NeuroImage 125 (2016) 120–130

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Resource atlases for multi-atlas brain segmentations with multiple ontology levels based on T1-weighted MRI☆

Dan Wu [a], Ting Ma [a], Can Ceritoglu [b], Yue Li [c], Jill Chotiyanonta [a], Zhipeng Hou [a], John Hsu [a], Xin Xu [a], Timothy Brown [b], Michael I. Miller [b,d], Susumu Mori [a,e,*]

[a] Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[b] Center for Imaging Science, Johns Hopkins University, Baltimore, MD, USA
[c] AnatomyWorks, LLC, Baltimore, MD, USA
[d] Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[e] F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 2 September 2015
Accepted 17 October 2015
Available online 21 October 2015

Keywords:
Multi-atlas
T1-weighted
Atlas creation strategy
Hierarchical ontology
Dynamic age-matching
Segmentation accuracy

ABSTRACT

Technologies for multi-atlas brain segmentation of T1-weighted MRI images have rapidly progressed in recent years, with highly promising results. This approach, however, relies on a large number of atlases with accurate and consistent structural identifications. Here, we introduce our atlas inventories (n = 90), which cover ages 4–82 years with unique hierarchical structural definitions (286 structures at the finest level). This multi-atlas library resource provides the flexibility to choose appropriate atlases for various studies with different age ranges and structure-definition criteria. In this paper, we describe the details of the atlas resources and demonstrate the improved accuracy achievable with a dynamic age-matching approach, in which atlases that most closely match the subject's age are dynamically selected. The advanced atlas creation strategy, together with atlas pre-selection principles, is expected to support the further development of multi-atlas image segmentation.

Introduction

Quantitative analysis of brain MRI data has played a pivotal role in many brain anatomical studies of development, aging, and various pathological conditions. For image-based quantification, the first and the most important step is to define corresponding brain locations across all participants of the study. One of the most common approaches, which is considered the gold standard in neuroanatomy, is the manual delineation of regions of interest (ROIs). However, because it is labor intensive, manual delineation is usually used for hypothesis-driven studies, in which a small number of target structures are preselected based on a hypothesis. Voxel-based analysis (VBA) is a widely used alternative approach, in which every single voxel is considered an ROI, and corresponding voxel locations are identified across all participants automatically using an image normalization method (Ashburner, 2009; Worsley et al., 1999). Whole-brain structural segmentation is an alternative approach, in which the voxels are joined based on a priori anatomical knowledge, e.g., voxels that belong to the caudate should be joined as one structure (Faria et al., 2010; Fischl et al., 2002; Heckemann et al., 2006; Joshi et al., 2004; Mori et al., 2005; Pham and Prince, 1999; Tu et al., 2008; Tustison et al., 2014; Tzourio-Mazoyer et al., 2002; Woolrich et al., 2009). If the entire brain is segmented into multiple structures, both VBA and segmentation-based methods provide the information about the anatomical features of the entire brain, but from very different granularity levels; in the segmentation-based approach, the information from more than 10^6 voxels in VBA is greatly reduced to the order of 10^2.

The meaning of "atlas" varies depending on research fields and study purposes and, thus, clarification is needed. The typical brain atlas consists of images (e.g. histology or MRI) and point-and-annotate labels, describing the locations and names of brain structures. For VBA, atlases mean MR images in specific orientations, positions, and matrix coordinates. They usually do not contain structural labels but the x, y, and z coordinates carry the common anatomical meaning. The images could be chosen within a study or external data such as those from International Consortium of Brain Mapping (ICBM). The MNI coordinates are one of the most widely used standard coordinates and the image could be single-subject such as MNI-Colin27 (Collins et al., 1998) or population-averaged such as linear and nonlinear MNI-ICBM152 (Fonov et al., 2011; Mazziotta et al., 2001). For the segmentation-based analysis (and throughout this paper), the atlas means MRI images with structural labels with three-dimensional boundaries, which carry

☆ Grant support: This publication was made possible by the following grants: NS084957, EB017638, EB015909, and NS086888.
* Corresponding author at: Johns Hopkins University School of Medicine, Traylor 330, 720 Rutland Avenue, Baltimore, MD 21205, USA. Fax: +1 410 614 1948.
E-mail address: susumu@mri.jhu.edu (S. Mori).

http://dx.doi.org/10.1016/j.neuroimage.2015.10.042
1053-8119/

anatomical references to define structures of interest. One of the simplest approaches to accomplish automated brain segmentation is to warp a single-subject atlas to each subject and transfer the structure labels (Fischl et al., 2002).

In recent years, a multiple-atlas approach has gained popularity due to its superior segmentation accuracy (Artaechevarria et al., 2009; Jia et al., 2012; Langerak et al., 2010; Lotjonen et al., 2010b; Sabuncu et al., 2010; van Rikxoort et al., 2010; Wang et al., 2012; Warfield et al., 2004). In this approach, rather than one reference brain atlas, multiple atlases with consistent structural segmentation are prepared, warped to a subject image, and the multiple segmentation results are fused to achieve the best estimation of the structural identification. Numerous publications have reported improved accuracy using this approach. Most of these previous papers have focused on the algorithms used to fuse the multiple segmentation results. The multiple-atlas approach, however, depends on the availability of atlases with accurate and consistent structural definitions, which is not only labor-intensive, but also has several important issues to be addressed.

For example, one of the most frequently asked questions is how many atlases are needed to improve the segmentation accuracy. Aljabar et al. (Aljabar et al., 2009) has shown that the segmentation accuracy reaches maximum with 15–25 atlases depending on the target structures. However, probably, there is no general answer to this question, because it depends on the anatomical variability among the atlases and the anatomical features of the subject. In an extreme case, if the anatomy of the subject is an outlier, with respect to the anatomical ranges covered by the multiple atlases, we cannot expect high accuracy regardless of the number of atlases used. This naturally extends to the notion of atlas pre-selection based on anatomical or non-anatomical features, such as the ventricle size or age-matching. Another important issue is the way in which anatomy is defined. The minimum definable units are determined by the available anatomical features and contrasts; it is difficult to sub-divide, for example, the globus pallidus (GP) into the internal and external portions, unless MRI provides contrasts to define these intra-GP structures. For segmentation-based analysis, the finer definitions may not always be better. For example, if the entire temporal lobe has 10% atrophy, a report of the volume loss of a dozen of substructures within the temporal lobe could not only hurt the statistical power but also result in misleading findings about the regional specificity of the abnormality if the atrophy was statistically detected only in a subset of the constituents. Thus, it is important to test regional specificity by examining larger super structures, such as lobes and hemispheres.

The purpose of this study is to establish a shared atlas resource to support multi-atlas segmentation algorithms for quantitative analyses of brain MR images and test the accuracy levels as well as the efficacy of atlas pre-selection approaches such as dynamic age-matching, which is enabled by the availability of the wide range of the age coverage. In this paper, we introduce our atlas inventories ($n = 90$) with unique hierarchical structural definitions, which cover ages 4–85 years. The atlas set can be combined with various available multi-atlas fusion algorithms. In addition, we tested the impact of age-matching on the accuracy of the segmentation.

Materials and methods

Overview of atlas creation strategy

For neuroanatomical studies, the manual delineation of structures by experienced neuroanatomists is considered the gold standard, against which the performance of automated segmentation is measured. However, a recent survey about protocols for the manual delineation of the hippocampus found as much as 250% of volume differences (Boccardi et al., 2011). These differences stem from anatomical definitions; it is not that one is correct and the others are inaccurate. In addition, anatomy often does not have boundaries. Even with histology, at a cellular-level definition, the boundary of a structure could change or overlap with adjacent structures depending on staining methods and specific cell types of interest. With only 15 MB of information stored in a typical T1-weighted anatomical image, cellular-level validation of structural boundaries is ill-posed and, thus, the segmentation process should be considered an engineering procedure that allows us to compare brains in a systematic manner, rather than a tool to find a ground truth. In this respect, we are interested in the precision (reproducibility) of the measurement as a tool, rather than in a quest to achieve the best possible accuracy (validity). The pre-segmented atlases serve as one of the references for structural definitions to pose anatomical boundaries with consistent criteria.

With manual delineation by an experienced rater, we invite several precision-related issues: intra-rater (delineate the same image multiple times); inter-rater (different raters delineate the same image); and inter-measurement (scan the same person multiple times) variability. With fully automated segmentation tools, we can eliminate the intra-rater and inter-rater variability. However, there is still a segmentation accuracy issue. Please note that there are two types of accuracy. The first is the accuracy of a segmentation protocol (e.g., the boundaries defined in an atlas), which, as mentioned above, is an elusive goal. The second accuracy question is whether the automated segmentation result delivers the boundary definitions that agree with the protocol. When we measure the performance of automated segmentation tools, this second type of accuracy issue and the inter-measurement precision become the two important questions. While these goals are usually the target of algorithm development, it is not widely recognized that the atlas creation strategy (i.e., segmentation strategy) also has an impact on the performance of the algorithm; thus, the algorithm and atlas development should evolve hand-in-hand. The comprehensive description of our protocols for structural definitions can be found at braingps mricloud.org/atlasrepo. Below, a couple of important strategies are described.

Atlas creation strategy I: the minimum unit

The minimum unit (or the maximum numbers of the unit) we can define is one voxel, which is the unit VBA uses. The most natural way to join voxels and define a structure is to use a hypothesis- or biology-driven approach. For example, one may be interested in the dentate gyrus of the hippocampus or the substantia nigra of Parkinson's disease patients. However, image resolution and available contrasts pose practical limitations to this approach; in reality, these image-based factors dictate the structure we can define. In Fig. 1A and B, images from diffusion tensor imaging (DTI) and T1 weighting are compared for the same person at the same slice level at the pons. All the defined structures in the pons are visually appreciable in the DTI image, while the pons lacks contrasts in the T1 image. If there are no anatomical clues to define these structures, such as the substructures in the pons of a T1-weighted image, it would be difficult to pre-segment these structures accurately in the atlas and to measure their volumes; the results reflect the algorithm parameters, such as elasticity, smoothing of the transformation field, and interpolation. Apparently, the definable minimum units depend on available contrasts and imaging modality. The image contrast is not the only anatomical clue. For example, as shown in Fig. 1C, the anterior limb and posterior limb of the internal capsule are both white matter structures, and their boundary cannot be defined by their contrasts, but they can be separated based on anatomical features (the sharp bend and feature-rich surrounding gray matter anatomy). Therefore, it is reasonable to separate them at the genu (the kink) of the internal capsule.

Atlas creation strategy II: incorporation of hierarchical definitions based on ontology

The definable minimum structures depend on the image modality. The DTI allows finer structural definitions of the white matter and T1-weighted images can define more gray matter structures due to the higher image resolution. This means pre-segmented brain atlases must be created for each modality. One way to introduce a systematic

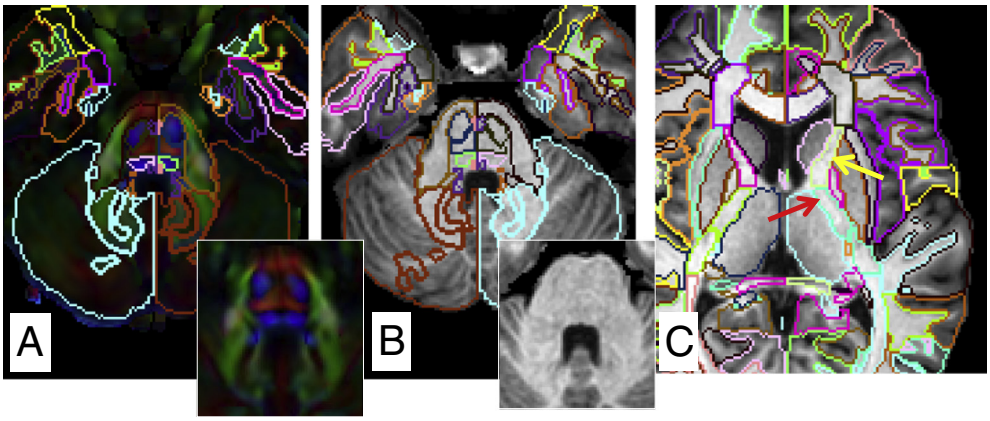

Fig. 1. Examples of the relationship between anatomical features and definable structures. The Direction-encoded colormap from DTI (A) and the T1-weighted image (B) obtained from the same person at the same slice level at the pons. Magnified views of the pons are shown in the insets. These images show the importance of anatomical contrasts to define structures of interest. On the other hand, availability of contrasts is not always a necessary factor to define a structure. For example, in (C), the anterior limb and the posterior limb of the internal capsule share the similar image intensities, but they can still be separated at the genu (the kink) of the internal capsule based on the anatomical clue; the sharp bend and feature-rich surrounding gray matter anatomy.

management of anatomical structure is to introduce a hierarchical structural relationship based on the ontology. For example, brain atlases and the anatomical literature (Mai et al., 2007a; Puelles et al., 2013) define how various structures should be hierarchically grouped. Based on the minimum units described above, higher hierarchical levels (superstructures) can be defined by joining the units, according to (Mai et al., 2007b). Fig. 2 shows images of our hierarchical relationship defined in five different levels. For the pons, for example, level 5 defines too many intra-pons structures for the T1 images. The volume measurements should employ level 4 and higher in this case. The multi-level hierarchical relationship is also important for examining the specificity of anatomical findings. For example, if one measures the hippocampus and finds a 10% volume loss, s/he may conclude "this disease causes hippocampal atrophy." However, the anatomical interpretation may change if the same amount of atrophy is found at higher hierarchical levels such as the entire limbic system, the gray matter, or the hemisphere. The hierarchical analysis, therefore, is important for examining the regional specificity of the findings. Our recent multi-ontology-level analysis confirmed expected results about the relationship between the precision (test–retest reproducibility) and ontology levels when they were used for automated multi-atlas brain segmentation (Djamanakova et al., 2014). As the level went up (thus, fewer defined structures and more voxel grouping), the precision went up, saturating at about 1.5% test–retest reproducibility using T1-weighted images with 1 mm resolution. At level 5, approximately 20% of the defined structures are smaller than 1000 mm^3, which resulted in substantially poorer precision, mostly in the 3–10% range. Of course, the gain in precision comes at the cost of the loss of region-specific information.

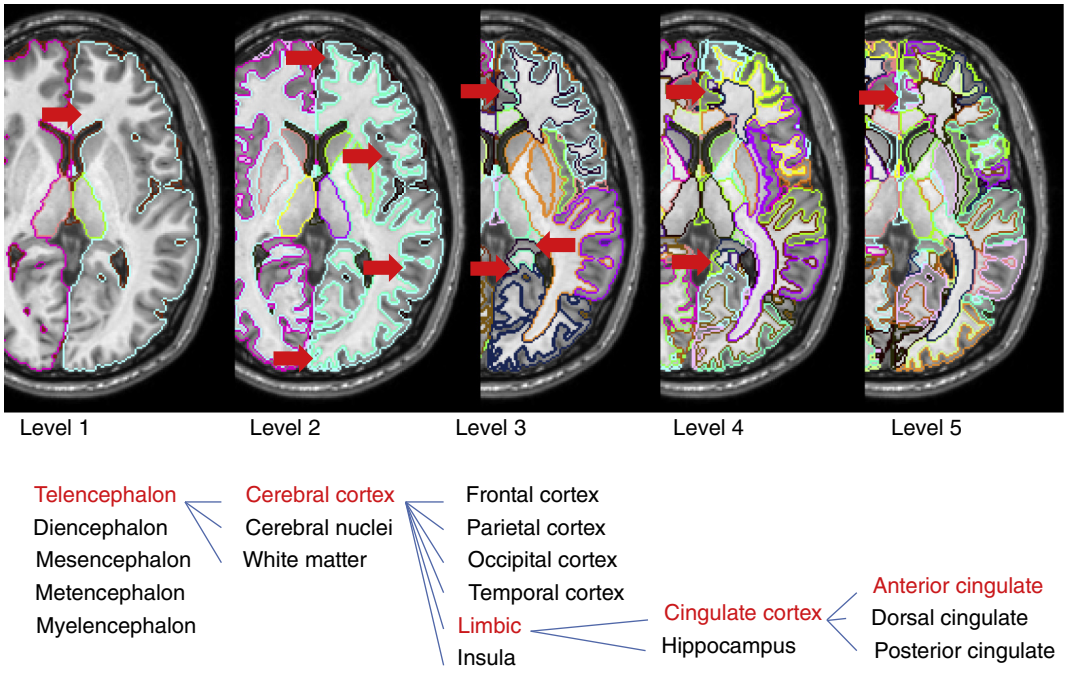

Fig. 2. Hierarchical relationship defined at five different levels. (A) An image representation of the five-level structural delineation. (B) An example of the hierarchical definitions in the telencephalon at level 1 to the anterior/dorsal/posterior cingulate cortex at level 5. The structures defined by a red color are indicated in the images, describing how the structures are defined in a finer unit as the level goes up.

The hierarchical relationship can be described by a structure-relationship table, with 286 structures listed in Appendix 1. In our system, this relationship can be readily modified by users and we currently offer two types of relationships (Type I and II), as will be described in the Results section.

Atlas creation

There are four different sources of data, each covering different age ranges with different protocols. All data were acquired using the MPRAGE sequences provided by the manufacturers, as specified below. The in-plane resolution was 1.0–1.1 mm and the second phase-encoding resolution was 1.0–1.5 mm.

Adult atlas

These data were acquired at JHU using 3 T Philips scanners. The Adult atlases consist of 47 subjects (41 normal subjects and six Alzheimer's disease patients) with an age range of 22–82 years (49.3 ± 20.9 years).

Pediatric atlas

Pediatric-I: These data were acquired at JHU using the same scanners as in the Adult Atlas. This atlas set consists of 19 pediatric cases with an age range of 8–12 years (10.0 ± 1.1 years).

Pediatric-II: These data were acquired at JHU using Siemens 3 T scanners. This atlas set consists of 11 pediatric cases with an age range of 8–18 years (11.5 ± 3.2 years).

Pediatric-III: These data were obtained from PING (Pediatric Imaging, Neurocognition, and Genetics, http://pingstudy.ucsd.edu/) and a mixture of images acquired on 3 T Siemens and Philips scanners. The scan information can be found at http://pingstudy.ucsd.edu/resources/neuroimaging-cores.html. This atlas set consists of 13 subjects with an age range of 4–8 years (6.7 ± 1.3 years).

A subset of the Adult atlases were initially segmented using our single-subject atlas (Oishi et al., 2009), which was warped to individual subjects using DiffeoMap (www.mristudio.org), followed by manual corrections based on the criteria described previously (Faria et al., 2015; Oishi et al., 2009). For pediatric and elderly atlases, substantial manual corrections were made to ensure accurate structural definitions (Faria et al., 2010; Oishi et al., 2013). Once a set of multiple atlases were established, all other data were segmented by multi-atlas segmentation with general age-matching (adult atlases for adult brains, and pediatric atlases for pediatric brains), followed by manual corrections. The list of the defined structures is in the Appendix, and their ontology-based hierarchical relationship based on (Mai et al., 2007a) (Version 6.12) is posted at www.mricloud.org. These images and hierarchical relationships can be interactively visualized by using RoiEditor (www.mristudio.org), as described in the Results section.

Atlas fusion algorithms

We tested two multi-atlas fusion algorithms in this study. Most of the analyses were performed using the fully-automated MriCloud pipeline (www.mricloud.org), which is based on the Diffeomorphic Multi-Atlas Likelihood Fusion (DMALF) algorithm (Tang et al., 2013). The pipeline uses Large Deformation Diffeomorphic Metric Mapping (LDDMM) for image registration (Ceritoglu et al., 2010). The input to the pipeline is unprocessed raw MPRAGE images and multiple atlases were registered to the input image using LDDMM, followed by DMALF-based atlas fusion. The details of DMALF and LDDMM can be found elsewhere (Tang et al., 2013; Tang et al., 2014). To investigate the generalizability of our findings (e.g., the effect of atlas age-matching), we also used a joint fusion algorithm (Wang et al., 2012), developed at PICSL (Penn Image Computing & Science Lab, www.picsl.upenn.edu/), for which we fed the same atlas sets, after LDDMM-based registration to the test data. The joint fusion-based atlas fusion is also available in www.mricloud.org.

MRI data used in this study

In this study and many others (Aljabar et al., 2009; Lotjonen et al., 2010a; Rohlfing et al., 2004), the manually corrected atlases were divided into the atlases and test data to test the segmentation accuracy. To evaluate the importance of age-matching, we tested two different approaches. First, we tested conventional "binned" atlas approaches; the "Pediatric atlas" with ten atlases randomly drawn from the Pediatric-I and Pediatric-II (age 8–12 years old, average = 10.1 ± 1.4 years old) and the "Adult atlas" with ten atlases from the Adult atlas set (age 22–82 year old, average = 49.8 ± 24.2 years old) were used. The large number of the atlases also enabled us to test a dynamically age-matched approach, in which the entire atlases (n = 90) were pooled, five data were chosen at 5, 8, 10, 15, 25, 40, 60, 80 years old as test data and ten atlases with closest age range were selected as age-matched atlases for each age level (Table 1). Narrower age levels were tested for the younger brains (5, 8, 10, 15 years old), compared to the adult brains (25, 40, 60, 80 years old), considering that the brain actively develops during early stages and gradually stabilizes in adulthood.

Data analysis

Because these test data (a part of the atlases) have manually-defined segmentation results, the accuracy of the automated multi-atlas segmentation can be measured from the extent of spatial matching with pre-defined structures in the atlases. Here we used DICE (Dice, 1945) measurement, which is the ratio between the intersection area A∩B and the union of AUB, where A and B denote the manual and the automated segmentations. Results were presented as mean ± standard error of the mean (SEM). The differences in DICE between the three atlas sets (pediatric, adult, and age-matched) were evaluated using one-way analysis of variance (ANOVA) for each age level and each structure, or two-way ANOVA for each structure and correction for age levels, followed by post-hoc Bonferroni's multiple comparison tests. A p value < 0.01 was considered significant and $p < 0.05$ was considered weakly significant. To assess the effects of age-matching and the fusion algorithm, we used a two-way ANOVA followed by post-hoc Bonferroni's tests. In this study, the following structures were used to measure the accuracy:

Deep gray matter structures: the caudate, the putamen, the globus pallidus, the hippocampus, and the amygdala, which were defined at Type I — Level 5.

Cortex: the entire cortex, which was defined at Type I — Level 2.

White matter structures: the cerebral peduncle, the anterior and posterior limb of internal capsule, and body of the corpus callosum, which were defined at Type I — Level 5.

Ventricles: the entire lateral ventricles, which were defined at Type I — Level 2.

These structures were chosen because they are relatively easier to define due to clear contrast boundaries.

Results

Ontology definition, reporting, and visualization system

The current version of the atlases, Version 6.12, defines 286 structures. As we expect the number of the atlases to increase with frequent updates in the future, we adopted a version control system by Git (www.github.com/git), which can be accessed through braingps.mricloud.org/git/gitweb.cgi. Through this system, users can have access to the latest, as

Table 1
MRI data used in the atlas age-matching study. Five subjects at 5, 8, 10, 15, 25, 40, 60, and 80 years of age were drawn from the atlas pool as test data. Age-matched atlases comprised ten atlases with the closest age range for each age level, as specified.

Subject (n = 5)	Age-matched atlas (n = 10)	Pediatric atlas (n = 10)	Adult atlas (n = 10)
5 years (range 4–5 years)	4–8 years	8–12 years	32–82 years
8 years	6–10 years	8–12 years	32–82 years
10 years	8–12 years	8–12 years	32–82 years
15 years (range 12–18 years)	10–20 years	8–12 years	32–82 years
25 years (range 25–26 years)	20–30 years	8–12 years	32–82 years
40 years (range 38–42 years)	30–50 years	8–12 years	32–82 years
60 years (range 57–63 years)	50–70 years	8–12 years	32–82 years
80 years (range 79–82 years)	70–80 years	8–12 years	32–82 years

well as previous versions of the atlases. The list of the 286 structures are shown in the Appendix. The comprehensive ontological relationship can be downloaded, viewed, and modified from the *atlas repository* (https://braingps.mricloud.org/atlasrepo) and also is included in the outputs of the image segmentation tools at MriCloud. As shown in Fig. 3, RoiEditor has an interface to read the ontology relationship file and interactively visualize the brain segmentation results at each ontology level. It also allows users to save a quantitative report (volumes and intensities of defined structures) and ROI definition files at each level. Currently, two types (Type I and II) of hierarchical relationships are offered and available in www.mricloud.org, in which the 286 structural units are combined to define different superstructures (Fig. 4). For example, at the lowest ontology level, Type I defines seven classical definitions of the brain ontology (telencephalon (right and left), diencephalon (right and left), mesencephalon, metencephalon, and myelencephalon), while Type II defines four structures that are widely used in clinical descriptions (hemispheres (right and left), cerebellum, and brainstem). For example, the metencephalon includes the pons and the cerebellum, which share the same developmental precursor, while the brainstem and the cerebellum are often defined as different entities in image research. The ontology relationship defined in the text file can be readily modified by users to create superstructures that could better serve users' clinical hypotheses.

The effect of the age-matching

The effects of age-matching on the segmentation accuracy are shown in Fig. 5 for the structures that had strong age effects. Compared to the pediatric atlas and the adult atlas, the age-matched atlas consistently showed superior performance across different age levels for the cerebral cortex, the caudate, and the ventricles. The statistical analysis detected significant differences between the age-matched approach and the binned atlases at several age levels (Fig. 5A). Considering test

Fig. 3. A screenshot of the RoiEditor (www.mristudio.org) interface to read the ontology relationship file and interactively visualize the brain segmentation results at each ontology level. It provides the ability to save a quantitative report (volumes and intensities of defined structures) and ROI definition files at each level.

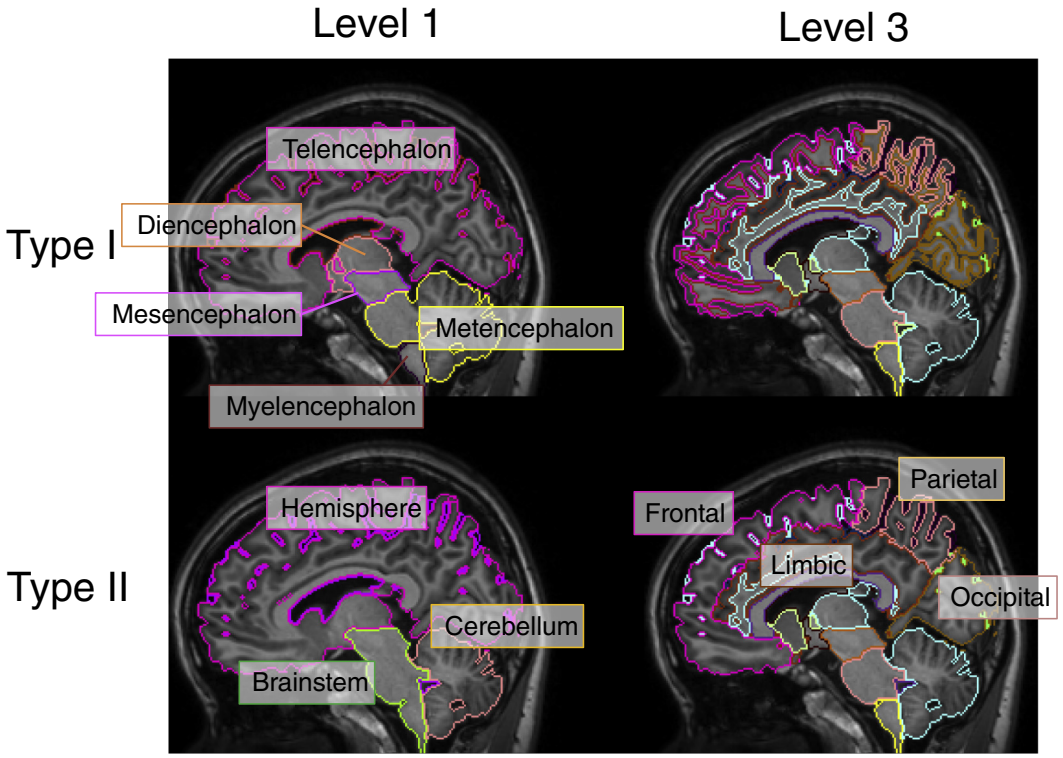

Fig. 4. Demonstration of two types (Type I and II) of hierarchical relationships at two ontology levels.

subjects at all ages, two-way ANOVA analysis showed a prominent effect of age-matching ($p < 0.001$ for the cortex and caudate, and $p < 0.01$ for the lateral ventricles). The pediatric atlas performed well for the 8- and 10-year-old subjects. It also achieved a high level of accuracy up to 40 years of age, but the performance quickly deteriorated for over 60 years of age where significant atrophy develops. The adult atlas performs generally well for adult brains up to teenagers (15-year-old). Please note that the DICE values of the ventricles tend to increase over age, which is likely due to their increased sizes in the elderly population; as is the nature of the DICE, smaller and narrower objects tend to have lower DICE.

We also compared the performance of two multi-atlas fusion methods — DMALF and joint fusion in the five-year-old subject group (Fig. 5B). Our observations were independent of the algorithms. Two-way ANOVA analysis was performed on the effects of the age and the fusion algorithm, which revealed a significant effect of age-matching ($p = 0.0007$, $p = 0.012$, and $p = 0.079$ in the cortex, caudate, and lateral ventricles, respectively), but no significance on fusion methods ($p > 0.3$). The result indicated the importance of age-based atlas pre-selection, and less impact of the fusion methods.

In the other deep gray matter structures, such as the hippocampus, amygdala, putamen, and globus pallidus, the improvement of DICE with the age-matched atlas was less prominent (Fig. 6). All atlas sets performed stably well across different age levels (DICE > 0.86 and 0.88 for the hippocampus and the amygdala, and >0.9 for the putamen and the globus pallidus), except when the pediatric atlas was used for subjects over 60 years of age or the adult atlas was used for subjects under 10 years of age. The observation was similar in the white matter structures (Fig. 7), such as the body of the corpus callosum (BCC), the cerebral peduncle (CP), the anterior limb of the internal capsule (ALIC), and the posterior limb of the internal capsule (PLIC). On average, DICE was over 0.87 if the atlas age was generally matched. ANOVA analysis indicated that these structures are less sensitive to atlas age-matching ($p > 0.05$, except for the hippocampus, amygdala, and BCC in the 80-year-old subjects). Overall, taking all the data together (Figs. 5–7), the effect of atlas age-matching was statistically significant ($p < 0.0001$), using a multiple ANOVA and correcting for the different structures and subject ages.

Discussion

Overview of the atlas creation

In this paper, we introduced our multiple atlas library of T1-weighted brain MR images. This library was developed to support multi-atlas image segmentation tools, which are currently a target of highly active research. As the atlas library serves as a teaching file for computer algorithms to judge which structures are located where and with what type of anatomical signatures, the availability of atlases with high-quality segmentation is essential. To achieve better segmentation accuracy, the atlas creation strategy (knowledge creation) could be as important as algorithm improvements (knowledge application). Many modern segmentation algorithms utilize not only location, but also intensity information to identify a structure of interest. If each target structure is defined in the atlas with a narrowly defined intensity histogram, such algorithms could place a high level of weighting on the image contrast to accurately define the boundary. Therefore, the way atlases are created certainly affects the performance of the multi-atlas algorithms. Another example is the effect of space occupancy. Many labeling algorithms assign labels aggressively (greedily) to voxels. If there are brain regions that are not labeled (vacant regions) in the atlases, the labels of nearby structures could leak into such regions. Therefore, it is important to assign structures even when such structures are not an interest of the research work.

One of the most common questions about brain atlases is, “which atlas is the most appropriate for my study?” For example, in the conventional atlas-based analysis, ICBM-152 in the MNI coordinate system is widely used, which was created by averaging images from 152 healthy adult volunteers. Is it appropriate to use this atlas for geriatric populations with pronounced atrophy? Or can we use this atlas as a target in

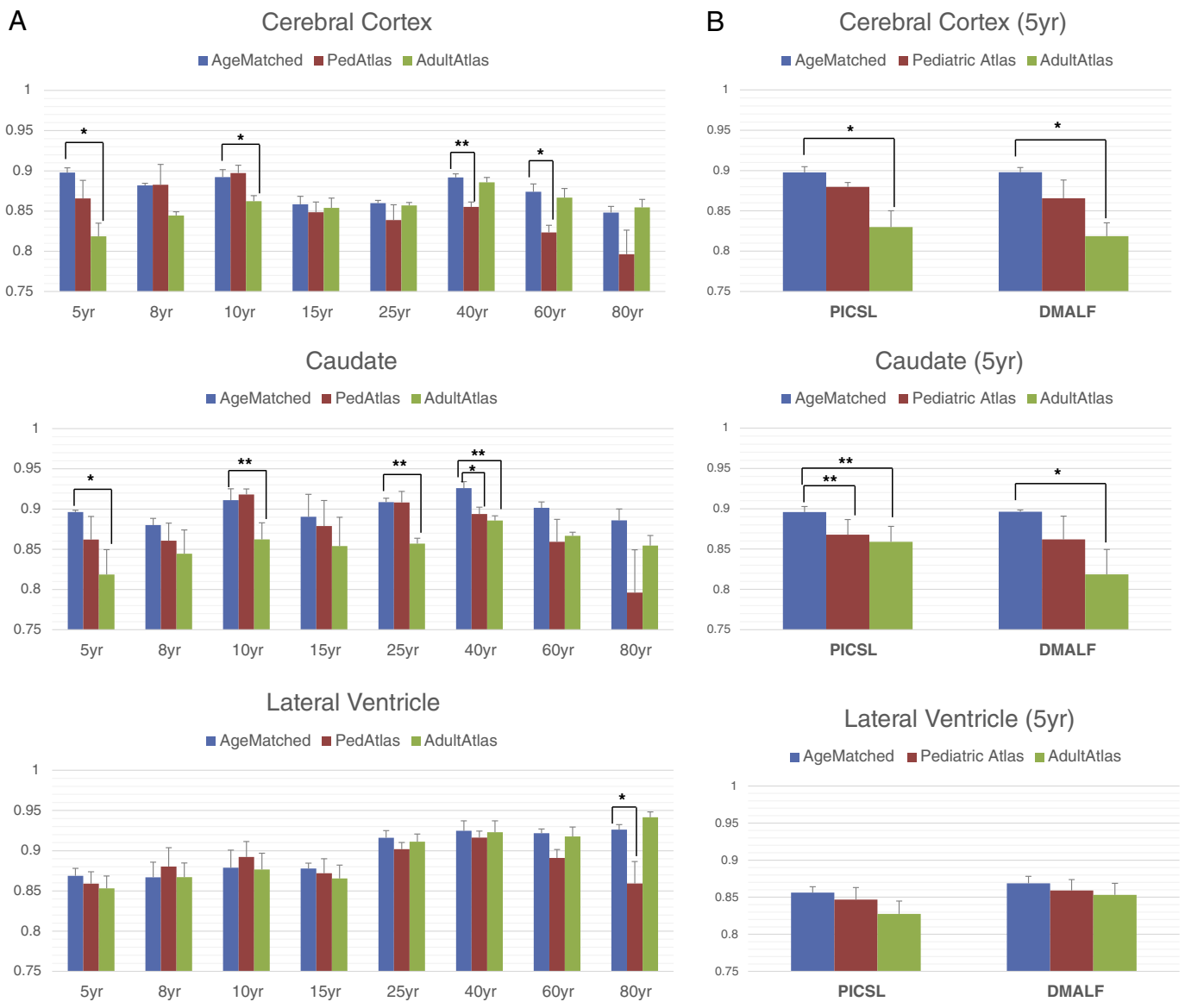

Fig. 5. Effects of atlas age-matching on multi-atlas segmentation accuracy in the cerebral cortex, the caudate, and the lateral ventricle, as evaluated by DICE (mean ± standard error of the mean). The DICE measures the spatial overlap and the higher the DICE, the higher the segmentation accuracy is. (A) DICE of the atlas sets that were matched to the subject age were compared with that of the pediatric atlas and adult atlas using test subjects at the ages of 5, 8, 10, 15, 25, 40, 60, and 80 years old (n = 5 each groups). The use of atlases is specified in Table 1. $^{*}p < 0.05$ and $^{**}p < 0.01$ from one-way ANOVA followed by post-hoc Bonferroni's tests. (B) Comparison of the segmentation accuracy using joint fusion and DMALF fusion methods in the five-year-old subject group. $^{*}p < 0.05$ and $^{**}p < 0.01$ from two-way ANOVA followed by post-hoc Bonferroni's tests.

studies of teenagers? In the multi-atlas regime, the images from the 152 subjects are not averaged to create one population-averaged atlas. One drawback of this approach is that all images from the 152 subjects need to be segmented with consistent criteria. However, we would gain a flexibility to choose atlases that are most appropriate for the study population from the 152 atlases. These pre-selection criteria could be subject attributes such as age, gender, or race, or could be image features, such as the level of atrophy. Among these criteria, our anatomical knowledge indicates that age should be one of the largest factors that influence the brain anatomy (Faria et al., 2010; Gogtay et al., 2004; Good et al., 2001; Huang et al., 2006; Shaw et al., 2008).

Atlas library as a resource

The first aim of this paper was to introduce our multiple-atlas library and provide detailed descriptions. In many cases, the notion of the brain structures is merely a concept because what we call "brain structures" or "anatomical names" often do not have clear boundaries in the cellular level. This is especially so for the white matter structures. For example, the corpus callosum can only be clearly demarcated at the mid-sagittal level, and there is no clear definition laterally; axons continue to the cortical regions. This can also be applied to many gray matter structures. For example, at a microscopic level, the caudate does not have a boundary at the lateral surface; there is only a continuum to the putamen, penetrated by the axons in the internal capsule. Therefore, it is often not possible to "accurately define" or "validate" a brain structure. In this respect, the role of the brain atlas is to apply a consistent anatomical definition to all study populations. The anatomical definitions in our atlas library is an extension of our previous work (Mori et al., 2008; Oishi et al., 2008), which was, in turn, defined based on the past literature as consistently as possible (Mazziotta et al., 1995; Toga and Mazziotta, 2002).

One unique aspect of our atlas is the availability of ontological relationships for the 286 defined structures. Based on these relationships, superstructures can be created to test the regional specificity of anatomical findings. This relationship is defined in a text file and can be readily

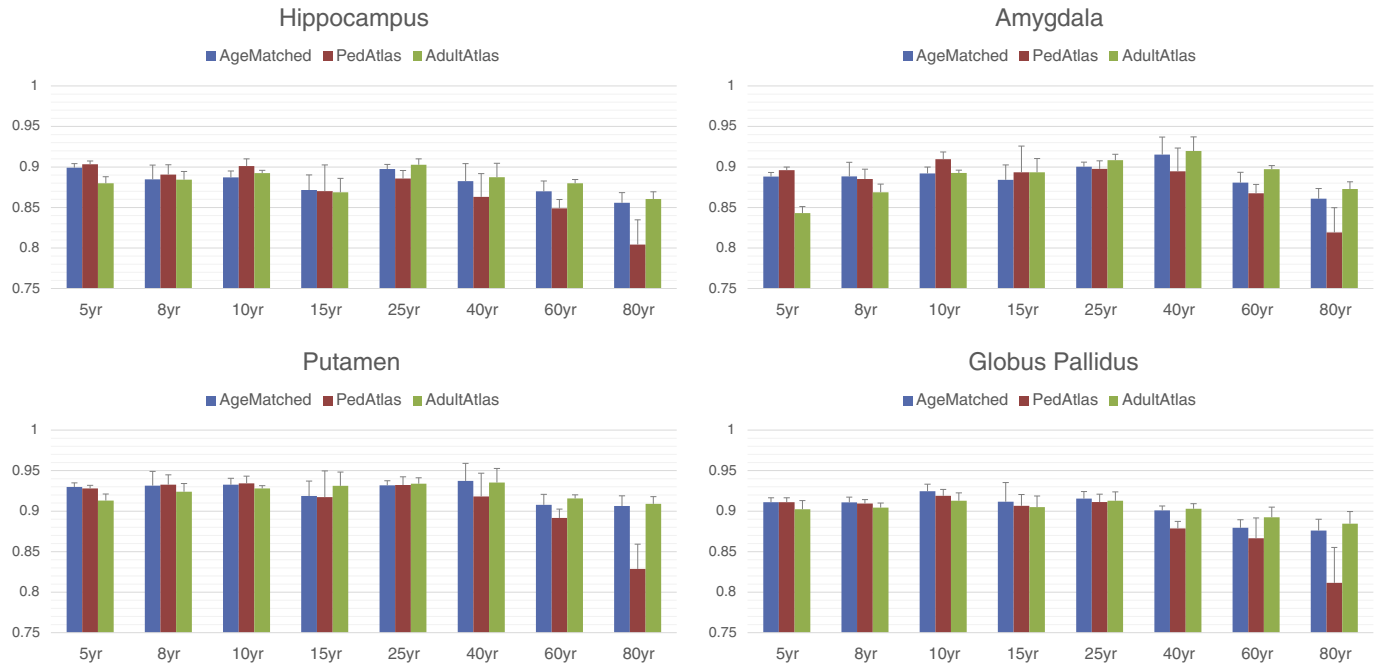

Fig. 6. Performance of the age-matched atlas, pediatric atlas, and adult atlas in four deep gray matter structures—the hippocampus, amygdala, putamen, and globus pallidus—using test subjects at the ages of 5, 8, 10, 15, 25, 40, 60, and 80 years old (n = 5 each groups). No statistically significant differences in DICE were found between the age-matched atlas and the pediatric atlas or adult atlas by post-hoc Bonferroni's tests following one-way ANOVA.

modified by users. The graphical interface by RoiEditor makes such modification and inspection straightforward. All resources are available through www.mristudio.org (RoiEditor) and www.mricloud.org (atlas resources).

One important issue about the proposed atlas resource is that it consists of data from multiple sources with different types of the scanners and image protocols. Ideally, atlases and test data share exactly the same data acquisition parameters but generating fully-segmented atlases for each study could be impractical. In many atlas-based analyses, including VBA, it is common to employ external atlases. For example, most of the MNI-ICBM atlas resources were based on MR images from 1.5 T scanners, while many modern MRI studies use 3 T scanners. Therefore, the mismatch of the data acquisition protocols between the external atlases and users' data is a common issue for many studies,

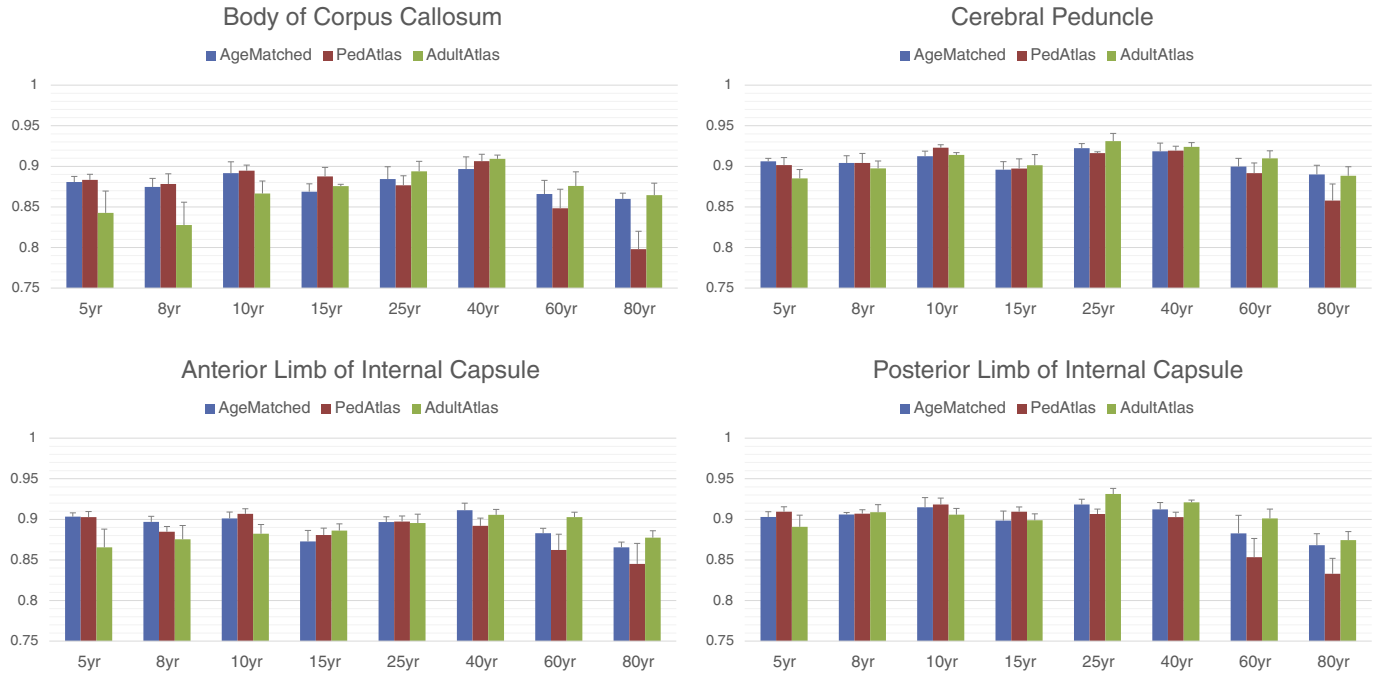

Fig. 7. Performance of the age-matched atlas, pediatric atlas, and adult atlas in four white matter structures — the body of the corpus callosum, the cerebral peduncle, the anterior limb of the internal capsule, and the posterior limb of the internal capsule, using test subjects at 5, 8, 10, 15, 25, 40, 60, and 80 years of age (n = 5 each groups). No statistically significant differences in DICE were found between the age-matched atlas and the pediatric atlas or adult atlas by post-hoc Bonferroni's tests following one-way ANOVA.

which should be minimized through image pre-processing or advanced registration tools. On the other hand, the heterogeneity of data acquisition parameters within a multiple-atlas library poses a unique problem. It is not immediately clear how important it is to acquire all data within a library with exactly the same parameters, given the variability in other demographic information (Liang et al., 2015). In reality, it is difficult to obtain an image library that spans the entire age-range with a single set of data acquisition parameters. One potential issue is, if one uses age-matched atlases for two subjects with very different ages, the two age-matched atlas sets could consist of different proportions of, for example, scanner manufacturers, which could introduce a certain amount of bias in the segmentation results. Although this (inclusion of substantially different age groups within a same study) rarely happens in conventional research studies, we need to be aware of this type of shortcomings of our atlas resources. In the future, as the number of atlases increase, it could become possible to pre-select atlases not only by their ages or anatomical features, but also data acquisition parameters.

Evaluation of age-matching

The second purpose of the study was to evaluate the effect of age-matching. Dynamic age-matching, in which atlases with the ages closest to the subject are dynamically selected based on the subject age, is a relatively new approach and it is possible only when a large number of atlases are available for a wide range of ages. In this paper, we tested the age-matching and conventional binned atlas sets (pediatric and adult atlases).

Atlas age-matching has been reported as one of the most important atlas pre-selection criteria (Aljabar et al., 2009). Brains of similar ages usually share similar anatomical features, such as the shape of the CSF space, ventricle sizes, and gray matter/white matter anatomy. The tissue contrast of age-matched brains are likely to be similar in structural MR images, because T1 and T2 relaxations change with age due to the myelination and water content in the tissues. Using age-matched atlases in mutli-atlas based segmentation could improve the image registration (and thus, segmentation) accuracy (Heckemann et al., 2010) between brains with similar morphometric and photometric features. Our results suggest that the effect of atlas age-matching is significant in many structures and is particularly important for early childhood and elderly brains, where the influences of brain growth or atrophy are most significant.

Among the structures we measured in our study, the segmentation accuracy of the cerebral cortex and the caudate are most dependent on age-matching. It is known that the cortex undergoes active growth from childhood to adulthood (Gogtay et al., 2004; Shaw et al., 2008), and that cortical atrophy accelerates with age (Raz et al., 1997; Scahill et al., 2003). This often manifest as the more pronounced CSF space in elderly brains. This leads to the necessity of a larger amount of image transformation to align two age-unmatched brains at the cortical surface, which may introduce registration errors. The caudate nucleus, which rapidly develops during brain maturation (Giedd et al., 1996; Huang et al.,

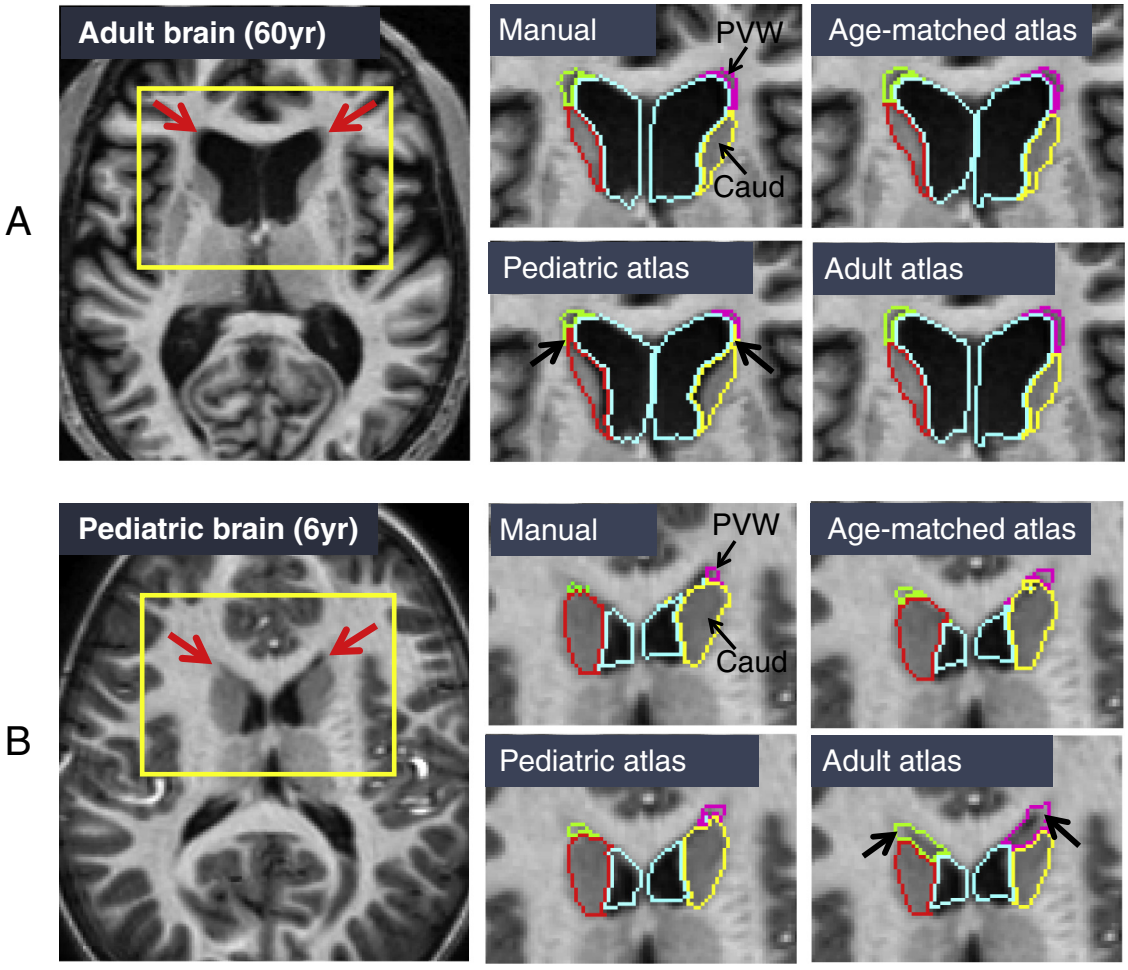

Fig. 8. Examples of segmentation errors in the caudate nuclei when an age-unmatched atlas was used. The dark region (periventricular white matter (PWM), indicated by red arrows) present in the elderly brain, is not seen in the younger brains. Due to this reason, if the pediatric atlas is used to segment the adult brain, the head of the caudate may intrude into the PWM; and if the adult atlas is used to segment the pediatric brain, the PWM may overlap part of the caudate.

2006), demonstrated a high level of atlas-age dependence for segmentation accuracy (Aljabar et al., 2009). The main reason is the presence of the hypo-intense white matter at the anterior horn of the lateral ventricles, which exists in almost all elderly populations above 60 to a different extent (Fig. 8). In the pediatric atlases, the darkening of these white matter regions is almost non-existent and if these pediatric atlases are applied to elderly populations, they are often labeled as a part of the caudate (Fig. 8A). In the opposite situation (elderly atlases used for pediatric cases), the labels that define the hypo-intense white matter areas invade a part of the caudate (Fig. 8B). This is an example of age-dependent anatomical features that would require age-matched atlases for better segmentation accuracy. The lateral ventricle is another structure, the shape of which has significant changes over development and aging (Coffey et al., 1992; Scahill et al., 2003). In the pediatric brain, a significant portion of the lateral ventricle spaces are closed and cannot be seen on T1 images. Diffeomorphic image registration cannot register structures if they do not exist, and, thus, it is understandable that the pediatric atlases could not accurately define the enlarged ventricles in the elderly populations.

As important as atlas pre-selection, the improvement in segmentation accuracy also depends on image registration and atlas-fusion strategies. It is important to note, therefore, that the reported effects of the age-matching could vary depending on the algorithms employed for the multi-atlas approaches.

Conclusion

We present multi-atlas inventories of 90 atlases, ranging from 4 to 82 years of age for T1-weighted brain MRI segmentation, which were established with accurate and consistent structural definition and hierarchical ontology, along with quantification and visualization tools. This large atlas database can be best used if combined with atlas pre-selection principles. Dynamic age-matching was shown to be a simple and efficient pre-selection approach that improved the segmentation accuracy for several brain structures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.10.042.

References

Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46, 726–738.

Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de-Solorzano, C., 2009. Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans. Med. Imaging 28, 1266–1277.

Ashburner, J., 2009. Computational anatomy with the SPM software. Magn. Reson. Imaging 27, 1163–1174.

Boccardi, M., Ganzola, R., Bocchetta, M., Pievani, M., Redolfi, A., Bartzokis, G., Camicioli, R., Csernansky, J.G., de Leon, M.J., deToledo-Morrell, L., Killiany, R.J., Lehericy, S., Pantel, J., Pruessner, J.C., Soininen, H., Watson, C., Duchesne, S., Jack Jr., C.R., Frisoni, G.B., 2011. Survey of protocols for the manual segmentation of the hippocampus: preparatory steps towards a joint EADC–ADNI harmonized protocol. J. Alzheimers Dis. 26 (Suppl. 3), 61–75.

Ceritoglu, C., Wang, L., Selemon, L.D., Csernansky, J.G., Miller, M.I., Ratnanather, J.T., 2010. Large deformation diffeomorphic metric mapping registration of reconstructed 3D histological section images and in vivo MR images. Front. Hum. Neurosci. 4, 43.

Coffey, C.E., Wilkinson, W.E., Parashos, I.A., Soady, S.A., Sullivan, R.J., Patterson, L.J., Figiel, G.S., Webb, M.C., Spritzer, C.E., Djang, W.T., 1992. Quantitative cerebral anatomy of the aging human brain: a cross-sectional study using magnetic resonance imaging. Neurology 42, 527–536.

Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C., 1998. Design and construction of a realistic digital brain phantom. IEEE Trans. Med. Imaging 17, 463–468.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Djamanakova, A., Tang, X., Li, X., Faria, A.V., Ceritoglu, C., Oishi, K., Hillis, A.E., Albert, M., Lyketsos, C., Miller, M.I., Mori, S., 2014. Tools for multiple granularity analysis of brain MRI data for individualized image analysis. NeuroImage 101, 168–176.

Faria, A.V., Zhang, J.Y., Oishi, K., Li, X., Jiang, H.Y., Akhter, K., Hermoye, L., Lee, S.K., Hoon, A., Stashinko, E., Miller, M.I., van Zijl, P.C.M., Mori, S., 2010. Atlas-based analysis of neurodevelopment from infancy to adulthood using diffusion tensor imaging and applications for automated abnormality detection. NeuroImage 52, 415–428.

Faria, A.V., Oishi, K., Yoshida, S., Hillis, A., Miller, M.I., Mori, S., 2015. Content-based image retrieval for brain MRI: an image-searching engine and population-based analysis to utilize past clinical data for future diagnosis. Neuroimage Clin. 7, 367–376.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., Brain Development Cooperative, G., 2011. Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54, 313–327.

Giedd, J.N., Snell, J.W., Lange, N., Rajapakse, J.C., Casey, B.J., Kozuch, P.L., Vaituzis, A.C., Vauss, Y.C., Hamburger, S.D., Kaysen, D., Rapoport, J.L., 1996. Quantitative magnetic resonance imaging of human brain development: ages 4–18. Cereb. Cortex 6, 551–560.

Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent, T.F., Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L., Thompson, P.M., 2004. Dynamic mapping of human cortical development during childhood through early adulthood. Proc. Natl. Acad. Sci. U. S. A. 101, 8174–8179.

Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14, 21–36.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33, 115–126.

Heckemann, R.A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J.V., Hammers, A., Initi, A.s.D.N., 2010. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. NeuroImage 51, 221–227.

Huang, H., Zhang, J., Wakana, S., Zhang, W., Ren, T., Richards, L.J., Yarowsky, P., Donohue, P., Graham, E., van Zijl, P.C., Mori, S., 2006. White and gray matter development in human fetal, newborn and pediatric brains. NeuroImage 33, 27–38.

Jia, H., Yap, P.T., Shen, D., 2012. Iterative multi-atlas-based multi-image segmentation with tree-based registration. NeuroImage 59, 422–430.

Joshi, S., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23 (Suppl. 1), S151–S160.

Langerak, T.R., van der Heide, U.A., Kotte, A.N.T.J., Viergever, M.A., van Vulpen, M., Pluim, J.P.W., 2010. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans. Med. Imaging 29, 2000–2008.

Liang, Z., He, X., Ceritoglu, C., Tang, X., Li, Y., Kutten, K.S., Oishi, K., Miller, M.I., Mori, S., Faria, A.V., 2015. Evaluation of cross-protocol stability of a fully automated brain multi-atlas parcellation tool. PLoS One 10, e0133533.

Lotjonen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., Alzheimer's Disease Neuroimaging, I., 2010a. Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage 49, 2352–2365.

Lotjonen, J.M.P., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., Initi, A.s.D.N., 2010b. Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage 49, 2352–2365.

Mai, J., Paxinos, G., Voss, T., 2007a. Atlas of Human Brain. Academic Press, San Diego.

Mai, J.K., Paxinos, G., Voss, T., 2007b. Atlas of the Human Brain. 3rd Edition. Academic Press.

Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium For Brain Mapping (ICBM). NeuroImage 2, 89–101.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B., 2001. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 1293–1322.

Mori, S., Wakana, S., Nagae-Poetscher, L.M., Van Zijl, P.C., 2005. MRI Atlas of Human White Matter. Elsevier, Amsterdam, The Netherlands.

Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A.V., Mahmood, A., Woods, R., Toga, A.W., Pike, G.B., Neto, P.R., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P., Mazziotta, J., 2008. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. NeuroImage 40, 570–582.

Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., Akhter, K., Hua, K., Woods, R., Toga, A.W., Pike, G.B., Rosa-Neto, P., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P.C., Mazziotta, J., Mori, S., 2008. Human brain white matter atlas: identification and assignment of common anatomical structures in superficial white matter. NeuroImage 43, 447–457.

Oishi, K., Faria, A., Jiang, H.Y., Li, X., Akhter, K., Zhang, J.Y., Hsu, J.T., Miller, M.I., van Zijl, P.C.M., Albert, M., Lyketsos, C.G., Woods, R., Toga, A.W., Pike, G.B., Rosa-Neto, P., Evans, A., Mazziotta, J., Mori, S., 2009. Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. NeuroImage 46, 486–499.

Oishi, K., Faria, A.V., Yoshida, S., Chang, L., Mori, S., 2013. Quantitative evaluation of brain development using anatomical MRI and diffusion tensor imaging. Int. J. Dev. Neurosci. 31, 512–524.

Pham, D.L., Prince, J.L., 1999. Adaptive fuzzy segmentation of magnetic resonance images. IEEE Trans. Med. Imaging 18, 737–752.

Puelles, L., Paxinos, G., Harrison, M., Watson, C., 2013. A developmental ontology for the mammalian brain based on the prosomeric model. Trends Neurosci. 36, 570–578.

Raz, N., Gunning, F.M., Head, D., Dupuis, J.H., McQuain, J., Briggs, S.D., Loken, W.J., Thornton, A.E., Acker, J.D., 1997. Selective aging of the human cerebral cortex observed in vivo: differential vulnerability of the prefrontal gray matter. Cereb. Cortex 7, 268–282.

Rohlfing, T., Russakoff, D.B., Maurer Jr., C.R., 2004. Performance-based classifier combination in atlas-based image segmentation using expectation–maximization parameter estimation. IEEE Trans. Med. Imaging 23, 983–994.

Sabuncu, M.R., Yeo, B.T.T., Van Leemput, K., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. IEEE Trans. Med. Imaging 29, 1714–1729.

Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N., Fox, N.C., 2003. A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. Arch. Neurol. 60, 989–994.

Shaw, P., Kabani, N.J., Lerch, J.P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., Clasen, L., Evans, A., Rapoport, J.L., Giedd, J.N., Wise, S.P., 2008. Neurodevelopmental trajectories of the human cerebral cortex. J. Neurosci. 28, 3586–3594.

Tang, X., Oishi, K., Faria, A.V., Hillis, A.E., Albert, M.S., Mori, S., Miller, M.I., 2013. Bayesian parameter estimation and segmentation in the multi-atlas random orbit model. PLoS One 8, e65591.

Tang, X., Yoshida, S., Hsu, J., Huisman, T.A., Faria, A.V., Oishi, K., Kutten, K., Poretti, A., Li, Y., Miller, M.I., Mori, S., 2014. Multi-contrast multi-atlas parcellation of diffusion tensor imaging of the human brain. PLoS One 9, e96985.

Toga, A.W., Mazziotta, J., 2002. Brain Mapping: The Methods. 2nd ed. Academic Press, San Diego, USA.

Tu, Z., Narr, K.L., Dollar, P., Dinov, I., Thompson, P.M., Toga, A.W., 2008. Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE Trans. Med. Imaging 27, 495–508.

Tustison, N.J., Cook, P.A., Klein, A., Song, G., Das, S.R., Duda, J.T., Kandel, B.M., van Strien, N., Stone, J.R., Gee, J.C., Avants, B.B., 2014. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. NeuroImage 99, 166–179.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273–289.

van Rikxoort, E.M., Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M.A., Pluim, J.P.W., van Ginneken, B., 2010. Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus. Med. Image Anal. 14, 39–49.

Wang, H., Suh, J.W., Das, S.R., Pluta, J., Craige, C., Yushkevich, P.A., 2012. Multi-atlas segmentation with joint label fusion. IEEE Trans. Pattern Anal. Mach. Intell. 35, 611–623.

Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23, 903–921.

Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. NeuroImage 45, S173–S186.

Worsley, K.J., Andermann, M., Koulis, T., MacDonald, D., Evans, A.C., 1999. Detecting changes in nonisotropic images. Hum. Brain Mapp. 8, 98–101.