manually delineated by experts.

**Level Set Label Fusion.** We use the level set method<sup>8</sup> to model the structures of thalamus in the training data. First, the level set function (LSF) is utilized to represent the shape of manually segmented labels. Figure 4 shows the shape of the thalamus (showed in Fig. 1) by using the level set method, where the manually segmented thalamus is described by the LSF, and its boundary is captured by its zero level-set.

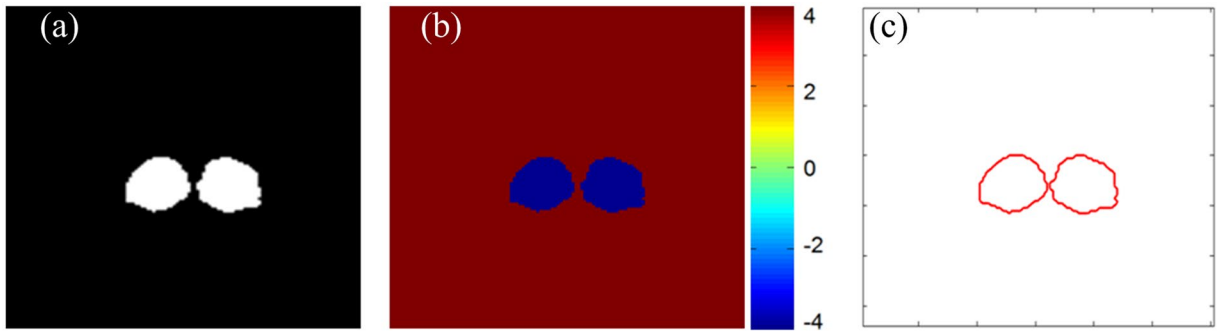
For a training data set with  $n$  labels, as we use ANTS and DRAMMS to register the atlases to target image, there are  $N = 2n$  separate LSF  $\{f_1, \dots, f_N\}$  where negative distances assigned to the inside and positive distances to the outside of object, are adopted to describe the boundaries of atlas shapes. The zero level contour of LSF  $\phi$  is the boundary of the thalamus in the target image, denoted by  $C$ . Our purpose is to find an optimal LSF  $\phi$  by using a label fusion technique under the proposed variational framework.

We propose a new and general variational level set formulation for label fusion, in which the energy  $E$  of LSF  $\phi$  is defined in a general form:

$$E(\phi) = \alpha F(\phi; \phi_1, \dots, \phi_N) + \beta D(\phi; I) + \gamma R(\phi) \quad (1)$$

where  $F(\phi; \phi_1, \dots, \phi_N)$  is the label fusion term, which can integrate the shape priors of different atlases;  $D(\phi; I)$  is the image based term with a given target image  $I$ , which will combine the regional intensity information of the target image; and  $R(\phi)$  is the regularization term that will smooth the shape of the thalamus in the level set evolution procedure.  $\alpha$ ,  $\beta$  and  $\gamma$  are the corresponding coefficients.

The label fusion term  $F(\phi; \phi_1, \dots, \phi_N)$  is defined by



**Figure 4.** 2D slice of an example of the shape modeling using the level-set method. (a) Manually segmented shape, (b) Level set function (LSF), (c) Zeros level set.

$$F(\phi; \phi_1, \dots, \phi_N) = \int_{\Omega} \sum_{i=1}^N \omega_i(x) \left| \phi(x) - \phi_i(x) \right|^2 dx \quad (2)$$

where  $\omega_i(x)$  is a local weight assigned to the  $i$ -th label represented by the level set function  $\phi_i$ , with  $\sum_{i=1}^N \omega_i(x) = 1$ ,  $x$  indexes the image pixels, and  $\Omega$  denotes the target image domain. Both ANTS and DRAMMS calculate the local similarity of each voxel in the registration procedure, where we use the local similarity as the local weight of the level set function. This strategy does not add computational burden to the algorithm, since the atlases are registered to the target image offline.

Minimizing label fusion energy  $F$ , the zero LSF  $\phi$  is forced to be close to the zero LSF  $\phi_1, \dots, \phi_N$ , which are the boundaries of the thalamus given by the warped labels  $L'_1, L'_2, \dots, L'_N$ . However, minimizing the energy  $F$  alone may result in an irregular shape of the contour obtained. Therefore, we need a regularization mechanism to maintain the regularity of the contour. We use the arc length of the zero level contour of the LSF  $\phi$  as the regularization.

$$R(\phi) = \int |\nabla H(\phi(x))| dx \quad (3)$$

This regularization term ensures the regularity of the boundary of the thalamus.

The traditional fusion method only combines the warped labels, without taking into account the target image information. These methods cannot correct the registration error and manually delineated errors effectively. We use the region-scalable fitting (RSF) model<sup>24</sup> as the image term, which is able to handle intensity variations within the foreground and background of the image due to its localization property. This image term can guide the motion of the contour toward the thalamus boundaries, if there are some registration errors and manual labelling errors.

Given a point  $y$ , consider its neighborhood  $O_y = \{x: |x - y| \leq \rho\}$ , which is divided by an object boundary  $C$  into two parts:  $O_y \cap \text{inside}(C)$  and  $O_y \cap \text{outside}(C)$ . The image intensities  $I(x)$  in  $O_y \cap \text{inside}(C)$  and  $O_y \cap \text{outside}(C)$  can be approximated by the two constants  $f_1(y)$  and  $f_2(y)$ , respectively. We seek an optimal contour  $C$  and fitting functions  $f_1$  and  $f_2$  such that the following energy

$$E_y(C, f_1(y), f_2(y); I) = \lambda_1 \int_{O_y \cap \text{inside}(C)} |I(x) - f_1(y)|^2 dx + \lambda_2 \int_{O_y \cap \text{outside}(C)} |I(x) - f_2(y)|^2 dx \quad (4)$$

is minimized for all  $y \in \Omega$  for a given target image  $I$ , where  $\lambda_1$  and  $\lambda_2$  are the weighting coefficients. Minimization of  $E_y(C, f_1(y), f_2(y); I)$  for all  $y$  can be achieved by minimizing the integration of  $E_y$  with respect to  $y$ , i.e., the energy  $D$  is defined by

$$D(C, f_1, f_2; I) = \int_{\Omega} E_y(C, f_1(y), f_2(y); I) dy \quad (5)$$

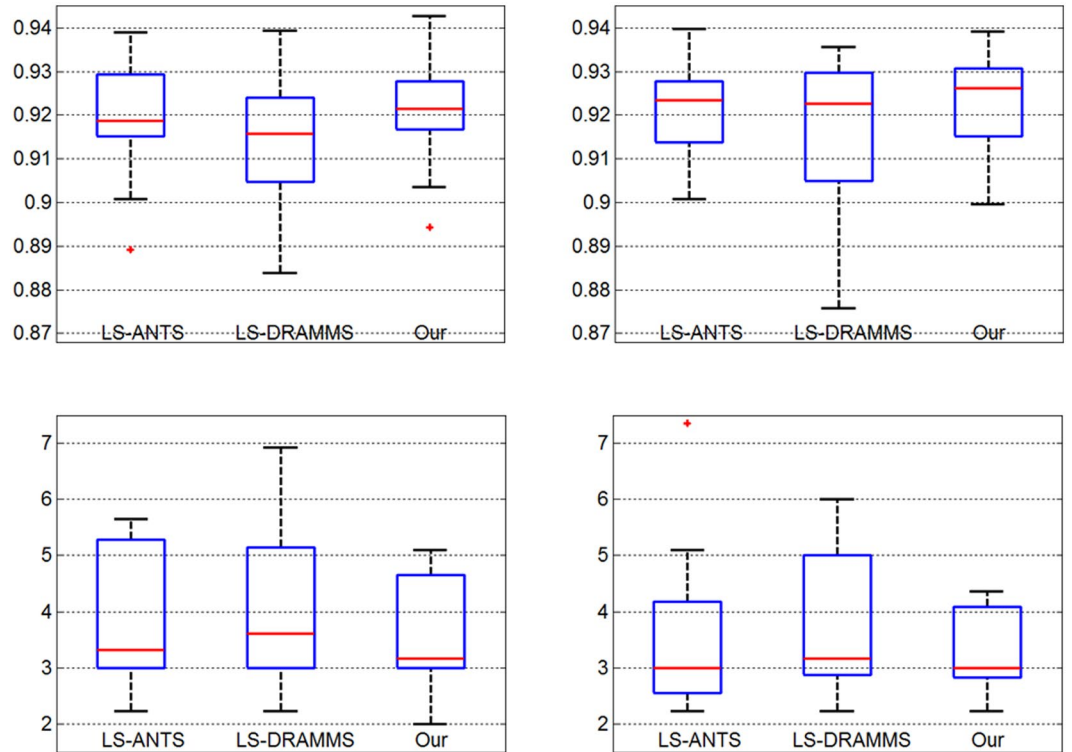
The energy  $D(C, f_1, f_2; I)$  can be expressed in an equivalent form in terms of LSF  $\phi$  and the fitting functions  $f_1$  and  $f_2$  as the following:

$$D(C, f_1, f_2; I) = \sum_{k=1}^2 \int \left( \int K_{\rho}(x - y) |I(x) - f_k(y)|^2 M_k(\phi(x)) dx \right) dy \quad (6)$$

where  $M_1(\phi(x)) = 1 - H(\phi(x))$  and  $M_2(\phi(x)) = H(\phi(x))$ , and where  $H$  is the Heaviside function. We approximate this as follows:

$$H_{\varepsilon}(z) = \frac{1}{2} \left[ 1 + \frac{2}{\pi} \arctan \left( \frac{z}{\varepsilon} \right) \right] \quad (7)$$

where  $K_{\rho}$  is the kernel function. In this paper we chose a Gaussian kernel:



**Figure 5.** Box-plot of performance metric on the MICCAI dataset. The left column is left thalamus and the right column is right thalamus. The top row contains the similarity index, higher values implying greater volumetric overlap with the manual segmentation, and the bottom row contains the Hausdorff distance, lower values implying a surface closer to the manual segmentation.

parameter	$\alpha$	$\beta$	$\gamma$	$\mu$	$\nu$	$\lambda_1$	$\lambda_2$
value	0.1	1	0.1	0.1	0.01	0.0001	0.0001

**Table 1.** The parameters used in all experiments.

$$K_\rho(u) = \frac{1}{(2\pi)^{n/2} \rho^n} e^{-|u|^2/2\rho^2} \tag{8}$$

with a scale parameter  $\rho > 0$ .

The above energy can be further expressed as

$$D(C, f_1, f_2; I) = \sum_{k=1}^2 \int \left( \int K_\rho(x - y) |I(x) - f_i(y)|^2 dy \right) M_i(\phi(x)) dx \tag{9}$$

which defines a specific form of the image based term in the proposed variation framework. The minimization of the energy  $E$  with respect to the LSF can be achieved by using the standard gradient descent method<sup>8</sup>:

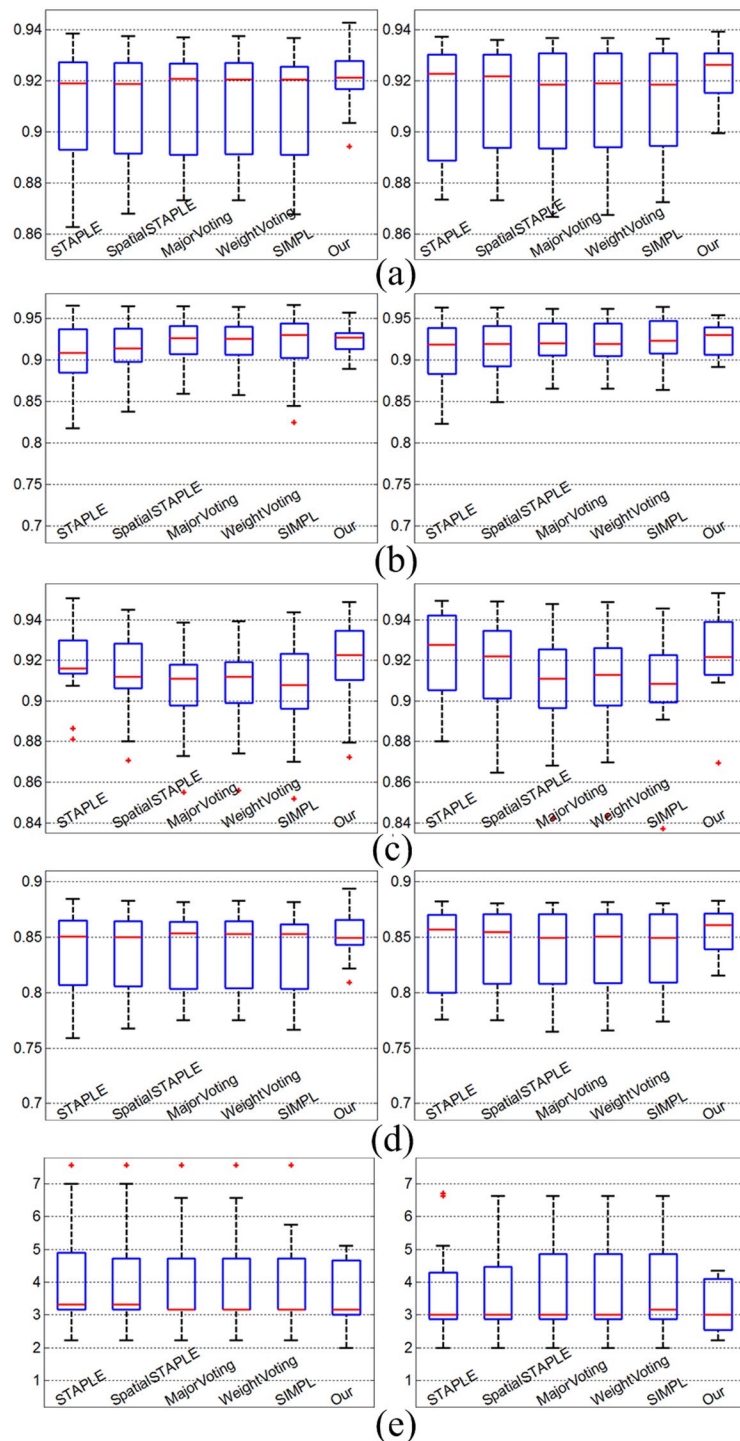
$$\frac{\partial \phi}{\partial t} = -\delta_\varepsilon(\phi)(\lambda_1 e_1 - \lambda_2 e_2) + \nu \delta_\varepsilon(\phi) \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) + \mu \left( \nabla^2 \phi - \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right) \tag{10}$$

where  $\delta_\varepsilon$  is the smoothed Dirac delta function, and  $e_1$  and  $e_2$  are the functions

$$e_i(x) = \int K_\rho(x - y) |I(x) - f_i(y)|^2 dy, \quad i = 1, 2 \tag{11}$$

### Parameter optimization

The parameter values of our proposed variational framework were chosen based on the 20 labeled images that were used for training. This was done by performing leave-one-out parameter-tuning experiments on those 20 images. For example, consider the case where image 1 was used as a target, and images 2 to 20 were used for training. To determine the optimal values of  $\alpha, \beta, \gamma, \mu, \nu, \lambda_1,$  and  $\lambda_2,$  20 parameter-tuning experiments were conducted.



**Figure 6.** Box-plot of performance metric on the MICCAI dataset. The left column is left thalamus and the right column is right thalamus. (a) Similarity index ( $SI$ ), higher values implying greater volumetric overlap with the manual segmentation, (b) Precision ( $P$ ), (c) Recall ( $R$ ), (d) Relative Overlap ( $RO$ ) and (e) Hausdorff Distances ( $HD$ ), lower values implying a surface closer to the manual segmentation.

First, the manual labels of images 2–20 were warped to the coordinate framework of image 1, and segmentation was performed using certain values of the parameters. The manual labels of image 1 were then used as the gold standard to compute the similarity index ( $SI$ ) for these parameters. This was repeated for several values of  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\mu$ ,  $\nu$ ,  $\lambda_1$ , and  $\lambda_2$ . Similar parameter-tuning experiments were then performed for images 2–20, each yielding a list of  $SI$ -parameter combinations. The parameters that gave the highest  $SI$  averaged over all parameter-tuning experiments were chosen. We fixed the parameters for all experiments in our proposed method, as listed in Table 1.

Similarity Index (mean $\pm$ standard deviation)		
Method	Left Thalamus	Right Thalamus
STAPLE	0.9106 $\pm$ 0.0224	0.9132 $\pm$ 0.0229
Spatial STAPLE	0.9109 $\pm$ 0.0218	0.9136 $\pm$ 0.0222
Major Voting	0.9118 $\pm$ 0.0205	0.9126 $\pm$ 0.0239
Weight Voting	0.9119 $\pm$ 0.0204	0.9128 $\pm$ 0.0237
SIMPLE	0.9100 $\pm$ 0.0225	0.9141 $\pm$ 0.0210
Our	0.9200 $\pm$ 0.0124	0.9239 $\pm$ 0.0100

**Table 2.** The Similarity Index (*SI*) on the MICCAI dataset (mean  $\pm$  standard deviation) for the various segmentation methods as compared with manual segmentations.

## Experiments results

In this section, several experiments were conducted to evaluate the proposed method. Since our method aims to improve segmentation accuracy in two ways, by incorporating multiple registration method into the conventional multiple atlas-based methods and by taking the advantage of level set framework, it is worthwhile to evaluate the contributions from these two different methods. First, we compared the proposed multiple registration method to label fusion using a single registration method. Second, we compared the proposed method to five state-of-the-art automatic segmentation methods: STAPLE<sup>25</sup>, Spatial STAPLE<sup>26</sup>, Major Voting<sup>27</sup>, Weight Voting<sup>28</sup>, and SIMPLE<sup>29</sup>. We use MASI Label Fusion software tools (<http://www.nitrc.org/projects/masi-fusion>). There are no parameters in Major Voting and Weight Voting, so we obtain the fusion results directly. For STAPLE, Spatial STAPLE, and SIMPLE, we use leave-one-out parameter-tuning experiments to get the best results, and then compare these with our proposed method. The Matlab source code for the Multiple Atlases Level Set Framework (MALSF) can be downloaded at <https://github.com/luzhentai/MALSF2.0>.

## Data and Experiments set

Thirty-five real brain MRI images were downloaded from <https://masi.vuse.vanderbilt.edu/>. This web site provides the online continuation of a segmentation contest held at the 2013 Medical Image Computing and Computer Assisted Intervention Challenge (MICCAI). The dataset consists of a de-faced T1-weighted structural MRI and an associated manually labeled volume with one label per voxel. Each volume (MRI and label) will be stored in a separate 3D NiFTI file. The dimensions are  $256 \times 256 \times 287$ , and the voxel dimensions are  $1 \times 1 \times 1$  mm. We randomly select fifteen images as targets and twenty images as training images.

All MR volumes are corrected for intensity inhomogeneity using the N4 bias field correction algorithm<sup>30</sup>. The intensities are then linearly normalized to the range of 0 to 255. The formula for the normalization is:  $\frac{I - I_{\min}}{I_{\max} - I_{\min}} \times 255$ , where  $I_{\min}$  and  $I_{\max}$  are the minimum and maximum intensity values in the volume, respectively.

To reduce computational burden, the segmentation is estimated only within the region of interest (ROI) surrounding the left and right thalamus. We use a very fast and simple approach that utilizes the union of all expert labels in the training database as the ROI. In this way, we ensure that the structure is completely included in the ROI.

## Quantitative evaluation of thalamus segmentation

To quantify the quality of our automatic segmentations, as well as of other state-of-the-art segmentation methods, we employed four widely used metrics, similarity index (*SI*), recall (*R*), relative overlap (*RO*), and precision (*P*), to measure the volumetric overlap of automatic segmentation with respect to the manual labels (i.e., ground truth). Moreover, we also measure the surface distance based on the Hausdorff distance (*HD*)<sup>28</sup>:

$$SI = \frac{2V(A \cap B)}{V(A) + V(B)} \quad P = \frac{V(A \cap B)}{V(B)} \quad R = \frac{V(A \cap B)}{V(A)} \quad RO = \frac{V(A \cap B)}{V(A \cup B)} \quad HD = \max(H_1, H_2) \quad (12)$$

where  $V(A)$  denotes the volume of the ground-truth segmentation,  $V(B)$  denotes the volume of automatic segmentation,  $H_1 = \max_{a \in A}(\min_{b \in B} d(a, b))$ ,  $H_2 = \max_{b \in B}(\min_{a \in A} d(a, b))$ , and  $d(a, b)$  is the Euclidean distance between two points  $a$  and  $b$ .

**Effect of using multiple registration method.** We performed segmentation on the MICCAI data using the following method:

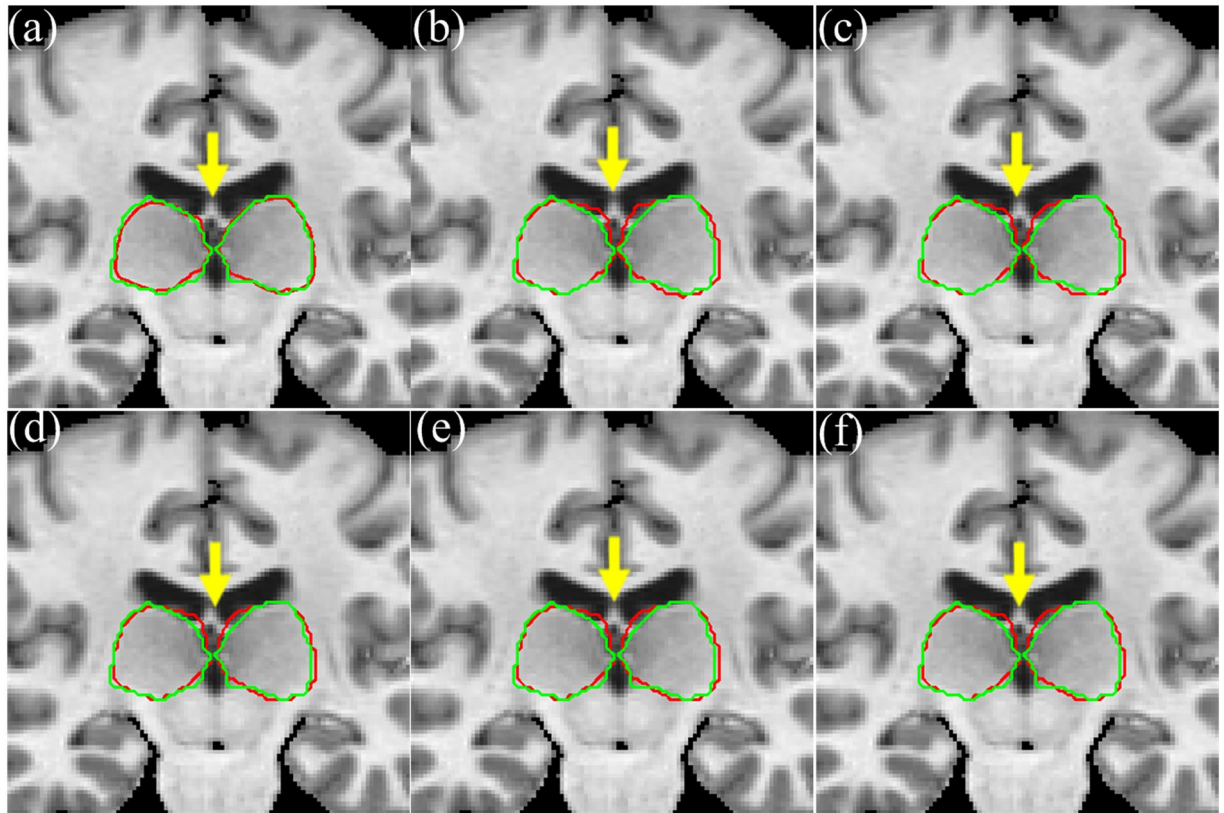
*LS-ANTS*: our level set based fusion strategy that uses a single ANTS registration method

*LS-DRAMMS*: our level set based fusion strategy that uses a single DRAMMS registration method

*Our*: our level set based fusion strategy that uses multiple registration methods (ANTS and DRAMMS)

Figure 5 shows summary box-plots of the similarity index (*SI*) and Hausdorff distance (*HD*) versus the manual delineate for each segmentation method on the MICCAI dataset. In top row, the red line is the median *SI* of fifteen test images, with higher values implying greater volumetric overlap with the manual segmentation. In bottom row, the red line is median Hausdorff distance, with lower values implying a surface closer to the manual segmentation. The results show that our method using multiple registrations methods is better than the one using a single





**Figure 7.** The segmentation results of MICCAI image No. 1006 by our method and five other methods. (a) Our method, (b) STAPLE, (c) Spatial STAPLE, (d) Major Voting, (e) Weight Voting, (f) SIMPLE. Automatic segmentation results are illustrated in red contour, the manually-delineated thalamus boundary shown by the green contour.

registration method. The use of the multiple registrations captures greater anatomical variability and improves robustness against occasional registration failures.

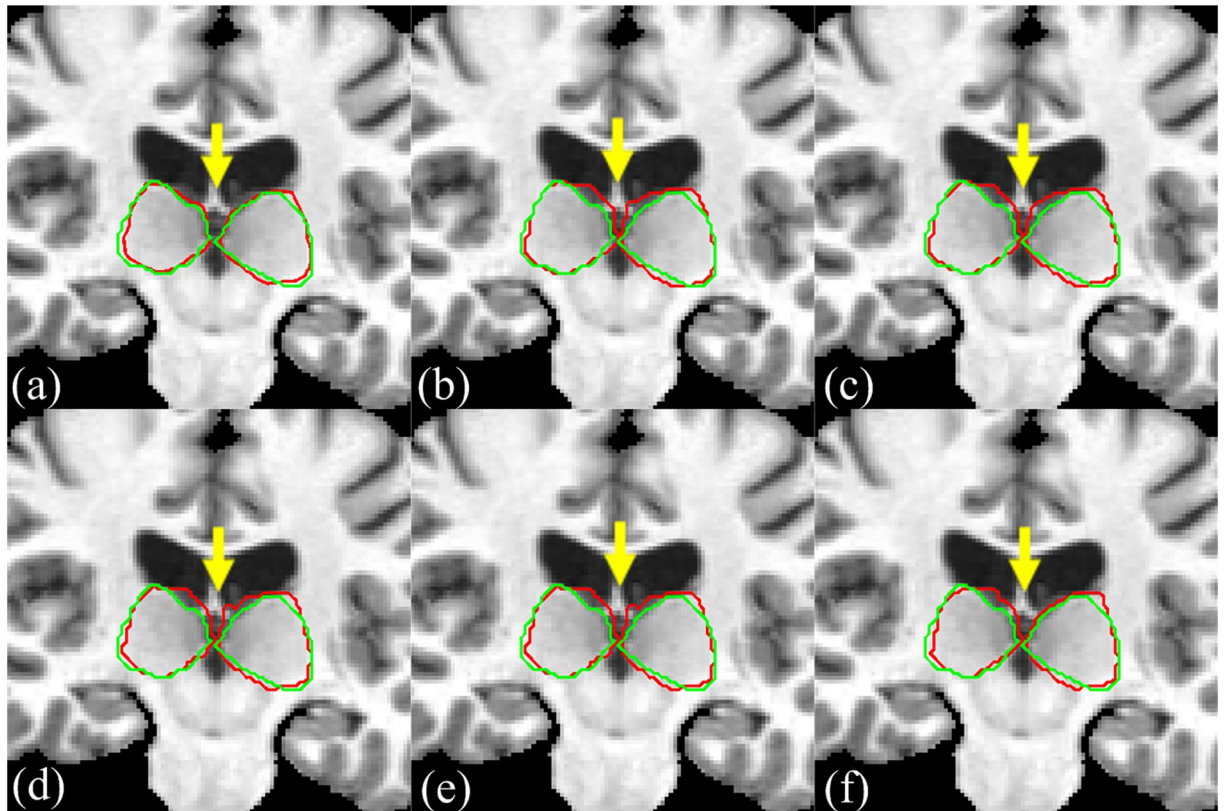
**Effect of using the level set fusion method.** To compare our results with those of recent automatic segmentation methods reported in the literature, we performed segmentation on the MICCAI data using several state-of-the-art fusion algorithms: STAPLE<sup>25</sup>, Spatial STAPLE<sup>26</sup>, Major Voting<sup>27</sup>, Weight Voting<sup>28</sup>, and SIMPLE<sup>29</sup>.

Table 2 shows the mean values and standard deviations of the similarity index for the MICCAI data. The mean similarity index of our segmentation is 0.9200 for the left thalamus and 0.9239 for the right thalamus. The proposed method clearly leads to a higher mean similarity index compared to the five other methods, which provide similar overlap scores separately. The volumes measured by our method are closer to the manual expert segmentations than to the volumes measured by the five other methods. The results also show that there is no significant volume difference between segmented left and right thalamus.

Figure 6 shows summary box-plots of the similarity index ( $SI$ ), precision ( $P$ ), recall ( $R$ ), relative overlap ( $RO$ ), and Hausdorff distance ( $HD$ ) versus the manual delineate for each segmentation method on the MICCAI dataset. The results show that our method using the level set fusion strategy is better than the state-of-the-art fusion algorithms.

For visual comparison, we show in Fig. 7 the segmentation results of MICCAI image No.1006 by our method and five other methods, along with the manual segmentation. It can be observed that the thalamus segmented by our method is more similar to the ground-truth than any other method, e.g., STAPLE, Spatial STAPLE, Major Voting, Weight Voting, and SIMPLE. Particularly, in the region of the yellow arrow, the five other methods cannot locate the left and right thalamus. Our method can correct the registration errors, because the region-scalable fitting (RSF) model in the level set fusion framework can guide the motion of the contour toward the thalamus boundaries.

We also show in Fig. 8 the segmentation results of MICCAI image No.1025 by our method and five other methods along with the manual segmentation. It can be clearly seen that the proposed method can accurately delineate the thalamus boundaries, especially for the regions near the lateral ventricle. The SIMPLE method can correct the registration errors in left thalamus as shown by the blue arrow, but it cannot correct the errors in right thalamus as shown by the yellow arrow. In addition to showing the two-dimensional slice of segmentation results, we present a three-dimensional visual inspection of the experimental results. In Fig. 9, we show the results of STAPLE, Spatial STAPLE, Major Voting, Weight Voting, SIMPLE and the proposed method. The five other



**Figure 8.** The segmentation results of MICCAI image No. 1025 by our method and five other methods. (a) Our method, (b) STAPLE, (c) Spatial STAPLE, (d) Major Voting, (e) Weight Voting, (f) SIMPLE. Automatic segmentation results are illustrated in red contour, the manually-delineated thalamus boundary shown by the green contour.

methods are able to perform reasonable segmentations, but our method is able to provide smooth and closed contours on the final segmentation.

## Discussion

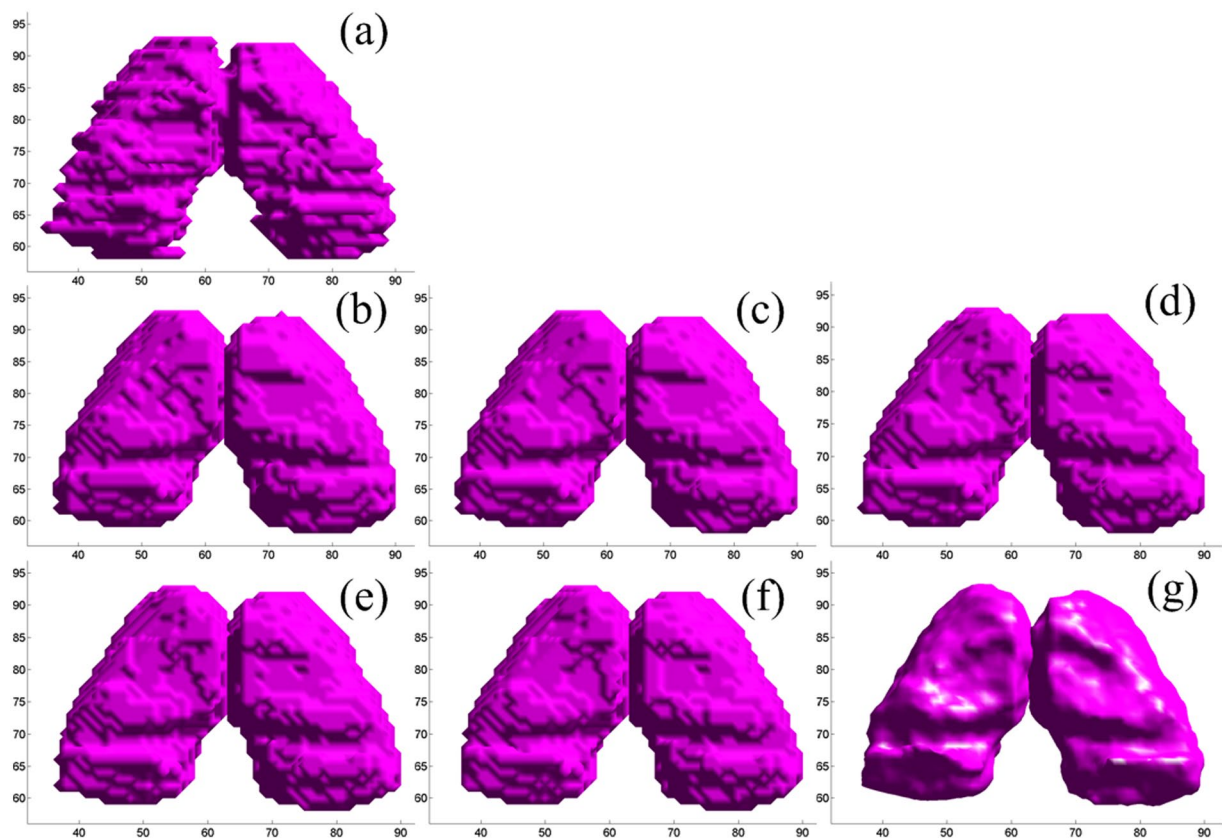
In this section important properties and possible alternative schemes of the proposed method are discussed. To the best of our knowledge, the proposed level set label fusion framework is new to the multi-atlas based segmentation. Our level set fusion framework is flexible and can be integrated with other types of registration methods and image based term for the possible improvement of thalamus segmentation. The choice of registration method will influence the overall performance of the segmentation the thalamus. One reason to use ANTS and DRAMMS as the registration methods in this paper is due to their topology preserving characteristic. Another reason is that both calculate the local similarity of each voxel in the registration procedure, and we can use the local similarity as the local weight when fusing the propagated labels.

The region-scalability of the image based term RSF is due to the kernel function with a scale parameter, which allows the use of intensity information in regions at a controllable scale, from small neighborhoods to the entire domain. As a result, this method can segment images with low contrast.

Currently, our method is implemented with Matlab 2012a on an Intel(R) Core(TM) 2.93 GHz CPU, the average time for segmenting a new image was approximately 2 min. However, our algorithm can be optimized and implemented using C/C++ in the future for significant improvement of the segmentation speed. The program can be further improved by using parallel programming techniques to reduce the segmentation time.

## Conclusion

The purpose of this work is to develop a new framework for the segmentation of the thalamus in magnetic resonance brain images. Segmentation of the thalamus is difficult and challenging because the structure boundaries may be blurry or even missing, and the surrounding background is full of irrelevant edges. To tackle these problems, we propose a novel method for segmenting the thalamus that combines multiple registration methods and a level set fusion strategy. The multiple registration methods are implemented by registering multiple atlas images to the target image using ANTS and DRAMMS. The incorporation of the level set fusion strategy enables the method to segment a thalamus with poor image contrast and to increase the robustness against errors in the registration procedure.



**Figure 9.** Surface rendering of ground truth and segmentation results (MICCAI, image No. 1006). (a) The manual segmentation, (b) STAPLE, (c) Spatial STAPLE, (d) Major Voting, (e) Weight Voting, (f) SIMPLE and (g) our proposed method.

Multi-atlas segmentation is strongly dependent on the accuracy of registration. Using multiple registration methods can combine the complementary advantages of the different registration methods and can effectively reduce the registration errors. In the level set framework, the image based term guides the level set evolves toward the desired boundary of thalamus, the fusion term fuses the propagated labels and the regularization term maintains a regular shape. The level set formulation takes into account both the image information and the regularity of the thalamus. It can fuse the labels and correct registration errors simultaneously.

The proposed method has been evaluated on a brain MR image database consisting of thirty-five patients and was further compared with several state-of-the-art brain MR segmentation algorithms using various evaluation metrics. Experimental results demonstrate that: 1) our multi-method outperforms label fusion using single registration method, and 2) the proposed method consistently achieves higher segmentation accuracy than any of the other methods under comparison.

If there is no correspondence between the target images and the atlas images, no label is propagated. This prevents an incorrect label from being introduced into the estimation of the final segmentation. Thus, further work is needed to improve the robustness of this level set label fusion method to a large diversity of pathologies and anatomical structures, such as lesions or tumors and to use this technique to potentially detect these pathological patterns.

## References

- Li, B. N., Chui, C. K., Chang, S. & Ong, S. H. Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation. *Comput Biol Med* **41**, 1–10, doi:[10.1016/j.combiomed.2010.10.007](https://doi.org/10.1016/j.combiomed.2010.10.007) (2011).
- Zhu, L., Gao, Y., Yezzi, A. & Tannenbaum, A. Automatic segmentation of the left atrium from MR images via variational region growing with a moments-based shape prior. *IEEE Trans Image Process* **22**, 5111–5122, doi:[10.1109/TIP.2013.2282049](https://doi.org/10.1109/TIP.2013.2282049) (2013).
- Naess-Schmidt, E. *et al.* Automatic thalamus and hippocampus segmentation from MP2RAGE: comparison of publicly available methods and implications for DTI quantification. *Int J Comput Assist Radiol Surg* **11**, 1979–1991, doi:[10.1007/s11548-016-1433-0](https://doi.org/10.1007/s11548-016-1433-0) (2016).
- Lankton, S. & Tannenbaum, A. Localizing region-based active contours. *IEEE Trans Image Process* **17**, 2029–2039, doi:[10.1109/TIP.2008.2004611](https://doi.org/10.1109/TIP.2008.2004611) (2008).
- Baillard, C., Hellier, P. & Barillot, C. Segmentation of brain 3D MR images using level sets and dense registration. *Med Image Anal* **5**, 185–194, doi:[S1361-8415\(01\)00039-1](https://doi.org/10.1016/S1361-8415(01)00039-1) (2001).
- Michailovich, O., Rathi, Y. & Tannenbaum, A. Image Segmentation Using Active Contours Driven by the Bhattacharyya Gradient Flow. *Image Processing, IEEE Transactions on* **16**, 2787–2801, doi:[10.1109/tip.2007.908073](https://doi.org/10.1109/tip.2007.908073) (2007).
- Oost, E. *et al.* Automated contour detection in X-ray left ventricular angiograms using multiview active appearance models and dynamic programming. *IEEE Trans Med Imaging* **25**, 1158–1171 (2006).

8. Li, C., Xu, C., Gui, C. & Fox, M. D. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans Image Process* **19**, 3243–3254, doi:10.1109/TIP.2010.2069690 (2010).
9. Liu, H., Chen, Y., Ho, H. P. & Shi, P. Geodesic active contours with adaptive neighboring influence. *Med Image Comput Comput Assist Interv* **8**, 741–748 (2005).
10. Lucas, B. C., Kazhdan, M. & Taylor, R. H. Multi-object geodesic active contours (MOGAC). *Med Image Comput Comput Assist Interv* **15**, 404–412 (2012).
11. Papandreou, G. & Maragos, P. Multigrid geometric active contour models. *IEEE Trans Image Process* **16**, 229–240 (2007).
12. Tao, W. Iterative narrowband-based graph cuts optimization for geodesic active contours with region forces (GACWRF). *IEEE Trans Image Process* **21**, 284–296, doi:10.1109/TIP.2011.2160955 (2012).
13. Hu, S., Coupe, P., Pruessner, J. C. & Collins, D. L. Appearance-based modeling for segmentation of hippocampus and amygdala using multi-contrast MR imaging. *Neuroimage* **58**, 549–559, doi:10.1016/j.neuroimage.2011.06.054 (2011).
14. Kim, M. *et al.* Automatic hippocampus segmentation of 7.0 Tesla MR images by combining multiple atlases and auto-context models. *Neuroimage* **83**, 335–345, doi:10.1016/j.neuroimage.2013.06.006 (2013).
15. Hongzhi, W. *et al.* Multi-Atlas Segmentation with Joint Label Fusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **35**, 611–623, doi:10.1109/tpami.2012.143 (2013).
16. Sjöberg, C. & Ahnesjö, A. Multi-atlas based segmentation using probabilistic label fusion with adaptive weighting of image similarity measures. *Computer Methods and Programs in Biomedicine* **110**, 308–319, doi:10.1016/j.cmpb.2012.12.006 (2013).
17. Cabezas, M., Oliver, A., Lladó, X., Freixenet, J. & Bach Cuadra, M. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine* **104**, e158–e177, doi:10.1016/j.cmpb.2011.07.015 (2011).
18. Warfield, S. K., Zou, K. H. & Wells, W. M. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* **23**, 903–921, doi:10.1109/TMI.2004.828354 (2004).
19. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044, doi:10.1016/j.neuroimage.2010.09.025 (2011).
20. Ou, Y., Sotiras, A., Paragios, N. & Davatzikos, C. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Medical Image Analysis* **15**, 622–639, doi:10.1016/j.media.2010.07.002 (2011).
21. Nguyen, P. T., Selley, D. E. & Sim-Selley, L. J. Statistical Parametric Mapping reveals ligand and region-specific activation of G-proteins by CB1 receptors and non-CB1 sites in the 3D reconstructed mouse brain. *Neuroimage* **52**, 1243–1251, doi:10.1016/j.neuroimage.2010.04.259 (2010).
22. Zhuang, X., Arridge, S., Hawkes, D. J. & Ourselin, S. A nonrigid registration framework using spatially encoded mutual information and free-form deformations. *IEEE Trans Med Imaging* **30**, 1819–1828, doi:10.1109/TMI.2011.2150240 (2011).
23. Amaral-Silva, H. T., Murta, L. O., Wichert-Ana, L., Sakamoto, A. C. & Azevedo-Marques, P. M. Medical image registration using Tsallis Entropy in Statistical Parametric Mapping (SPM). *Conf Proc IEEE Eng Med Biol Soc* **2010**, 6276–6279, doi:10.1109/IEMBS.2010.5628080 (2010).
24. Li, C., Kao, C. Y., Gore, J. C. & Ding, Z. Minimization of region-scalable fitting energy for image segmentation. *IEEE Trans Image Process* **17**, 1940–1949, doi:10.1109/TIP.2008.2002304 (2008).
25. Commowick, O., Akhondi-Asl, A. & Warfield, S. K. Estimating a reference standard segmentation with spatially varying performance parameters: local MAP STAPLE. *IEEE Trans Med Imaging* **31**, 1593–1606, doi:10.1109/TMI.2012.2197406 (2012).
26. Asman, A. J. & Landman, B. A. Formulating spatially varying performance in the statistical fusion framework. *IEEE Trans Med Imaging* **31**, 1326–1336, doi:10.1109/TMI.2012.2190992 (2012).
27. Wu, G. *et al.* A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Med Image Anal* **18**, 881–890, doi:10.1016/j.media.2013.10.013 (2014).
28. Cabezas, M., Oliver, A., Llado, X., Freixenet, J. & Cuadra, M. B. A review of atlas-based segmentation for magnetic resonance brain images. *Comput Methods Programs Biomed* **104**, e158–177, doi:10.1016/j.cmpb.2011.07.015 (2011).
29. Langerak, T. R. *et al.* Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Trans Med Imaging* **29**, 2000–2008, doi:10.1109/TMI.2010.2057442 (2010).
30. Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* **29**, 1310–1320, doi:10.1109/TMI.2010.2046908 (2010).

## Acknowledgements

This work was supported in part by Guangdong Natural Science Foundation (No. 2014A030313316, No. 2016A030313574).

## Author Contributions

Minghui Zhang and Zhentai Lu are co-first authors, and they contributed equally to this work. Specifically, Minghui Zhang and Zhentai Lu performed the experiment and wrote the paper. Qianjin Feng validated the experiment. Yu Zhang checked the results and reviewed the paper.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**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) 2017