

Atlas-based automatic mouse brain image segmentation revisited: model complexity vs. image registration

Jordan Bai^a, Thi Lan Huong Trinh^a, Kai-Hsiang Chuang^{b, c, d}, Anqi Qiu^{a, c, e, *}

^aDepartment of Bioengineering, National University of Singapore, Singapore

^bSingapore Bioimaging Consortium, the Agency for Science, Technology and Research, Singapore

^cClinical Imaging Research Center, National University of Singapore, Singapore

^dDepartment of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

^eSingapore Institute for Clinical Sciences, the Agency for Science, Technology and Research, Singapore

Received 28 June 2011; revised 8 December 2011; accepted 14 February 2012

Abstract

Although many atlas-based segmentation methods have been developed and validated for the human brain, limited work has been done for the mouse brain. This paper investigated roles of image registration and segmentation model complexity in the mouse brain segmentation. We employed four segmentation models [single atlas, multiatlas, simultaneous truth and performance level estimation (STAPLE) and Markov random field (MRF) via four different image registration algorithms (affine, B-spline free-form deformation (FFD), Demons and large deformation diffeomorphic metric mapping (LDDMM)] for delineating 19 structures from in vivo magnetic resonance microscopy images. We validated their accuracies against manual segmentation. Our results revealed that LDDMM outperformed Demons, FFD and affine in any of the segmentation models. Under the same registration, increasing segmentation model complexity from single atlas to multiatlas, STAPLE or MRF significantly improved the segmentation accuracy. Interestingly, the multiatlas-based segmentation using nonlinear registrations (FFD, Demons and LDDMM) had similar performance to their STAPLE counterparts, while they both outperformed their MRF counterparts. Furthermore, when the single-atlas affine segmentation was used as reference, the improvement due to nonlinear registrations (FFD, Demons and LDDMM) in the single-atlas segmentation model was greater than that due to increasing model complexity (multiatlas, STAPLE and MRF affine segmentation). Hence, we concluded that image registration plays a more crucial role in the atlas-based automatic mouse brain segmentation as compared to model complexity. Multiple atlases with LDDMM can best improve the segmentation accuracy in the mouse brain among all segmentation models tested in this study.

© 2012 Elsevier Inc. All rights reserved.

Keywords: Diffeomorphic mapping; Markov random field; Mouse brain segmentation

1. Introduction

In vivo magnetic resonance microscopy (MRM) of the mouse brain has been widely used to understand basic neuroscience questions [1,2], examine the influence of particular gene on the brain [3], test effectiveness of treatment [4–7] and study anatomical or morphological changes related to neurodegenerative diseases [8–10]. Automatic delineation of mouse anatomical structures of interest directly from in vivo MRM images plays an

important role in assessing structural morphology as well as its longitudinal changes. While many automatic segmentation methods have been developed for the human brain [11–17], only a few of them, including a single-atlas segmentation method [18,19] and Markov random field (MRF) models built based on image registration [18,20], have been adapted for the mouse brain. Since there is less morphological variation of the mouse brain compared to the human brain, it raises the question of how the image registration and the complexity of segmentation models affect the accuracy of the mouse brain segmentation. This paper revisited atlas-based automatic segmentation techniques, in the context of mouse brain in vivo MRM images, focusing on the effects of image registration and the complexity of segmentation models.

* Corresponding author. Department of Bioengineering, National University of Singapore, Singapore 117576. Tel.: +65 6516 7002; fax: +65 6872 3069.

E-mail address: bieqa@nus.edu.sg (A. Qiu).

Automatic atlas-based segmentation approaches have been widely used in human brain imaging studies. One of the earliest approaches is the single-atlas-based method that passes the structural label information of an atlas image to a subject's image via image registration [21]. Certainly, the performance of the segmentation is highly dependent on registration accuracy. It has also been shown that segmentation accuracy is biased towards the choice of the atlas [22–24]. Hence, researchers recently extended the single-atlas-based approach to the multiatlas-based approach, where either a few of the atlases are chosen from a set of atlases if their anatomy is more similar to that of the subject or the single-atlas method is applied using each of the atlases and the resulting segmentations combined through classifier fusion. Majority voting is widely used as the simplest classifier fusion approach [15,23]. We henceforth referred to this as multiatlas segmentation. Additional to majority voting, the simultaneous truth and performance level estimation (STAPLE) employed advanced Bayesian modeling to merge multiple segmentation results from the single-atlas segmentation approach [25,26]. Furthermore, MRF has been introduced to incorporate spatial information to aid in structural classification through the construction of a probabilistic atlas [14]. The complexity of this model has been increased tremendously, but it has several advantages: handling a characteristic spatial pattern of neuroanatomical structures and dealing with spatially heterogeneous MR imaging properties of tissues.

To date, there have been applications of the single-atlas and MRF-based segmentation to the mouse brain. Ma et al. [19] segmented MRM images of 10 adult C57BL/6J male mice using the single-atlas-based approach with affine and nonlinear registrations. No comparison, however, was made with other methods. In addition, Ali et al. [20] built the MRF model based on T2-weighted, proton density, diffusion-weighted image contrasts for the mouse brain segmentation. Bae [18] extended this approach by integrating empirical probabilistic models via support vector machines (SVM) into the MRF model (eMRF). Both the MRF and eMRF models based on affine image registration showed superior performance to the single-atlas-based segmentation [18,27]. However, there is evidence suggesting that the accuracy of the single-atlas-based segmentation can be greatly improved for specific structures if better registration techniques were employed [10]. If such improvement enables the single-atlas-based approach to achieve accuracy comparable to that of the MRF approaches, the former simpler approach may suffice to produce equally accurate segmentations. Taking this a step further, better registration could also improve the performance of those complex models, such as the multiatlas, STAPLE, MRF or eMRF methods. Furthermore, although the multiatlas-based approach has been shown to perform well for human brain images, there has thus far been neither application nor validation of the method in the context of the mouse brain image segmentation.

In this paper, we developed the single-atlas, multiatlas, STAPLE and MRF segmentation methods via advanced image registration algorithms, B-spline-based free-form deformation (FFD) [28], Demons [29,30] and large deformation diffeomorphic metric mapping (LDDMM) [31,32] for delineating 19 mouse brain structures from in vivo T2-weighted MRM images and validated their accuracy against manual segmentation. FFD and Demons were chosen because of their popularity in a variety of applications involving image registration [27,28,33–35] and their availability in the Insight Segmentation and Registration Toolkit (ITK). LDDMM algorithm was chosen because of its robustness in delineating specific structures of the human and mouse brains and good registration quality [10,36–38]. We examined the roles of registration algorithms (affine, FFD, Demons, LDDMM) under individual segmentation models as well as the importance of the complexity of segmentation models (single atlas, multiatlas, STAPLE, MRF) in the mouse segmentation. The multiatlas, STAPLE and MRF segmentation approaches aptly represent varying degrees of probabilistic model complexity with multiatlas modeling, STAPLE integrating multiple segmentation decisions, and MRF incorporating intensity likelihood and spatial priors.

2. Materials and methods

2.1. Image acquisition, preprocessing and manual segmentation

Five adult male C57BL/6 mice (23–28 g) were initially anesthetized by 1.5% isoflurane mixed in 2:1 air:O₂. The anesthetized animals were secured in a stereotactic holder with breathing and temperature monitored during imaging. The isoflurane level was varied between 1% and 1.5% to maintain a stable breathing rate of 60–90 breaths per minute, and the rectal temperature was maintained with a water bath. Three-dimensional T2-weighted turbo spin echo images (repetition time/echo time=2000/46 ms, field of view=9 mm×13 mm×25 mm, matrix=88×140×256, voxel size=100×98×98 μm and echo train length of 13) were acquired using a Bruker 7-T/20-cm ClinScan MRI (Bruker BioSpin MRI GmbH, Ettlingen, Germany) with a volume coil for transmit and a Bruker mouse brain 2×2 array coil for receive. The acquisition time was 1 h and 23 min. The animal imaging procedure was approved by the Institutional Animal Care and Use Committees of the Biomedical Sciences Institutes, Agency for Science, Technology and Research, Singapore.

We manually extracted the brain region from the intensity-inhomogeneity-corrected T2-weighted MR images [39] and normalized the image intensity to the range between 0 and 255. When considering the T2 atlas given in Ref. [40] as reference image, our brain images were rigidly transformed to it by maximizing intensity cross-correlation of the reference image and our image. Since the reference image

given in Ref. [40] is in the Franklin–Paxinos space, our T2 images are rigidly aligned to this common space to facilitate manual segmentation of the neuroanatomical structures. Using the Franklin–Paxinos atlas [41] as reference, we manually labeled 19 neuroanatomical structures (listed in Table 1) according to the definitions given in Ref. [20] for each of the five T2-weighted images (Fig. 1). Details of the delineation protocol have been described in the Supplementary Material. We considered the manual segmentation as a gold standard to evaluate the accuracy of automatic segmentation methods in this study.

2.2. Affine transformation

Affine transformation, A , is a linear transformation and has 12 degrees of freedom in three dimensions, accounting for translation, rotation, scaling and shearing. It can be found through a variational problem that maximizes the similarity between the reference image I^{atlas} and subject image $I^{subject}$. Several similarity measures, such as intensity difference, cross-correlation and mutual information, have thus far been widely used for this purpose [42,43].

2.3. FFD

FFD seeks a second-order differentiable transformation that aligns control points of one image to the other. To achieve the second-order differentiability, the transformation is characterized via a linear combination of B-splines in the form of:

$$\phi(x) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 c_{l+u,m+v,n+w} B_l(u) B_m(v) B_n(w), \quad (1)$$

where B_l represents the l th basis function of the B-spline. u , v and w , respectively, denote the lattice coordinates of the x neighbors. c denotes the coefficients of the B-splines. FFD seeks these coefficients, c , via maximizing the mutual information between the two sets of control points [28].

2.4. Demons deformable registration

The Demons algorithm [29,30] applies the concept of diffusion models to image registration to seek a diffeo-

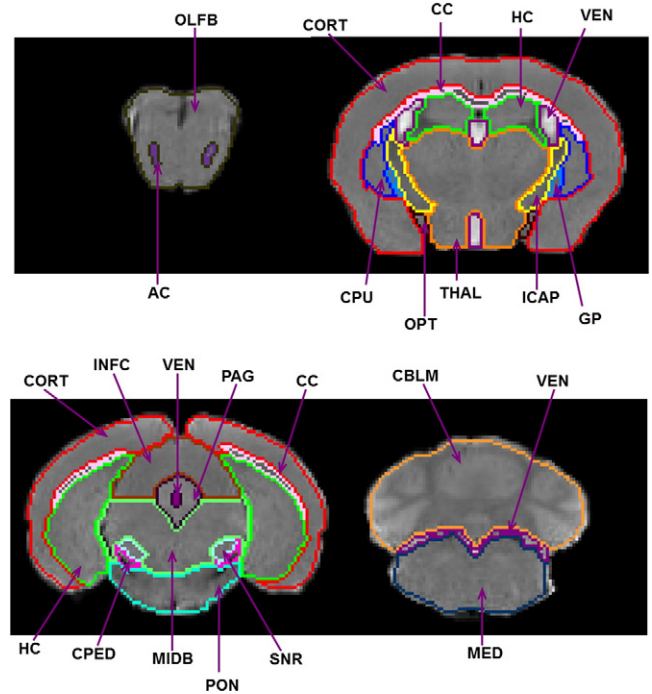


Fig. 1. Illustration of T2-weighted mouse brain image manual labels. Label abbreviations are listed in Table 1.

morphic transformation (one-to-one, reversible smooth transformations that preserve topology) that deforms one anatomy to the other. The boundaries of the reference image I^{atlas} are considered to be semipermeable membranes, while the subject image $I^{subject}$ is made to ‘diffuse’ through the membrane of I^{atlas} . The voxels in I^{atlas} are local forces to drive this diffusion process for displacing the voxels of $I^{subject}$ to match I^{atlas} . This displacement is updated iteratively and given by:

$$\phi^n(x) = \phi^{n-1}(x) - \frac{(I^{subject}(x + \phi^{n-1}(x)) - I^{atlas}(x)) \nabla I^{atlas}(x)}{(I^{subject}(x + \phi^{n-1}(x)) - I^{atlas}(x))^2 + \|\nabla I^{atlas}(x)\|^2}, \quad (2)$$

where $\phi^n(x)$ represents the displacement vector field in iteration n , with x as the voxel location, and $I^{subject}(x + \phi^{n-1}(x))$ represents the deformed subject image in the previous iteration. $\nabla I^{atlas}(x)$ is the intensity gradient of the atlas image which is kept constant for all the iterations. Gaussian smoothing of the deformation field is performed in every iteration to enforce an elastic-like behavior [29,30].

2.5. LDDMM

The LDDMM algorithm seeks a diffeomorphic transformation that deforms one anatomy to the other and defines a metric that quantifies the geodesic distance between two anatomies in a metric space. We assume that the anatomies can be generated one from the other via a flow of

Table 1
Structures and their abbreviations

Gray matter & CSF		White matter	
Cerebral cortex	CORT	Cerebral peduncle	CPED
Hippocampus	HC	Internal capsule	ICAP
Caudoputamen	CPU	Medulla oblongata	MY
Globus pallidus	GP	Midbrain	MIDB
Periaqueductal gray	PAG	Anterior commissure	AC
Inferior colliculus	INFC	Pontine nuclei	PON
Thalamus	TH	Optic tract	OPT
Cerebellum	CBLM	Corpus callosum	CC
Ventricular system	VEN		
Substantia nigra	SNR		
Olfactory bulb	OLFB		

diffeomorphisms, solutions of ordinary differential equations, $\phi = \nu_t(\phi_t)$, $t \in [0,1]$, with $\phi_0 = id$, identity maps, and associated velocity fields, ν_t . For a pair of images, I^{atlas} and $I^{subject}$, a diffeomorphic map ϕ transforms one to the other $I^{subject} = \phi_1 I^{atlas}$ at time $t=1$. Such a variational problem can thus be defined as:

$$J(\nu_t) = \inf_{\nu_t: \phi_t = \nu_t(\phi_t), \phi_0 = id} \int_0^1 \|(-\alpha \nabla^2 + \gamma id_{3 \times 3}) \nu_t\|^2 dt + \|\phi_1 \cdot I^{atlas} - I^{subject}\|^2, \quad (3)$$

where $\nabla^2 = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + \frac{\partial}{\partial z^2}$ is the Laplacian operator and $id_{3 \times 3}$ is a 3×3 identical matrix. The first term quantifies the metric distance that is the length of the geodesic curve, $\phi_t \cdot I^{atlas}$, $t \in [0,1]$, through the metric space generated from connecting I^{atlas} and $I^{subject}$. ν_t are smooth vector fields with norm $|(-\alpha \nabla^2 + \gamma id_{3 \times 3}) \cdot|$ to ensure that the solutions are diffeomorphisms. The coefficient α enforces smoothness, large values ensuring solutions are diffeomorphic, and the coefficient γ is chosen to be positive so that $-\alpha \nabla^2 + \gamma id_{3 \times 3}$ is nonsingular. The α/γ ratio affects the elasticity of the transformation. The matching quality improves as the ratio decreases [44]. The second term quantifies the intensity dissimilarity between deformed I^{atlas} and $I^{subject}$. Detailed mathematical derivation and implementation were described elsewhere [31].

2.6. Single-atlas, multiatlas, STAPLE and MRF-based segmentation methods via affine, FFD, Demons and LDDMM

As illustrated in Fig. 2, given a set of atlases, $\mathcal{I}^{atlas} = \{(I^{atlas,i}, L^{atlas,i}), i=1,2,\dots,N\}$, with atlas images, $I^{atlas,i}$, and their structural labels, $L^{atlas,i}$, we sought optimal transformations between $I^{atlas,i}$ and a subject's image, $I^{subject}$, such that the labels of the atlas image can be automatically transformed to the subject image. Assume ϕ^i to be the transformation between $I^{atlas,i}$ and $I^{subject}$. Thus, the image segmentation of $I^{subject}$, $L^{subject}$, at location x can be approximated as:

$$L^{subject}(x) = \max_l p(L^{subject}(x) = l | I^{subject}, I^{atlas,i}, L^{atlas,i}, \phi^i, i = 1, 2, \dots, N) \approx \max_l \frac{1}{N} \sum_{i=1}^N 1_{\{\phi^i \cdot L^{atlas,i}(x) = l\}}, \quad (4)$$

where l is the label of a structure, such as the cortex, hippocampus and so on. $1_{condition}$ is an indicator function, that is, when the condition is true, the outcome is 1. Otherwise, the outcome is 0. In other words, the highest occurring transformed atlas label $\phi^i \cdot L^{atlas,i}(x)$ at each voxel location is chosen as the final label in this label fusion method known as majority voting. In the single-atlas-based segmentation method, the atlas set contained only one image and one label, while in the multiatlas-based segmentation

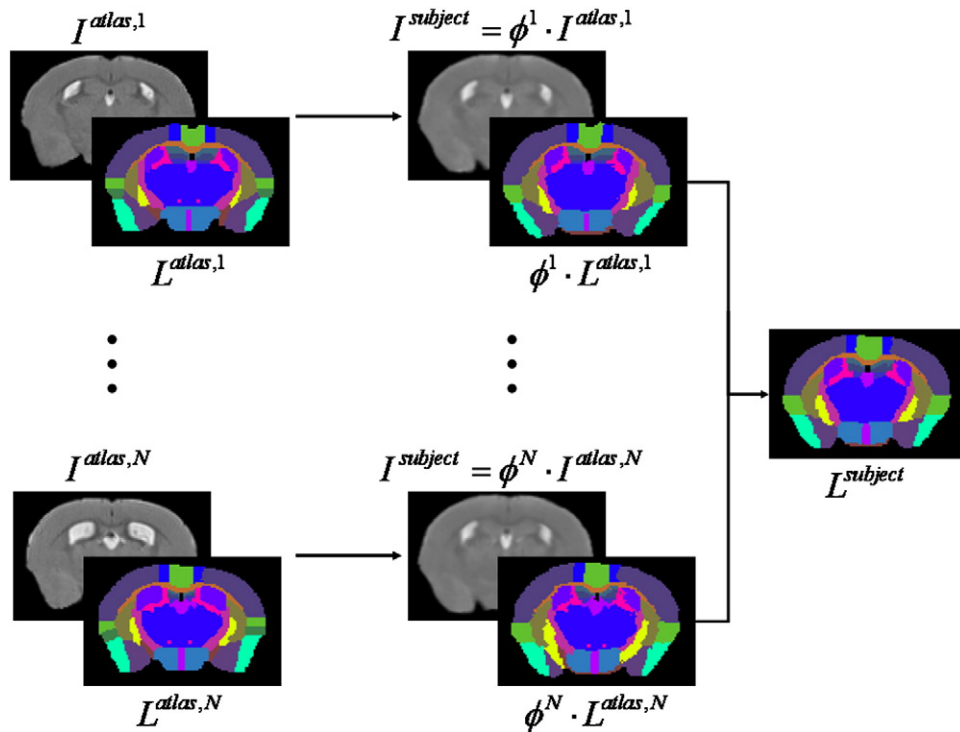


Fig. 2. Schematic of single- or multiatlas-based segmentation methods. $I^{atlas,i}$ and $L^{atlas,i}$ are the i^{th} atlas image and its label. ϕ^i is the transformation between $I^{atlas,i}$ and a subject's image, $I^{subject}$. $L^{subject}$ is the label of image $I^{subject}$.

method, the atlas set contained multiple atlas images and their labels.

While the multiatlas-based segmentation method combines the segmentation results from multiple single-atlas segmentation processes via majority voting, the STAPLE approach merges these results via Bayesian modeling that maximizes the probability, $p(L^{subject}(x)=l|I^{subject}, I^{atlas,i}, L^{atlas,i}, \phi^i, i=1,2, \dots, N)$. The mathematical derivation of this approach is detailed in Refs. [25,26], and the software is publically available at <http://crl.med.harvard.edu/>.

In the MRF approach, the probabilistic distribution of the voxel label is modeled jointly with the consideration of a neighborhood of the voxel [14,18,20]. Thus, Eq. (4) can be evaluated as:

$$L^{subject}(x) = \max_l p(L^{subject}(x) = l | I^{subject}, I^{atlas,i}, L^{atlas,i}, \quad (5)$$

$$\phi^i, i = 1, 2, \dots, N) = \max_l p(I^{subject}(x) | L^{atlas}(x) = l) p$$

$$(L^{atlas}(x) = l) \prod_{k=1}^{N_v} p(L^{subject}(x_k) | L^{atlas}(x) = l, x_k),$$

where $p(I^{subject}(x) | L^{atlas}(x)=l)$ is the likelihood of observing an image intensity $I^{subject}(x)$ for a particular label l , $p(L^{atlas}(x)=l)$ is the prior probability of a particular label l occurring at that voxel, while $\prod_{k=1}^{N_v} p(L^{subject}(x_k) | L^{atlas}(x)=l, x_k)$ models the prior probability of the neighborhood label configuration using MRF, where x_k is the neighborhood location and N_v is the size of the neighborhood. This information is trained with the aligned atlas images, and details of implementation have been thoroughly described in existing literatures [14,18,20]. In brief, $p(I^{subject}(x) | L^{atlas}(x)=l)$ is modeled using a univariate Gaussian for each label l at voxel location x , characterized by the maximum likelihood estimates of mean $\mu_l(x)$ and variance $\sigma_l^2(x)$ derived from the training data. $p(L^{atlas}(x)=l)$ is evaluated as the ratio of the number of times the voxel at location x is labeled as a particular label l by the transformed atlases to the total number of atlases. Finally, $p(L^{subject}(x_k) | L^{atlas}(x)=l, x_k)$ is estimated for each voxel location x with an adjoining voxel x_k in the first-order neighborhood ($N_v=6$) by the number of times the particular pairwise label configuration has occurred in the transformed atlases.

In our study, we applied affine, FFD, Demons and LDDMM transformations for the single-atlas-, multiatlas-, STAPLE- and MRF-based segmentations. In the affine transformation case, ϕ^i in Eqs. (4) and (5) are defined as $\phi^i=A^i$, where A^i represents the affine transformation between $I^{atlas,i}$ and $I^{subject}$ found by maximizing cross-correlation between the two images. This was implemented using Linear Image Registration Tool from Oxford Centre for Functional MRI of the Brain (FMRIB) [42,43]. ϕ^i in both Eqs. (4) and (5) are defined as the combination of A^i and the nonlinear transformations in Eq. (1) for FFD, Eq. (2) for Demons and Eq. (3) for LDDMM. FFD was implemented by optimizing the deformation vector field on

control points of arbitrarily selected lattice size of $5 \times 5 \times 5$ found by minimizing mean squared differences between the two images. The Demons transformation displacement vector field was obtained after 200 iterations of the optimization in Eq. (2). The number of iterations was chosen such that the changes of the deformation field in two subsequent iterations were sufficiently small. Both FFD and Demons were implemented using ITK. For the LDDMM case, ϕ^i in Eqs. (4) and (5) is defined as $\phi^i = \phi^{i,step2} \cdot \phi^{i,step1} \cdot A^i$, where $\phi^{i,step1}$ and $\phi^{i,step2}$ represent the sequential LDDMM transformations given in Eq. (3), to improve local alignments. In our study, we took a two-step cascading approach with a decreasing α/γ of 0.005 and 0.001 in the LDDMM mapping to gradually improve the matching quality.

2.7. Quantitative evaluation of segmentation accuracy

We evaluated the accuracies of the various segmentation approaches against the manual segmentations. The volume overlap percentage (VOP) between the automatic segmentation (L^A) and the manual segmentation (L^M) was computed as [14,18,20,22,27]:

$$VOP_l(L^A, L^M) = \frac{V(L^A \cap L^M)}{[V(L^A) + V(L^M)]/2} \times 100\%, \quad (6)$$

for structure l . $V(L)$ is the volume of the structure labeled as l . Hence, VOP gave the similarity measure between the automatic and manual segmentation, where larger VOP indicated higher accuracy of the automatic segmentation. We adopted the leave-one-out cross-validation approach to evaluate the accuracies of the different segmentation approaches. In the single-atlas-based approach, we selected one brain as atlas and segmented the rest of mouse brains in this study. We repeated this process by choosing every subject in this study as atlas and computed the VOP value averaged across all the experiments to avoid potential bias in the segmentation due to the choice of the atlas. Similarly, in the multiatlas and MRF segmentations, we selected one brain as a test set and segmented it using the rest of the brains in the study. We repeated this process by choosing every subject in this study as a test set and computed the VOP value averaged across all the experiments.

One-tailed paired t test was carried out on the mean VOP values to evaluate the significance of any improvement in accuracy when comparing one approach to the other.

3. Results

3.1. Effects of registration

Fig. 3A illustrates the accuracies of the single-atlas segmentation approach using affine, FFD, Demons and LDDMM registrations. Their comparisons based on mean VOP values indicated that nonlinear (FFD, Demons and LDDMM) registrations improved the segmentation accuracy for all 19 structures when compared to the affine alignment.

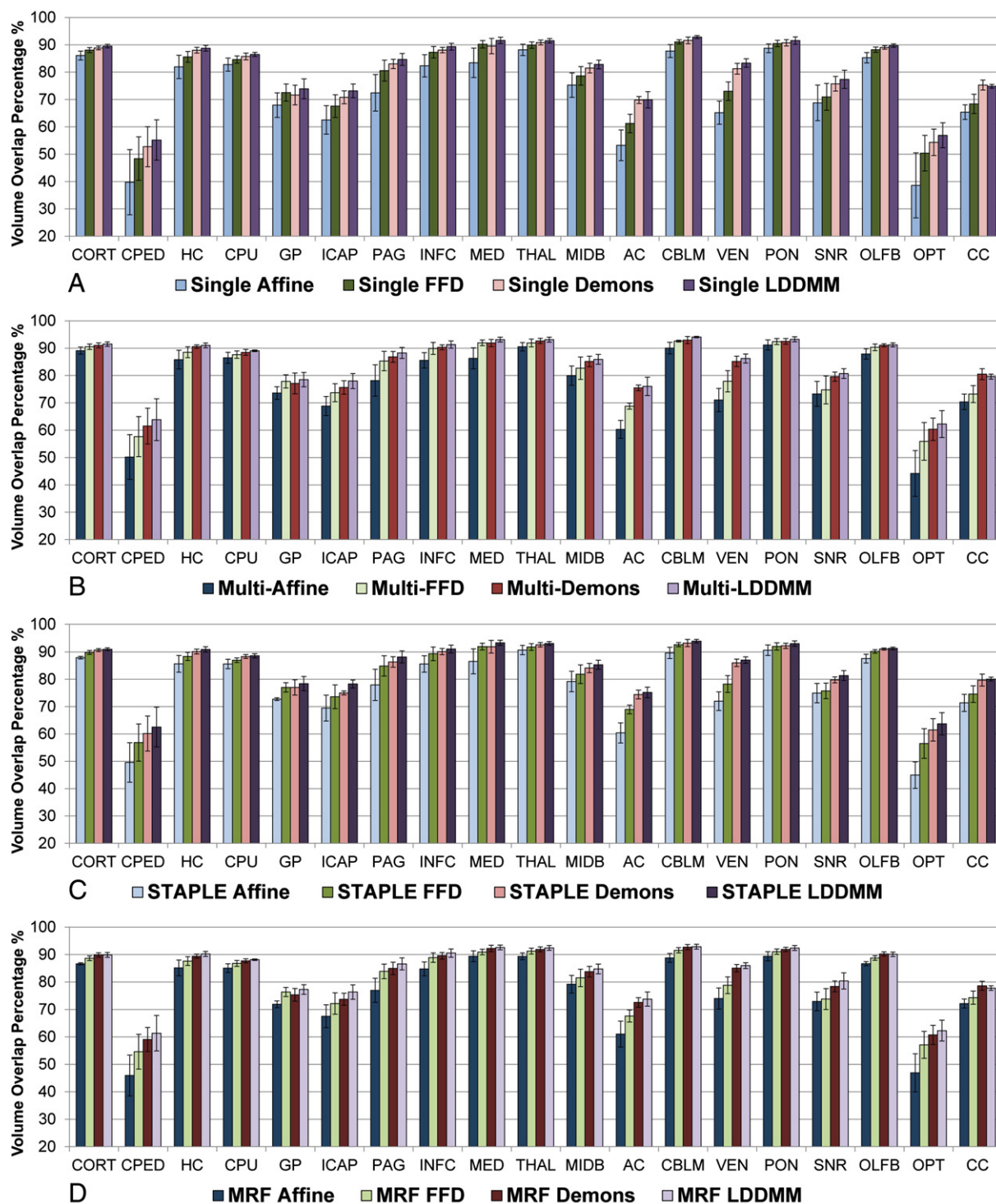


Fig. 3. Segmentation accuracy comparisons of registration algorithms for different models. (A) Volume overlap percentage obtained using the single-atlas affine (light blue), FFD (dark green), Demons (pale red) and LDDMM (purple) methods; (B) volume overlap percentage obtained using the multiatlas affine (dark blue), FFD (pale green), Demons (red) and LDDMM (pale purple) methods; (C) volume overlap percentage obtained using the STAPLE affine (light blue), FFD (green), Demons (light red) and LDDMM (dark purple) methods; and (D) volume overlap percentage obtained using the MRF affine (blue), FFD (light green), Demons (dark red) and LDDMM (pale purple) methods. The full names of the structures are listed in Table 1.

The single-atlas LDDMM approach performed the best followed by Demons, FFD and the affine-based counterparts (all $P < .001$).

Similarly, Fig. 3B–D illustrates the accuracies of using affine, FFD, Demons and LDDMM registrations in the multiatlas, STAPLE and MRF segmentation models.

Statistically, a similar trend was observed with the LDDMM approach performing best, followed by Demons, FFD and the affine-based counterparts for multiatlas, STAPLE and MRF segmentation methods (all $P < .001$). Table 2 lists mean VOP values over 19 structures for each approach.

3.2. Effects of segmentation model complexity

We investigated the effects of the model complexity on segmentation by comparing the performance of the single-atlas, multiatlas, STAPLE and MRF approaches. Compared to the single-atlas segmentation, the multiatlas segmentation improved the mean accuracy of segmentation by 4.59% (S.D.: 1.99%) for affine, 4.02% (S.D.: 2.11%) for FFD, 3.69% (S.D.: 1.86%) for Demons and 3.34% (S.D.: 1.92%) for LDDMM. The STAPLE approach improved the mean accuracy of segmentation by 4.56% (S.D.: 2.21%) for affine, 3.91% (S.D.: 2.15%) for FFD, 3.45% (S.D.: 1.76%) for Demons and 3.31% (S.D.: 1.88%) for LDDMM, while the MRF approach improved the mean accuracy by 4.12% (S.D.: 2.58%) for affine, 3.08% (S.D.: 2.14%) for FFD, 2.58% (S.D.: 1.56%) for Demons and 2.23% (S.D.: 1.67%) for LDDMM. Statistically, the multiatlas, STAPLE and MRF approaches significantly improved the segmentation accuracy when compared to the single-atlas approach under the same type of registration (all $P < .001$).

Surprisingly, increasing the model complexity from the multiatlas approach to STAPLE or to MRF did not result in any significant improvement in segmentation accuracy. For the affine transformation, there was no significantly enhanced improvement using MRF over multiatlas segmentation ($P = .299$). Under the FFD, Demons or LDDMM transformations, the performance of MRF was statistically significantly worse than that of the multiatlas segmentation and STAPLE (all $P < .001$) (Fig. 4). The comparison between the multiatlas approach and STAPLE showed

that there was no significant improvement using the STAPLE over the multiatlas segmentation under affine transformation ($P = .880$), FFD ($P = .424$), Demons ($P = .112$) and LDDMM ($P = .842$).

3.3. Interactive effects of registration and model complexity

The above results suggested that both registration and model complexity significantly influence the segmentation accuracy. We further determined the role of the registration and model complexity in improving the segmentation accuracy. Fig. 5 shows improvement of the VOP values of the single-atlas FFD, single-atlas Demons, single-atlas LDDMM, multiatlas affine, affine STAPLE and affine MRF segmentations with respect to those of the single-atlas affine segmentation. This indicated that both single-atlas Demons and LDDMM segmentations improved the accuracy significantly more than multiatlas, STAPLE and MRF models of affine-based segmentation ($P < .001$), reaffirming the prowess of improved registration technique in segmentation. In the case of FFD, it yielded higher mean segmentation accuracy than all affine-transformation-based methods (Table 2), although the improvement over affine-based single-atlas segmentation was not significantly greater than multiatlas ($P = .332$), STAPLE ($P = .311$) and MRF ($P = .055$) models of affine-based segmentation.

4. Discussion

This study revisited atlas-based automatic segmentation approaches for the mouse brain, including single-atlas-, multiatlas-, STAPLE- and MRF-based segmentation approaches using different registration methods (affine, FFD, Demons, LDDMM). Our results revealed that LDDMM outperformed FFD, Demons and affine in all single-atlas, multiatlas, STAPLE and MRF segmentation approaches. Under the same transformation, increasing model complexity from single-atlas to multiatlas, STAPLE or MRF improved the segmentation accuracy. When considering the single-atlas affine segmentation as reference, the improvement due to Demons and LDDMM registrations was statistically greater than that due to increasing model complexity of multiatlas, STAPLE and MRF. Hence, we concluded that image registration plays a more crucial role in the atlas-based automatic mouse brain segmentation, although multiple atlases can further improve the segmentation accuracy in the mouse brain.

As the single-atlas-based approach depends solely on image registration to label subjects' images, the segmentation accuracy is directly affected by how well the atlas image is aligned to subjects' images. Our segmentation results suggested that the LDDMM image registration algorithm performed best among the four registration algorithms examined in this study, including affine, FFD, Demons and LDDMM. This is congruent with the findings that showed the stationary LDDMM approach outperformed diffeomorphic

Table 2
Volume overlap percentage mean and standard deviation (S.D.) over 19 structures

Method	Mean (%)	S.D. (%)
Single-atlas segmentation via affine	72.40	15.11
Single-atlas segmentation via FFD	77.21	13.04
Single-atlas segmentation via Demons	79.93	11.46
Single-atlas segmentation via LDDMM	81.23	11.09
Multiatlas segmentation via affine	76.99	13.39
Multiatlas segmentation via FFD	81.24	11.18
Multiatlas segmentation via Demons	83.62	9.70
Multiatlas segmentation via LDDMM	84.57	9.32
STAPLE segmentation via affine	76.96	13.13
STAPLE segmentation via FFD	81.12	11.01
STAPLE segmentation via Demons	83.38	9.77
STAPLE segmentation via LDDMM	84.55	9.26
MRF segmentation via affine	76.51	13.11
MRF segmentation via FFD	80.29	11.06
MRF segmentation via Demons	82.51	10.03
MRF segmentation via LDDMM	83.46	9.50

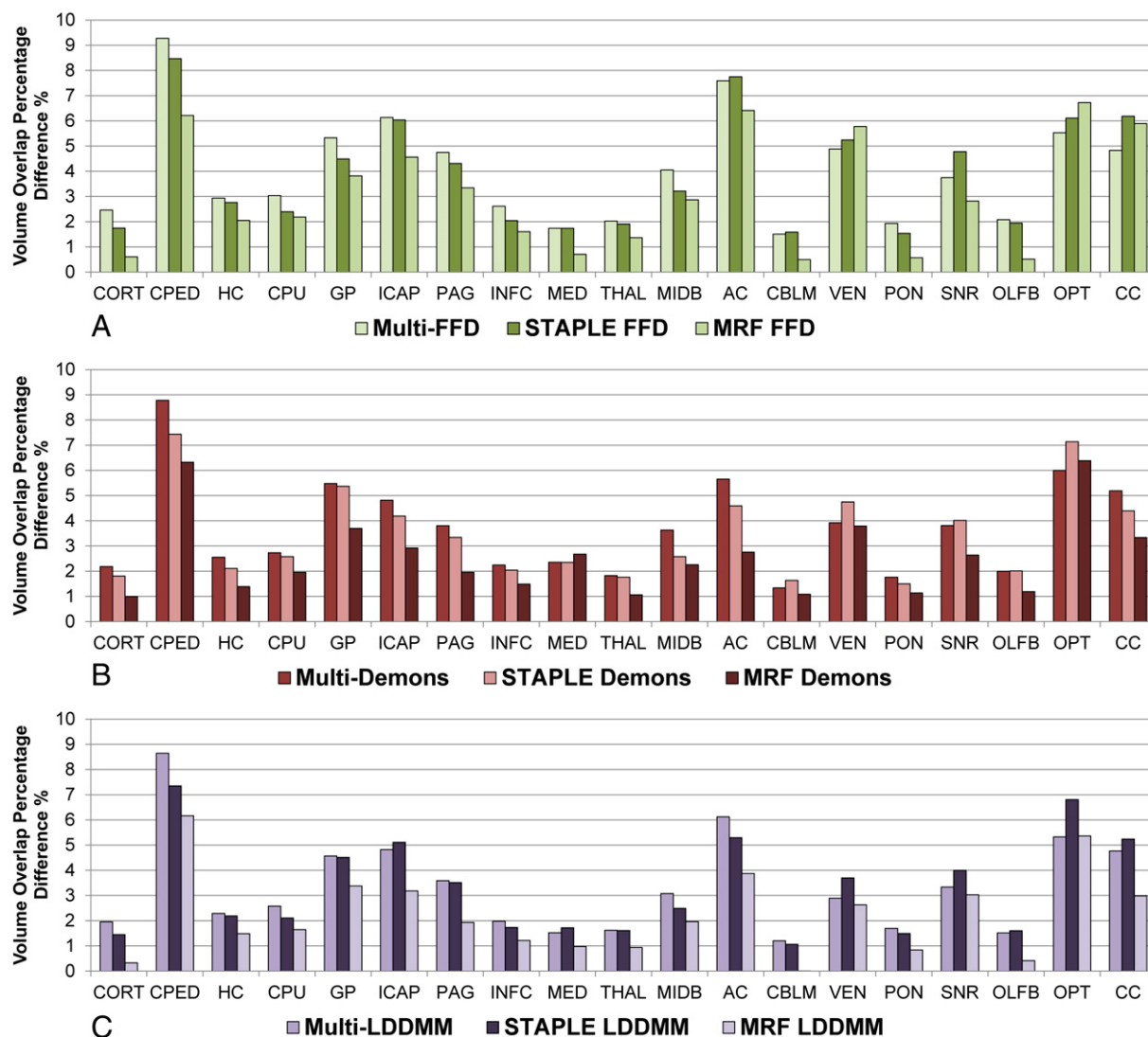


Fig. 4. Segmentation accuracy improvement through increasing model complexity under different nonlinear registration algorithms. Figure illustrates volume overlap percentage differences between the single-atlas segmentation and its multitlas and MRF counterparts obtained using (A) FFD (pale green/green/light green), (B) Demons (red/pale red/dark red) and (C) LDDMM (pale purple/purple/light purple) for the 19 structures listed in Table 1.

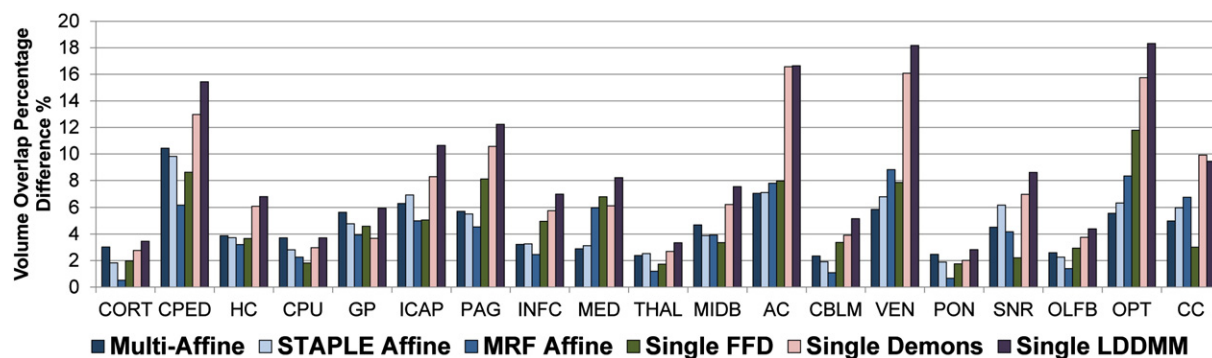


Fig. 5. Segmentation accuracy improvement through nonlinear registration versus complex models. Figure illustrates volume overlap percentage differences between the single-atlas affine method and the multitlas affine (dark blue), STAPLE affine (pale blue), MRF affine (blue), single-atlas FFD (dark green), single-atlas Demons (pale red) and single-atlas LDDMM (purple) segmentations for the 19 structures listed in Table 1.

Demons in terms of registration accuracy [45]. In our study, we employed the leave-one-out cross-validation approach to remove potential bias to the choice of a specific atlas in the segmentation, further indicating the robustness of the single-atlas FFD, Demons and LDDMM segmentation approach for the mouse brain. This evidence was also observed when the multiatlas, FFD and MRF segmentation models were used.

Increasing model complexity from single-atlas to multi-atlas, STAPLE or MRF segmentations, while keeping the registration approach the same, significantly improves the segmentation accuracy. Multiatlas segmentation is able to reduce the registration errors in single-atlas-based approach by averaging them out through mapping different atlases to the same subject. The STAPLE algorithm further incorporates estimated atlas performances to improve on the segmentation accuracies. The MRF model, on the other hand, makes use of additional information, including intensity and spatial priors, to increase its segmentation accuracy. Compared with existing literature [20], our MRF model had a slightly lower mean VOP (76.51% vs. 77.91%). Both models were based on affine transformation, while the subject size, image resolution and anatomical labels used were similar. Hence, the better accuracy reported in Ref. [20] could be attributed to the further increase in model complexity by incorporating multispectral MR intensity information in their study. Ref. [18] extended the multispectral, affine-based MRF approach by integrating empirical probabilistic models via SVM into the MRF model (eMRF) and was able to increase the mean VOP to 80.09% using the same data set as in Ref. [20]. It is often thought that MRF with nonlinear registration could perform best in the mouse brain segmentation among all methods tested in this study.

Surprisingly, our study showed that multiatlas segmentation and STAPLE performed better than the MRF model, despite the latter's increased complexity, when FFD, Demons or LDDMM was used. This may be partly because intensity information from only one image modality (T2) was incorporated in the MRF model, limiting its ability to characterize detailed anatomical differences across subjects. Moreover, one could expect that the limited samples in our study can also cause worse performance of the MRF FFD, Demons or LDDMM segmentation than that of the multiatlas counterparts. Nevertheless, we applied the same analysis for segmenting the thalamus and striatal structures from 20 human infant brain images and revealed the same results (multiatlas LDDMM: 85.02% vs. MRF LDDMM: 82.12%).

In addition, increasing model complexity for segmentation, such as incorporating multiple atlases and spatial prior information [18,20], would not necessarily improve the segmentation accuracy as much if anatomical variations across subjects cannot be properly characterized. This was supported by comparisons of the accuracy improvement, over the single-atlas affine method, using single-atlas LDDMM method with that using the multiatlas affine, STAPLE and MRF methods in our study. The affine transformation can only characterize differences in global

features across subjects, such as brain size and global shape. Even though multiple atlases were used in the segmentation, they only averaged out the alignment errors in the brain boundary rather than in the interior of the brain. On the other hand, the MRF model for affine transformation was trained with poorly aligned structures, providing nonaccurate probabilistic and spatial prior information. Thus, the multi-atlas, STAPLE and MRF affine methods did not show as much improvement in the segmentation accuracy when compared to the single-atlas LDDMM method. Compared with multispectral MRF [20] and eMRF [18], although the complexity of the segmentation model was significantly increased, the improvement in accuracy achieved was at best comparable with that using the single-atlas Demons (79.93%) and LDDMM (81.23%) methods by just looking at the reported mean VOP values (multispectral MRF: 77.91%; eMRF: 80.09%).

In conclusion, our study suggested that the multiatlas LDDMM method performed the best in the mouse brain segmentation among the single-atlas, multiatlas and MRF methods. Increasing complexity of registration or segmentation models can improve the accuracy of the mouse brain delineation. Nevertheless, MRF with nonlinear registration (Demons or LDDMM) did not outperform the multiatlas segmentation approach, suggesting that there is a need to understand the nature of the segmentation model and registration for their integration in the mouse brain segmentation.

Acknowledgment

The work was supported by grants A*STAR SERC 082-101-0025 and A*STAR SICS-09/1/1/001, a center grant from the National Medical Research Council (NMRC/CG/NUHS/2010), the Young Investigator Award at the National University of Singapore (NUSYIA FY10 P07) and the National University of Singapore MOE AcRF Tier 1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at [doi:10.1016/j.mri.2012.02.010](https://doi.org/10.1016/j.mri.2012.02.010).

References

- [1] Larvaron P, et al. In vivo analysis of the post-natal development of normal mouse brain by DTL. *NMR Biomed* 2007;20:413–21.
- [2] Maheswaran S, et al. Longitudinal regional brain volume changes quantified in normal aging and Alzheimer's APP × PS1 mice using MRI. *Brain Res* May 2009;1270:19–32.
- [3] Mueggler T, et al. Age-dependent impairment of somatosensory response in the amyloid precursor protein 23 transgenic mouse model of Alzheimer's disease. *J Neurosci* 2003;23:8231–6.
- [4] Fomchenko E, Holland E. Mouse models of brain tumors and their applications in preclinical trials. *Clin Cancer Res* 2006;12:5288–97.

- [5] Gaisler-Salomon I, et al. How high-resolution basal-state functional imaging can guide the development of new pharmacotherapies for schizophrenia. *Schizophr Bull* 2009;35:1037–44.
- [6] Suarez S, et al. Brain activation by short-term nicotine exposure in anesthetized wild-type and beta2-nicotinic receptors knockout mice: a BOLD fMRI study. *Psychopharmacology (Berl)* 2009;202:599–610.
- [7] Breton E, et al. In vivo preclinical low-field MRI monitoring of tumor growth following a suicide-gene therapy in an orthotopic mice model of human glioblastoma. *C R Biol* 2010;333:220–5.
- [8] Jack CJ, et al. Magnetic resonance imaging of Alzheimer's pathology in the brains of living transgenic mice: a new tool in Alzheimer's disease research. *Neuroscientist* 2007;13:38–48.
- [9] Boretius S, et al. In vivo MRI of altered brain anatomy and fiber connectivity in adult pax6 deficient mice. *Cereb Cortex* 2009;19:2838–47.
- [10] Zhang J, et al. Longitudinal characterization of brain atrophy of a Huntington's disease mouse model by automated morphological analyses of magnetic resonance images. *Neuroimage* 2010;49:2340–51.
- [11] Anbeek P, et al. Probabilistic segmentation of brain tissue in MR imaging. *Neuroimage* 2005;27:795–804.
- [12] Babalola KO, et al. An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage* 2009;47:12.
- [13] Brouwer RM, et al. Segmentation of MRI brain scans using non-uniform partial volume densities. *Neuroimage* 2010;49:467–77.
- [14] Fischl B, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 2002;33:341–55.
- [15] Heckemann RA, et al. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* 2006;33:115–26.
- [16] Klein A, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* Jul 2009;46:786–802.
- [17] Wu J, Chung ACS. A novel framework for segmentation of deep brain structures based on Markov dependence tree. *Neuroimage* 2009;46:1027–36.
- [18] Bae MH, et al. Automated segmentation of mouse brain images using extended MRF. *Neuroimage* 2009;46:717–25.
- [19] Ma Y, et al. A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. *Neuroscience* 2005;135:1203–15.
- [20] Ali AA, et al. Automated segmentation of neuroanatomical structures in multispectral MR microscopy of the mouse brain. *Neuroimage* 2005;27:425–35.
- [21] Iosifescu D, et al. An automated registration algorithm for measuring MRI subcortical brain structures. *Neuroimage* 1997;6:13–25.
- [22] Aljabar P, et al. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage* 2009;46:726–38.
- [23] Artaechevarria X, et al. Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans Med Imaging* 2009;28:1266–77.
- [24] Lötjönen J, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 2010;49:2352–65.
- [25] Rohlfing T, et al. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans Med Imaging* 2004;23:983–94.
- [26] Warfield SK, et al. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 2004;23:903–21.
- [27] Badea A, et al. Morphometric analysis of the C57BL/6J mouse brain. *Neuroimage* 2007;37:683–93.
- [28] Rueckert D, et al. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 1999;18:712–21.
- [29] Thirion JP. Fast non-rigid matching of 3D medical images. *Med Robot Comput Aided Surg* 1995:47–54.
- [30] Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal* 1998;2:243–60.
- [31] Beg MF, et al. Computing metrics via geodesics on flows of diffeomorphisms. *Int J Comput Vis* 2005;61:19.
- [32] Miller MI, Qiu A. The emerging discipline of computational functional anatomy. *Neuroimage* 2009;45:S16–39.
- [33] Hojjat SP, et al. Micro-computed tomography-based highly automated 3D segmentation of the rat spine for quantitative analysis of metastatic disease. *J Neurosurg Spine* 2010;13:367–70.
- [34] Werner R, et al. Estimation of motion fields by non-linear registration for local lung motion analysis in 4D CT image data. *Int J Comput Assist Radiol Surg* 2010.
- [35] Maheswaran S, et al. Analysis of serial magnetic resonance images of mouse brains using image registration. *Neuroimage* 2009;44:692–700.
- [36] Miller M, et al. Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. *Proc Natl Acad Sci U S A* 2005;102:9685–90.
- [37] Du J, et al. Whole brain diffeomorphic metric mapping via integration of sulcal and gyral curves, cortical surfaces, and images. *Neuroimage* 2011;56:162–73.
- [38] Zhong J, et al. Quantitative evaluation of LDDMM, FreeSurfer, and CARET for cortical surface mapping. *Neuroimage* 2010;52:131–41.
- [39] Sled JG, et al. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 1998;17:87–97.
- [40] MacKenzie-Graham A, et al. A multimodal, multidimensional atlas of the C57BL/6J mouse brain. *J Anat* 2004;204:93–102.
- [41] Franklin KBJ, Paxinos G. The mouse brain in stereotaxic coordinates, compact. 3rd ed. New York: Academic Press; 2008.
- [42] Jenkinson M, et al. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 2002;17:825–41.
- [43] Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal* 2001;5:143–56.
- [44] Ceritoglu C, et al. Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. *Neuroimage* 2009;47:618–27.
- [45] Hernandez M, et al. Comparing algorithms for diffeomorphic registration: stationary LDDMM and diffeomorphic Demons. *Proc. MFCAWorkshop of MICCAI'08*; 2008.