

# Subject-Specific Sparse Dictionary Learning for Atlas-Based Brain MRI Segmentation

Snehashis Roy, Qing He, Elizabeth Sweeney, Aaron Carass, Daniel S. Reich, Jerry L. Prince, and Dzung L. Pham

**Abstract**—Quantitative measurements from segmentations of human brain magnetic resonance (MR) images provide important biomarkers for normal aging and disease progression. In this paper, we propose a patch-based tissue classification method from MR images that uses a sparse dictionary learning approach and atlas priors. Training data for the method consists of an atlas MR image, prior information maps depicting where different tissues are expected to be located, and a hard segmentation. Unlike most atlas-based classification methods that require deformable registration of the atlas priors to the subject, only affine registration is required between the subject and training atlas. A subject-specific patch dictionary is created by learning relevant patches from the atlas. Then the subject patches are modeled as sparse combinations of learned atlas patches leading to tissue memberships at each voxel. The combination of prior information in an example-based framework enables us to distinguish tissues having similar intensities but different spatial locations. We demonstrate the efficacy of the approach on the application of whole-brain tissue segmentation in subjects with healthy anatomy and normal pressure hydrocephalus, as well as lesion segmentation in multiple sclerosis patients. For each application, quantitative comparisons are made against publicly available state-of-the-art approaches.

**Index Terms**—Brain, dictionary, histogram matching, magnetic resonance imaging (MRI), patches, segmentation, sparsity.

## I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is a widely used noninvasive modality to image the human brain. Post-processing of MR images can provide information for understanding normal aging, as well as for monitoring and predicting many diseases. For example, quantitative measurements of brain tissues, such as gray matter (GM), white matter (WM),

or the ventricles are important biomarkers in aging, dementia, hypertension [1], while WM lesions and GM volumes are associated with the progression of Alzheimer’s disease and multiple sclerosis (MS) [2]. Therefore, segmentation of multiple tissues as well as lesions from MR images is important in research and potentially clinical settings. Beyond volumetric analysis, segmentation is also a fundamental step in many other image analysis procedures, such as voxel-based morphometry, cortical surface reconstruction, and disease classification.

In this paper, we describe a novel framework for dictionary-based multiclass segmentation of MR brain images. We call this method *subject-specific sparse dictionary learning*, or S3DL. This framework can be employed for a variety of binary or multilabel segmentation tasks. Here, we demonstrate its use for  $T_2$  lesion segmentation in MS patients, as well as brain tissue segmentation on healthy subjects and patients. Although measurements computed from these segmentations may have relevance in clinical research, we focus primarily on describing the mathematical framework from which the results are computed and evaluating its efficacy compared to similar algorithms.

Previous work on brain image segmentation has employed a variety of models to capture the intensity distributions of different anatomical structures. Finite mixture models are the basis of many image segmentation algorithms, where the intensity histogram is fitted with a number of distributions, such as Gaussian, (SPM [3], FSL [4], EMS [5], or LoAd [6]), or Rician [7]. Other algorithms model the tissue intensities using fuzzy C-means (FCM), such as FANTASM [8], [9] or TOADS [10], [11]. Methods such as FreeSurfer [12] incorporate prior information on the spatial locations of the tissue with statistical priors that capture their spatial variability.

Instead of trying to fit image intensities in individual voxels into predefined parametric models, example based approaches rely on exploiting similar looking patches from expert segmented images within training data. A patch is defined as a 3-D subimage centered around a voxel. Consisting of  $3 \times 3 \times 3$  or larger voxel neighborhoods, the use of patches allows modeling of local textural features. The combined use of training data and patch features generally provides superior results to the aforementioned models based on the intensity distributions of single voxels. The main idea of many example-based MR segmentation methods is patch-based label fusion [13], where the training data takes the form of an *atlas* consisting of MR images and their manual segmentation. It is implicitly assumed that a subject patch can be represented by a linear combination of many atlas patches found in a search window around that patch [14]–[17]. A patch from a subject MR image is first matched to similar looking relevant patches (or “examples”) from the atlas MR image by either nonlocal means [18] or sparse convex

Manuscript received January 14, 2015; revised April 28, 2015 and May 26, 2015; accepted May 27, 2015. Date of publication June 1, 2015; date of current version September 1, 2015. This work was supported by the Department of Defense at the Center for Neuroscience and Regenerative Medicine, and in part by the Intramural Research Program of the NINDS/NIH, and under Grant NIH/NINDS R01NS070906, Grant NIH/NIBIB R21EB012765, Grant NIH/NIBIB 1R01EB017743, Grant NIH/NIA T32AG021334, Grant NIH/NINDS R01NS085211, and Grant NIH/NINDS R01NS060910.

S. Roy, Q. He, and D. L. Pham are with the Center for Neuroscience and Regenerative Medicine, Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD 20817 USA (e-mail: snehashis.roy@nih.gov; qing.he@nih.gov; dzung.pham@nih.gov).

E. Sweeney is with the Department of Biostatistics, The Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: elizabethmargaretsweeney@gmail.com).

A. Carass and J. L. Prince are with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: aaron\_carass@jhu.edu; prince@jhu.edu).

D. S. Reich is with the Translational Neuroradiology Unit, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, Bethesda, MD 20824 USA (e-mail: daniel.reich@nih.gov).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JBHI.2015.2439242

combination [19], [20] via a patch dictionary [21]. Then corresponding label patches from the atlas segmentation image are combined similarly to estimate the label of the subject patch. The atlas is usually registered to the subject to facilitate the windowed search for relevant patches [22]. Such example-based patch matching algorithms have been extensively applied in binary segmentation of the hippocampus [22], amygdala [23], and ventricles [24] from  $T_1$ -weighted ( $T_1$ -w) brain MR images. Although three-class tissue segmentation methods have been proposed [19], [25] to segment cerebro-spinal fluid (CSF), GM and WM from  $T_1$ -w images, they do not distinguish between cortical and subcortical structures. A label-consistent multiclass dictionary-based segmentation method has also been proposed [26] for subcortical GM segmentation. As cortical and subcortical structures in the brain have similar intensity values in a  $T_1$ -w image, additional information is required to distinguish them.

Our S3DL framework is an example-based approach, using patches as features and utilizing training data in the form of an MR image with a known segmentation. It generalizes the approach proposed by [21] in that it employs dictionary learning to reduce the atlas size by selecting only the most relevant patches, leading to improved computational efficiency and accuracy. Unlike [21], S3DL can simultaneously segment multiple tissue classes while being informed by atlas priors. These priors inform the algorithm of where different anatomical structures are likely to occur within the image space. From the training data, a subject-specific dictionary is learnt using subject image features, so that it can represent the training subject patches optimally. Then for each subject patch, its weight is computed from the learnt dictionary using a sparse optimization procedure. Corresponding atlas hard segmentation labels are weighted to generate a tissue membership of the subject patch. In this way, we can perform binary or whole brain multiclass segmentation using the same framework. Our approach is, therefore, capable of segmenting structures with anatomically separate structures with similar intensities, such as cortical and subcortical GM.

Deformable registration-based label fusion techniques [27], [28] are often used to segment multiple tissue classes. These methods employ multiple atlases with labels that are first registered nonlinearly to a subject. The atlas label maps are then also transformed into the subject space and combined either by majority voting [29], probabilistic measures [30], [31], or local similarity criteria [32]. The performance of these methods typically depends on the accuracy of registrations between the subject and atlases. In addition, large numbers of atlases are often required, which increases the computational burden.

An advantage of our approach is that deformable registration between the subject and atlas is not needed, thereby avoiding the registration-related issues of computational expense and inaccurate correspondences that occur when the discrepancies between the atlas and the target are large. Subject priors are adaptively modified so as to account for the wide variability in the tissue shapes. A preliminary version of this paper has been reported earlier [33]. The following provides additional details of the algorithm and new results compared to our previous conference article.

## II. METHOD

This section is organized as follows. First, we define notation and describe in Section II-A how our approach utilizes an atlas consisting of the original MR image, its segmentation, and spatial priors for each tissue component. Then in Section II-B, we show how the atlas and subject images are used within a sparse patch matching framework to segment a subject image. Dictionary learning is used to reduce redundancy in the atlas and extract important patch features. In Section II-C, we show how the priors are iteratively updated to increase their relevance with respect to the subject data. Finally, we describe the dataset in Section II-D.

### A. Training data

Our algorithm minimally requires one set of MR images with a known segmentation to serve as training data. We call this training data an *atlas*. Formally, an *atlas* can be defined as a  $(m + 1)$ -tuple of images,  $\{a_1, \dots, a_{m+1}\}$ , where  $a_1$  to  $a_m$  denotes  $m$ -channel input MR images, such as  $T_1$ -w,  $T_2$ -w, or FLAIR. The image  $a_{m+1}$  denotes the hard segmentation with  $n$  multiple labels. For binary segmentations (e.g., lesion segmentation),  $n = 1$ . When  $n > 1$  and the desired labels have overlapping intensity distributions, prior information is desirable to differentiate structures based on their spatial location. In such cases, the atlas is augmented to include  $n$  priors,  $\{a_{m+1}, \dots, a_{n+m+1}\}$ , depicting the probability that each label could occur at each voxel location. In that case,  $a_{n+m+1}$  denotes the hard segmentation. These priors can be computed using a variety of approaches but practically, a simple blurring of the known atlas segmentation suffices [10]. An example of a single channel ( $m = 1$ )  $T_1$ -w atlas image  $a_1$ , tissue priors, and hard segmentation  $a_{n+m+1}$  are shown in Fig. 1. Additional priors for subcortical GM and WM would typically be available but are not shown in this figure. Although the notations and equations can be generalized for any  $m$ , we will assume single channel input ( $m = 1$ ) for the rest of this section to keep the notations succinct. However, in Section III-B, we use four channels,  $T_1$ ,  $T_2$ , PD, and FLAIR, i.e.,  $m = 4$ .

Given a subject MR image to be segmented, denoted by  $s_1$ , we must transform the atlas with respect to its geometric space and intensity scale to align it with the subject. Atlas  $a_1$  is affinely registered to  $s_1$ , and the priors  $\{a_2, \dots, a_{n+1}\}$  are transformed to the subject space by the same affine transformation [34]. Note that only a rough initial alignment is necessary because of the prior adaptation performed by S3DL, as described in Section II-C. The transformed priors are then multiplied by a scalar  $w$  to control their influence on the final segmentation and are denoted by  $\{s_2, \dots, s_{n+1}\}$ . MR images  $a_1$  and  $s_1$  are also intensity normalized so that the modes of their WM intensities are unity. The WM intensity modes were found automatically based on a smooth kernel density estimator of the histograms [35].

For every image, 3-D patches are transformed into 1-D vectors where  $d$  denotes the dimension of each patch from an image. For example, for  $3 \times 3 \times 3$  patches,  $d = 27$ . A subject

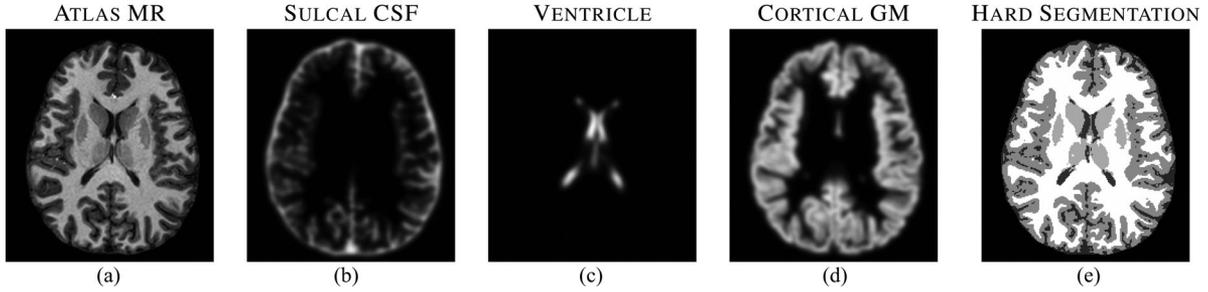


Fig. 1. (a) Atlas image, and its (b) sulcal CSF, (c) ventricle, (d) cortical GM priors, and (e) the hard segmentation.

patch-based feature is denoted by  $\mathbf{b}(j) \in \mathbb{R}^{(n+1)d}$ ,  $j = 1, \dots, N$ , which is the concatenation of subject MR and prior image (i.e.,  $s_1, \dots, s_{n+1}$ ) patches. Similarly, an atlas patch feature is the concatenation of patches from atlas MR and priors ( $a_1, \dots, a_{n+1}$ ). An atlas patch collection is defined as  $A_1 \in \mathbb{R}^{(n+1)d \times M}$ , where the  $i$ th column of  $A_1$  is an atlas feature.  $M$  and  $N$  denote the total number of atlas and subject patches. From now on, we also refer to the  $(n+1)d \times 1$  feature vector as a “patch.” All features are normalized so that their  $\ell_2$  norms are unity [19]. The atlas segmentation image is also decomposed into patches, which form the columns of the segmentation patch collection  $A_2$ .

### B. Sparse Dictionary Learning

The atlas data are used in a machine learning framework to segment the subject image by matching patch-based features between the subject images and the atlas images. Because the atlas has a known anatomy, features in the atlas that are similar to a subject feature contribute information about the anatomy of the subject. We assume that for every subject patch  $\mathbf{b}(j)$ , a small number of similar looking patches can always be found from the collection of atlas patches  $A_1$  [19], [20]. Sparse matching enforces the condition that every subject patch can be matched to only a few atlas patches, written as

$$\begin{aligned} \mathbf{b}(j) &\approx A_1 \mathbf{x}(j), \quad \text{for some } \mathbf{x}(j) \in \mathbb{R}^M \\ \|\mathbf{x}(j)\|_0 &\ll M \quad \forall j \text{ and } \mathbf{x}(j) \geq \mathbf{0} \end{aligned} \quad (1)$$

where  $\mathbf{x}(j)$  contains weights for each atlas patch in the collection. The weights are large where the corresponding atlas patch is similar to the subject patch. Previous methods have enforced the similarity in spatial locations by searching for the similar patches in a small window around the  $j$ th voxel [22]. We obviate the need for such windowed searching by including priors in the features. The nonnegativity constraints in the weight  $\mathbf{x}(j)$  enforces the similarity in *texture* between the subject patch and the chosen atlas patches.

The combinatorics of the  $\ell_0$  problem in (1) makes it infeasible to solve directly, but it can be transformed into an  $\ell_1$  minimization problem

$$\begin{aligned} \hat{\mathbf{x}}(j) &= \arg \min_{\mathbf{x} \geq \mathbf{0}} \{ \|\mathbf{b}(j) - A_1 \mathbf{x}(j)\|_2^2 + \lambda \|\mathbf{x}(j)\|_1 \} \\ \text{subject to } &\|\mathbf{f}(i)\|_2^2 = 1 \end{aligned} \quad (2)$$

$\mathbf{f}(i)$ ,  $i = 1 \dots, M$  are the columns of  $A_1$ . However,  $\mathbf{x}(j)$  is a  $M \times 1$  vector, where  $M$  is the number of atlas patches, typically  $M \sim 10^7$ . Thus, solving such a large optimization for every subject patch is computationally intensive. We use sparse dictionary learning to estimate a dictionary of smaller length  $D_1 \in \mathbb{R}^{(n+1)d \times L}$ ,  $L \ll M$ , from  $A_1$ , which can be used instead of  $A_1$  in (2) to solve for  $\mathbf{x}(j)$ .

The advantage of learning a dictionary is twofold. First, since the dictionary can represent a set of training patches optimally [36], it can, therefore, produce a lower reconstruction error than other approaches and achieve higher accuracy. This was shown in [21] where dictionary learning methods exhibited superior performance over nonlocal-means-based approaches. Second, the computational burden of (2) for every subject patch is reduced because of the obvious reduction in data size.

The dictionary learning approach we use is an alternating minimization algorithm [37] that solves the following problem:

$$\begin{aligned} &\{\hat{\mathbf{x}}(j), \hat{D}_1\} \\ &= \arg \min_{\mathbf{x} \geq \mathbf{0}, D_1} \sum_{j=1}^N \{ \|\mathbf{b}(j) - D_1 \mathbf{x}(j)\|_2^2 + \lambda \|\mathbf{x}(j)\|_1 \} \\ &\text{s.t. } \|\mathbf{f}(i)\|_2^2 = 1 \end{aligned} \quad (3)$$

where  $\mathbf{f}(i)$ ,  $i = 1 \dots, L$  now are the columns of  $D_1$ . Equation (3) can be solved in two alternating steps. First, keeping  $D_1$  fixed, we solve for  $\mathbf{x}(j)$  for each  $j$ , as in (2). Then keeping  $\mathbf{x}(j)$  fixed, we solve

$$\hat{D}_1 = \arg \min_{D_1} \sum_{j=1}^N \|\mathbf{b}(j) - D_1 \mathbf{x}(j)\|_2^2. \quad (4)$$

A gradient descent approach leads to the following update:

$$D_1^{(t+1)} = D_1^{(t)} + \eta \sum_{j=1}^N (\mathbf{b}(j) - D_1^{(t)} \mathbf{x}(j)) \mathbf{x}(j)^T \quad (5)$$

where  $\eta$  is the step size and  $t$  denotes iteration numbers. We note that  $\eta$  should be chosen carefully so that  $D_1^{(t)} > 0$ , since the columns of  $D_1$  contain MR intensities and priors.  $D_1^{(0)}$  is generated using  $L$  randomly chosen columns of  $A_1$ . The segmentation dictionary  $D_2$  is generated using the corresponding columns of  $A_2$ . Convergence is achieved when  $\|D_1^{(t+1)} - D_1^{(t)}\|_2^2 < 0.01$ . An example of a learnt dictionary is shown in Fig. 2, where  $D_1^{(0)}$

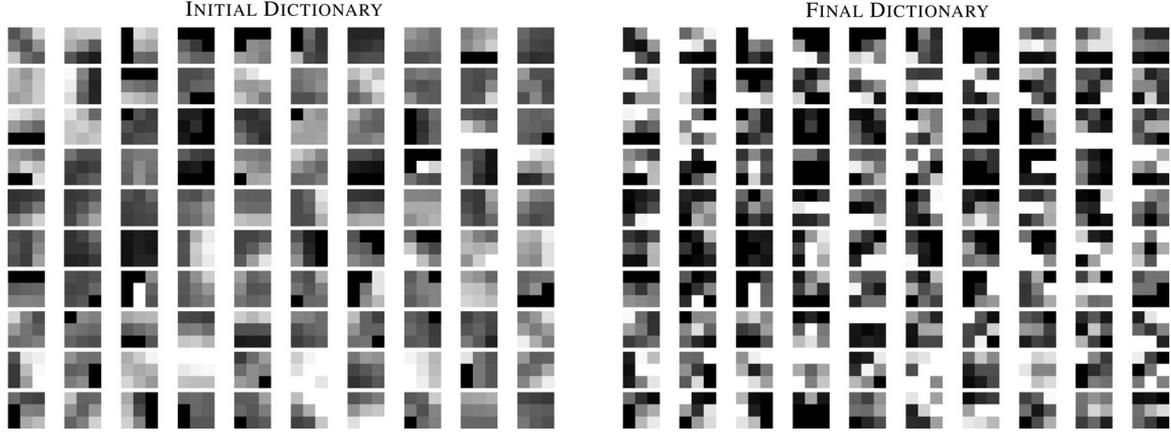


Fig. 2. Left image shows middle sections of 100 randomly chosen  $3 \times 3 \times 3$  patches from  $D_1^{(0)}$ , while on the right are the same atlas patches learnt from the subject after five iterations of (5).

is compared with  $D_1^{(5)}$ . Clearly, after learning from the subject patches, there are more edges in  $D_1^{(5)}$  patches, compared to the “flat”-looking patches in  $D_1^{(0)}$ .

Once the dictionary has been learned, (2) is solved for every subject patch  $\mathbf{b}(j)$  substituting  $\widehat{D}_1$  instead of  $A_1$  to solve for the atlas weights  $\mathbf{x}(j)$ . Because every atom in the learnt dictionary  $D_1$  has a corresponding segmentation patch in  $D_2$  that correspond to labels, these can be used to define a label for the subject patch. Empirically, we have found that the mean of  $\|\mathbf{x}(j)\|_1$  is close to unity and variance is usually very small ( $\sim 0.005$ ). Therefore, we can weight the segmentation labels according to their atlas weights in  $\mathbf{x}(j)$  to generate tissue memberships

$$\mathbf{p}_k(j) = (\mathbb{1}_{D_2}(k)) \frac{\mathbf{x}(j)}{\|\mathbf{x}(j)\|_1}, k = 2, \dots, n \quad (6)$$

where  $\mathbb{1}_{D_2}(k)$  denotes the indicator matrix having the same size as  $D_2$ , whose elements are 1 if the corresponding element in  $D_2$  is  $k$ , 0 otherwise. The index  $k = 2, \dots, n + 1$  denotes  $n$  unique tissue labels. Note that we only take the central voxel of  $\mathbf{p}_k(j)$  to generate the full membership image  $p_k$ .

### C. Adaptive Priors

The previously described steps produce a segmentation that is influenced by the geometry of the spatial priors. However, because the priors are derived from a different brain image, it may not be ideally suited to the subject image because of pathology or simply the variability of the brain geometry. Thus, instead of using a fixed prior based on the initial atlas-to-subject registration, we dynamically update within an iterative loop. The priors  $\{s_2, \dots, s_{n+1}\}$  at each iteration are replaced by a Gaussian blurred version of the obtained memberships  $p_k$ , similar to the approach of Shiee *et al.* [38]. The blurring relaxes the localization of the tissues in the memberships, allowing for greater freedom in the segmentation computed at the next step. The algorithm can be written as follows:

- 1) At  $t = 0$ , we start with  $\{a_1, \dots, a_{n+1}, s_1, s_2^{(0)}, \dots, s_{n+1}^{(0)}\}$ , where  $s_k^{(0)}$  are the registered atlas priors,  $k = 2, \dots, n + 1$ .

- 2) Generate dictionaries  $\widehat{D}_1$  and  $D_2$  from (5) using  $\{a_1, \dots, a_{n+1}, s_1, s_2^{(0)}, \dots, s_{n+1}^{(0)}\}$ . The subject patches are denoted by  $\mathbf{b}^{(0)}(j)$ .
- 3) At  $t \leftarrow t + 1$ , for each subject patch  $\mathbf{b}^{(t)}(j)$ , generate the sparse coefficient  $\mathbf{x}^{(t)}(j)$  using  $\widehat{D}_1$  from (2).
- 4) Generate memberships  $\{p_2^{(t)}, \dots, p_{n+1}^{(t)}\}$  using  $\mathbf{x}^{(t)}(j)$ s from (6).
- 5) Generate new adaptive priors  $s_k^{(t)} \leftarrow G_\sigma * p_k^{(t)}, k = 2, \dots, n + 1$ .  $\sigma = 3$  mm is chosen empirically.
- 6) Generate  $\mathbf{b}^{(t+1)}(j)$  using the updated priors  $\{s_1, s_2^{(t)}, \dots, s_{n+1}^{(t)}\}$ .
- 7) Stop if  $\frac{1}{N} \sum_{j=1}^N \|\mathbf{x}^{(t)}(j) - \mathbf{x}^{(t-1)}(j)\| < \epsilon$ , else go to step 3.

Fig. 3 shows the effect of iteratively updating the priors via memberships. Since the atlas is registered to the subject using affine registration, the strong GM prior in the middle of CSF (red boxes) introduces nonzero GM membership. However, because the prior is updated at each iteration, the effect of that misalignment is mitigated and the CSF memberships are increased in those areas.

### D. Evaluation

To evaluate and optimize the dictionary size selection, we employed simulated images from Brainweb [39]. Brainweb allows simulation of MR brain images with a known underlying anatomy. It further allows adjustment of the amount of noise, providing an excellent testbed for image segmentation algorithms.

We describe three applications of the method. First, to demonstrate the efficacy of only the patch-based dictionary learning approach without priors, we applied the algorithm to the segmentation of lesions from a pool of 122 subjects (57 male) with MS. The average age of the participants was 44 years (range 22–67), with an average disease duration nine years (range 0–38). Median expanded disability status scale across the patients was 3 (range 0–40). Subjects had  $T_1$ -w MPRAGE ( $0.82 \times 0.82 \times 1.1$  mm<sup>3</sup>),  $T_2$ -w, PD-w, and FLAIR scans

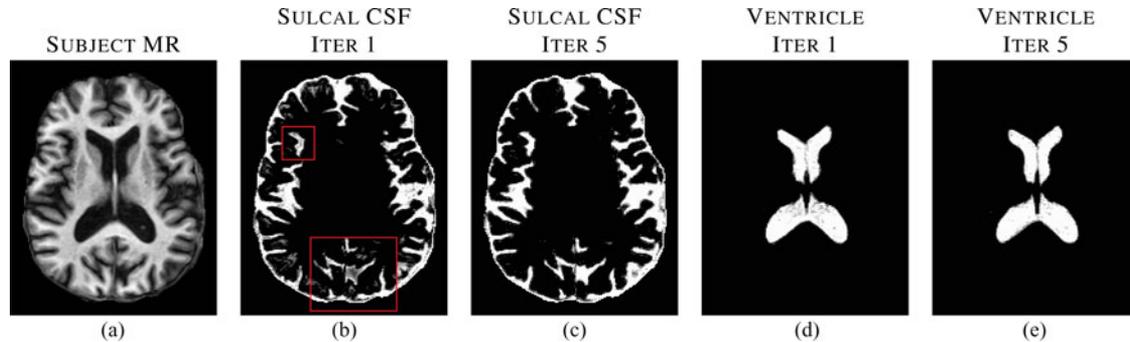


Fig. 3. (a) MR image of a subject, (b) and (c) sulcal CSF and (d) and (e) ventricle memberships ( $p_k$ ) for the first and fifth iterations of the algorithm (see Section II-C). Red boxes indicate where the CSF memberships change while updating the adaptive priors.

( $0.82 \times 0.82 \times 2.2 \text{ mm}^3$ ) acquired on a 3T Philips scanner (TR/TE=10.21/6 ms, flip angle  $8^\circ$ ), along with a manual delineation of lesions by an expert rater. We compare our method against two model-based lesion segmentation methods, LesionTOADS [11] (<http://www.nitrc.org/projects/toads-cruise>) and OASIS [40]. LesionTOADS is an FCM-based lesion segmentation method improving upon [10], where the lesions are modeled as outliers of the intensity distribution. OASIS uses a logistic regression model incorporating MR image intensities to produce a voxelwise membership for lesions.

Two datasets were used to validate the application of S3DL to multiclass brain segmentation. For these two datasets, the following tissue classes were assumed: GM, WM, subcortical GM, and ventricles. We tested the algorithm on 18 (14 male) publicly available IBSR (<http://www.nitrc.org/projects/ibsr>) subjects with normal anatomy, acquired with a  $T_1$ -w sequence. Image resolution varied between  $0.9375 \times 0.9375 \times 1.5 \text{ mm}^3$ ,  $1 \times 1 \times 1.5 \text{ mm}^3$ , and  $0.837 \times 0.837 \times 1.5 \text{ mm}^3$ . Mean age is 38 years (range 7–71). Manual segmentations of 43 structures are available for each subject. For the purpose of this study, we combined the labels into four classes, GM, WM, ventricles, and subcortical GM. Since sulcal CSF is usually underrepresented in the manual segmentations, it is not considered, as was done in [41]. We compared our method with two state-of-the-art model-based whole-brain segmentation methods, LoAd [6], and TOADS [10], and a nonlocal patch-based label fusion method (<http://volbrain.upv.es/>) [24]. TOADS uses statistical atlases in an FCM [42] framework, while LoAd uses Gaussian mixture models [5] with a refinement for better segmentation of convoluted cortical folds. The nonlocal-means-based method [24] finds relevant patches in a neighborhood, and combines the labels using weights based on a similarity criteria. Additional comparisons to FreeSurfer [12] are made in a prior conference publication [33].

In order to test the prior adaptation, we experimented on ten subjects with normal pressure hydrocephalus (NPH) (not publicly available), acquired with a structural  $T_1$ -w MPRAGE sequence at  $0.42 \times 0.42 \times 0.9 \text{ mm}^3$  resolution in a Siemens 3T scanner, (TR/TI/TE = 2110/1100/3.24 ms, flip angle  $8^\circ$ ). Many of these subjects exhibit substantial ventriculomegaly, which can be challenging to segment for atlas-based approaches. The ventricle segmentation on patients showed the effects of adaptive

priors when there is significant atrophy in the ventricles compared to a healthy subject. Manual delineations from an expert rater were available on only the ventricles, since ventricle volume is one of the important biomarkers for the disease. We show that when the anatomy between the atlas and the subject is significantly different (i.e., enlarged ventricles), our method is still capable of obtaining accurate ventricular segmentations.

For all the data,  $T_1$ -w scans were first registered to the MNI atlas [43] by rigid registration via FLIRT [34], then corrected for intensity inhomogeneity by N4 [44] and skull-stripped using SPECTRE [45]. For the MS data,  $T_2$ , PD, and FLAIR images were aligned to the  $T_1$ -w image by a rigid registration.

Quantitative evaluation was carried out using the Dice coefficient [46] as the metric for similarity between whole-brain segmentations. For lesion segmentations, we also use absolute lesion volume difference (VD), lesion true positive rate (LTPR), lesion false positive rate (LFPR), and average symmetric surface distance (SD) [47], [48]. VD is the absolute VD between the automatic and manual segmentations divided by the volume of manual segmentation. LTPR is measured by dividing the number of lesions in the manual segmentation that overlap with a lesion in the automatic one with the number of overall lesions in the truth. This evaluates whether all lesions are detected or not. Note that rather than using a lesion segmentation mask to calculate the sensitivity, we use lesion count, which is independent of boundary accuracy. LFPR is measured by dividing the number of lesions in the automatic segmentation that do not overlap with any lesion in the truth with the number of lesions in the automatic segmentation. To calculate SD, the boundary voxels of the segmentation and truth are first determined. For each voxel on one boundary, the closest voxel on the other boundary is determined using unsigned Euclidean distance. The average of all these distances, i.e., between truth and segmentation and vice versa, gives the averages symmetric SD. More details on these metrics can be found in [48].

### III. RESULTS

S3DL was implemented in MATLAB (R2013a, The MathWorks, Natick, MA, USA) using parallel computation. For whole-brain segmentation (as in Sections III-C and III-D), the total runtime was typically 20 min on 2.7 GHz 12-core AMD

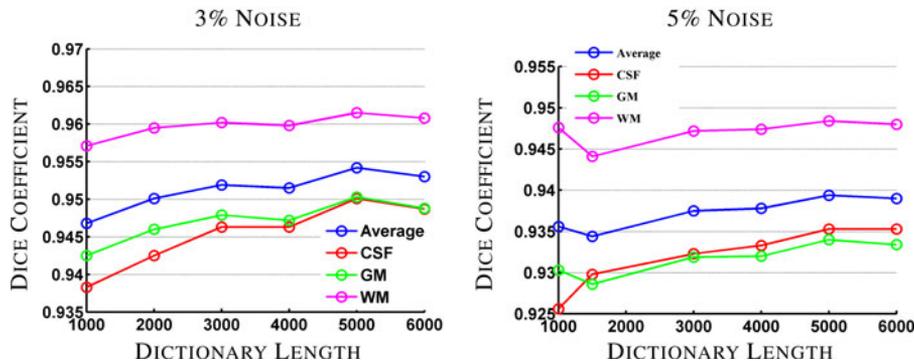


Fig. 4. Dice coefficients of three tissue classes and their weighted average between the true segmentation and dictionary based segmentation for phantoms with 3% and 5% noise are shown. The atlas is chosen as the phantom with 0% noise.

processors for one  $181 \times 217 \times 181$  sized  $1 \text{ mm}^3$  image, of which approximately 10 min were spent on learning the dictionary. The runtime is reduced to 15 min for lesion segmentation (see Section III-B), for which the dictionary learning takes approximately 5 min. In comparison, OASIS [40] and LesionTOADS [11] take about 5 and 20 min, respectively. SparseLab [49] was used to solve (2). We used  $3 \times 3 \times 3$  patches in all our experiments. To choose optimal  $w$ , we segmented one of the IBSR subjects using a range of  $w \in [0.01, 0.21]$  and found that the segmentation is stable in the range of  $[0.05, 0.15]$ ; therefore, we chose the prior weight  $w$  as 0.10. More details are available in the supplemental material. This selection ensures the priors do not have too much influence on the segmentation when the anatomy is widely different between subject and atlas, as illustrated in NPH patients, described in Section III-D. The parameters  $\lambda$  in (2) and  $\eta$  in (5) were chosen as 0.01.

#### A. Effect of Dictionary Size

To determine the effect of dictionary size ( $L$ ) on the segmentation, we varied  $L$  while segmenting Brainweb simulated images [39]. If the dictionary size is too small, the dictionary atoms may not represent the spectrum of subject patches well. If it is too large, the computation time increases. Furthermore, it has been shown that if the learnt dictionary  $\hat{D}_1$  is “incoherent” (i.e., off-diagonal entries of the matrix  $\hat{D}_1^T \hat{D}_1$  are sufficiently small), then it is possible to optimally reconstruct an unknown patch  $\mathbf{b}_j$  via (2) (with  $A_1 \equiv \hat{D}_1$ ) if it is a noisy perturbation of one or more patches from  $\hat{D}_1$  [36]. A proper dictionary size, therefore, ensures more accurate representation of the atlas.

We chose  $a_1$  as a Brainweb phantom with 0% noise and  $s_1$  as the same phantom with either 3% or 5% noise, with 0% inhomogeneity.  $\{a_2, a_3, a_4\}$  were chosen as CSF, GM, and WM tissue memberships, and  $a_5$  is the true three-class segmentation, available also from Brainweb. Priors were not used in this experiment. We segmented  $s_1$  using various dictionary sizes ranging from 1000 to 6000 and compared against the known segmentation. Dice coefficients of three tissue classes CSF, GM, and WM, and their weighted average (weighted by the corresponding volumes) between  $a_5$  and the segmentations are shown in Fig. 4. The average Dice coefficient increases with dictionary size, but near  $L = 5000$ , it plateaus at 0.965 for 3% noise (or

0.94 for 5% noise). The standard deviations of Dice coefficients, based on a set of ten instances of the phantom with 5% noise, are 0.007, 0.003, 0.006, for CSF, GM and WM. A similar experiment on IBSR data is performed where one subject is segmented with multiple dictionary lengths using another subject as atlas. Based on the accuracy of segmentation as well as to keep runtime short, an optimal dictionary size of 5000 was also observed. More details are available in the supplemental material. We chose a dictionary size  $L = 5000$  for the remainder of the experiments.

#### B. Lesion Segmentation

In this section, we validated only the dictionary learning aspect without priors on a binary segmentation application, similar to [21].  $T_1$ -w MPRAGE,  $T_2$ , PD-w, and FLAIR images are used to segment WM lesions from MS patients. Using patches from the subject and the atlas, a dictionary was learnt via (5), lesion memberships were obtained as before via (6), and a threshold was established to obtain a lesion mask. To reduce lesion false positives, multiple atlases were used to generate lesion memberships and they were averaged to create a more robust estimate. Selection of the number of atlases is described below.

Since only the WM lesions are of interest, it is possible to add a WM atlas prior to automatically detect lesions within WM. However, adding more features increases memory and runtime; therefore, we simply obtained the lesion segmentations by masking the membership images with a WM mask obtained from LesionTOADS [11] to remove false positives, since we are interested in segmenting WM MS lesions.

We used cross validation to find the optimal threshold for generating lesion masks. Four random subjects were chosen from the pool of 122 patients with MS. For each subject, three memberships were generated using the other three as atlases and were averaged to generate a mean membership. The mean membership was thresholded at various thresholds and average Dice coefficients (averaged over four subjects) are plotted in Fig. 5(a). The maximum Dice is obtained at a threshold of 0.80, which is used for the rest of the experiments.

To demonstrate the effect of the number of atlases on the accuracy of the lesion segmentation, we chose the previous four subjects as atlases and generated four memberships for each

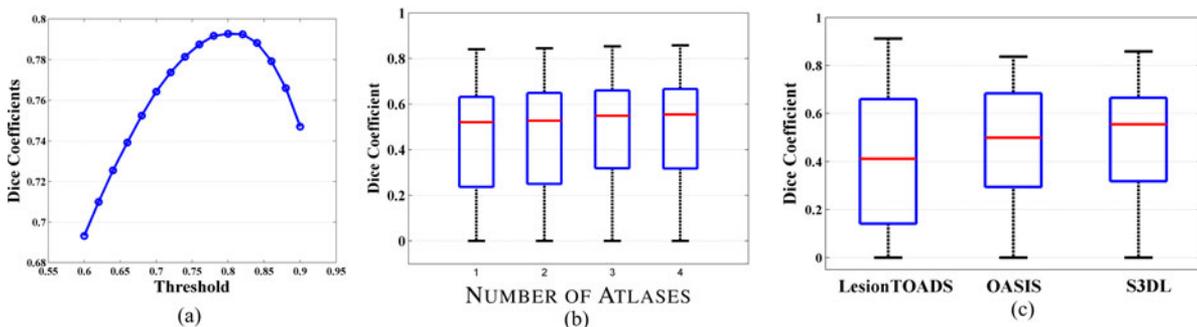


Fig. 5. (a) Average Dice coefficients (for four MS subjects) between manual lesion segmentation and thresholded S3DL generated memberships are plotted for different thresholds on the lesion memberships. (b) Boxplots show Dice coefficients for 118 MS subjects between manual lesion segmentation and thresholded mean lesion memberships. Horizontal axis shows number of atlas the memberships are averaged over, ranging from one to four. The median Dice coefficients are 0.521, 0.528, 0.549, and 0.555 for one to four atlases. (c) Boxplots show Dice coefficients between manual lesion segmentation and three automatic methods, LesionTOADS [11], OASIS [40], and S3DL with four atlases. The median Dice coefficients are 0.411, 0.499, and 0.555, respectively.

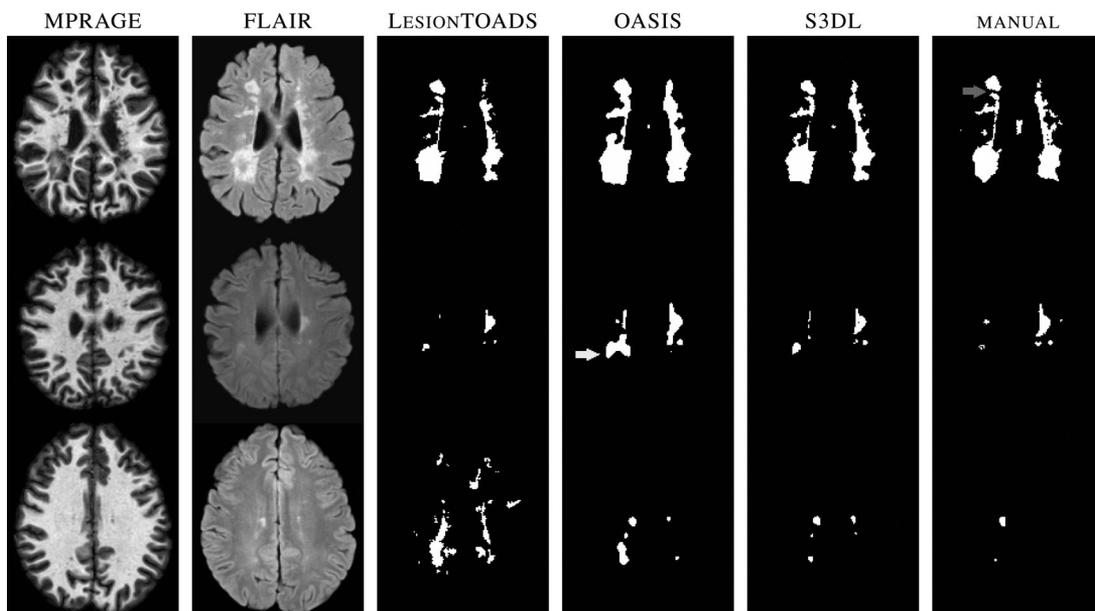


Fig. 6. Examples of lesion segmentation on three subjects with MS are shown in each row. MPRAGE, FLAIR,  $T_2$ -w and PD-w images are used for segmentation, although only MPRAGE and FLAIR scans are shown. Top row shows a subject where all three methods perform comparably, although OASIS produces smooth segmentation which in turn overestimates subtle changes in lesions (red arrow). Middle row shows an example where OASIS grossly overestimates lesions (yellow arrow). Bottom row shows gross over-estimation of LesionTOADS.

of the remaining 118 subjects. Dice coefficients between the thresholded mean segmentations (threshold = 0.80), averaged over number of atlases ranging from one to four, are shown in Fig. 5(b). The median Dice coefficients are 0.521, 0.528, 0.549, and 0.555, respectively, for the one to four atlases used to average memberships, respectively. Both three and four atlases produce significant improvement in Dice compared to using one or two atlases ( $p < 10^{-4}$  via Wilcoxon signed-rank test), although the Dice improvement in itself is small between three to four atlases (0.006). Subsequent comparisons against other methods are reported using four atlases.

Examples of lesion segmentations are shown in Fig. 6 for three methods, LesionTOADS [11], OASIS [40], and S3DL. The optimal parameters for OASIS were found by cross validation on the same four subjects used for training. LesionTOADS

being an unsupervised method does not need any training data. However, an optimized set of “compactness parameters” [9] were used for LesionTOADS, accounting for the amount of lesions present in the subject. OASIS generally produces the most smooth segmentations compared to LesionTOADS and S3DL, and sometimes cannot detect subtle changes in lesion boundary (see top row in Fig. 6). If the lesion load is high, all three methods are comparable, e.g., top row in Fig. 6, where the Dice coefficients are 0.589, 0.634, and 0.670 for LesionTOADS, OASIS, and S3DL, respectively. If the lesion load is low, LesionTOADS generally overestimates lesions, e.g., bottom row of Fig. 6, Dice coefficients 0.081, 0.153, 0.307.

Fig. 5(c) shows a comparison of Dice coefficients between manual and automated segmentations. The median Dice coefficients are 0.411, 0.499, and 0.555, for LesionTOADS, OASIS,

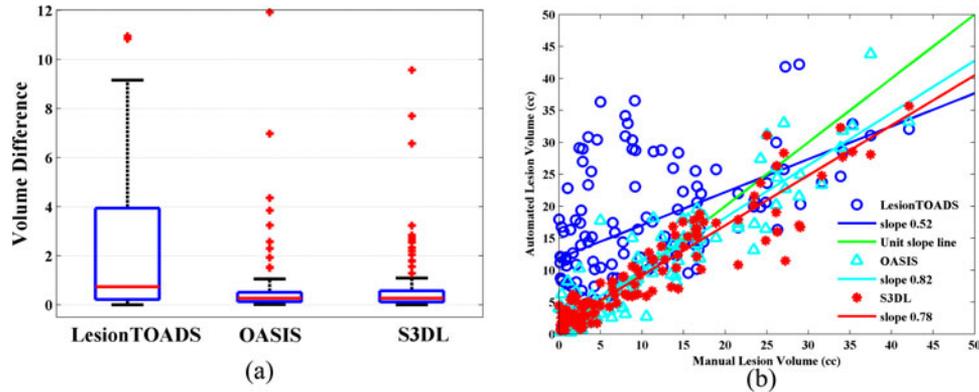


Fig. 7. (a) Boxplots show VD for 118 MS subjects for three automatic methods, LesionTOADS, OASIS, and S3DL. The median VDs are 0.729, 0.268, and 0.267, respectively. (b) Automatic lesion volumes are plotted against manual lesion volumes. Each point indicates a subject. Solid lines are best linear fits of the scatterplots.

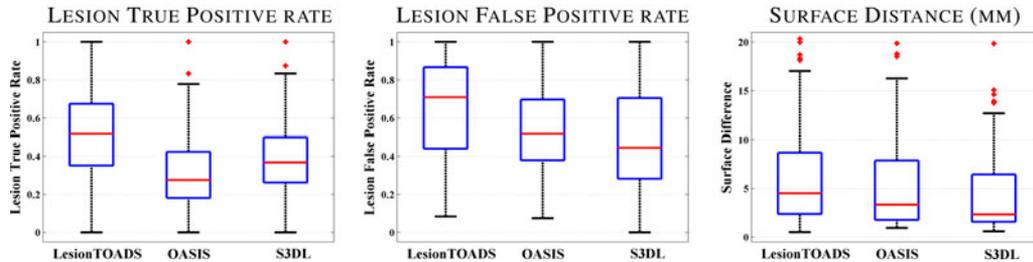


Fig. 8. Boxplots show LTPR, LFPR, and absolute symmetric SDs in millimeters between automatic segmentation and manual segmentation for 118 MS subjects. (See Section II-D for definitions.)

and S3DL, respectively. A nonparametric Wilcoxon sign-rank test shows that S3DL is significantly better than the others ( $p < 0.01$  for both). We note that the magnitude of Dice coefficients are relatively low, and they are even exactly zero for two subjects, which have very small lesion load. Since lesions are small objects with a highly ambiguous boundary, Dice may not be an appropriate measure of accuracy, when lesion load is small [48].

Fig. 7(a) shows the lesion VD. Ideally, if the segmentation is perfect, VD should be zero. The median VD for LesionTOADS, OASIS, and S3DL are 1.014, 0.268, and 0.267, respectively, indicating that LesionTOADS, OASIS, and S3DL differs from true lesion volumes by 101%, 26.8%, and 26.7% on an average. A nonparametric test reveals that VD is significantly lower in OASIS and S3DL ( $p < 10^{-6}$ ) compared to LesionTOADS, while there is no significant difference ( $p > 0.05$ ) between OASIS and S3DL. Fig. 7(b) shows scatter plots of manual versus automatic lesion volumes for each subject. Solid green, magenta, and red lines indicate best linear fits of the points. The intercepts are 11.87, 1.86, and 1.33 cc, for LesionTOADS, OASIS, and S3DL, respectively. The slopes of the best linear fits are 0.518, 0.818, 0.783. From Fig. 7(b), it is also evident that when lesion load is small (less than 15 cc), LesionTOADS overestimates the lesion most, followed by OASIS, and S3DL. Median lesion volume on the manual segmentations was 7.79 cc, while it was 15.99, 7.10, and 6.28 cc for LesionTOADS, OASIS, and

S3DL, with p-values being  $2 \times 10^{-10}$ , 0.091, and 0.144, respectively. Therefore, S3DL and OASIS produce statistically similar lesion volumes with the manual, while LesionTOADS produces significantly higher lesion volume.

We plotted the LTPR and LFPR in Fig. 8. LTPR denotes the ratio between number of detected lesions and number of overall lesions in the truth. This metric evaluates if all true lesions are detected. LTPR for S3DL is significantly higher than OASIS ( $p < 10^{-4}$ ), but significantly lower than LesionTOADS ( $p < 10^{-4}$ ), with median values being 0.516, 0.275, and 0.368, respectively. The median LFPR for the three methods are 0.708, 0.516, and 0.444, respectively. S3DL produces significantly lower LFPR ( $p < 10^{-4}$ ) compared to both OASIS and LesionTOADS. We also examined the performance of the three algorithms as a function of low, medium, and high lesion volume. S3DL had at least equivalent or significantly better performance in all three classes of lesion load compared to the other algorithms in terms of Dice and LFPR.

Average symmetric SDs are also shown in Fig. 8. The median SDs are 4.50, 3.32, and 2.32 mm for LesionTOADS, OASIS, and S3DL, respectively. S3DL has significantly lower SD than the other two methods ( $p < 10^{-6}$ ). The maximum SDs for all three methods are more than 25 mm, indicating some gross errors in the segmentations. These errors mostly occurs near the cerebellum and frontal lobe, where the FLAIR images can have hyperintensities in the cortex.

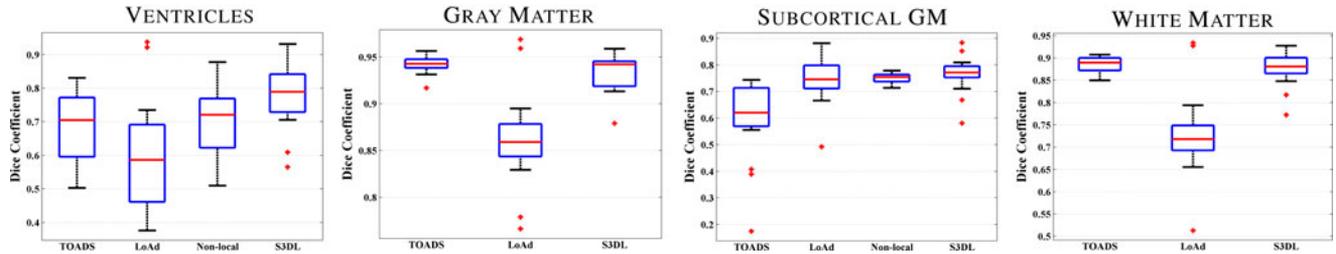


Fig. 9. Dice coefficients between truth segmentation and automatically generated segmentations for four methods, TOADS [10], LoAd [6], nonlocal-patch-based method [24], and S3DL, are shown for four tissue classes on 18 IBSR data.

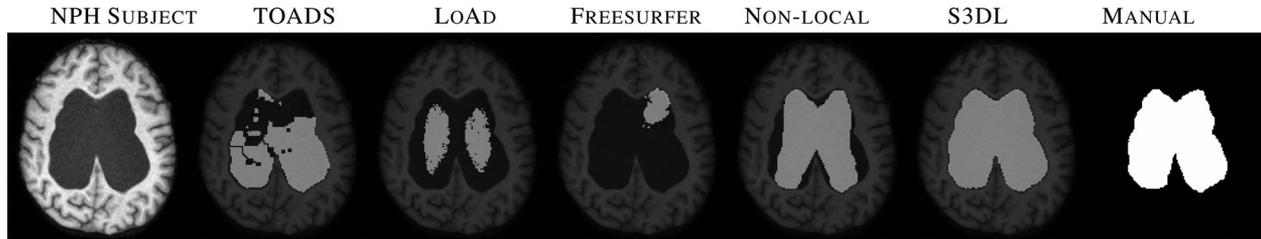


Fig. 10. Subject with NPH is segmented using TOADS [10], LoAd [6], Freesurfer [12], nonlocal method [24], and S3DL. Rightmost column shows manual delineation of the ventricles.

### C. Experiments on IBSR Data

We evaluated the algorithm on 18 subjects from IBSR V2 database (<http://www.nitrc.org/projects/ibsr>), which have manually segmented labels. We segmented the images into five classes, sulcal CSF, GM, WM, ventricle and subcortical GM (e.g., caudate, putamen, thalamus etc), and combined sulcal CSF and GM memberships together after segmentation because the provided manual segmentations excludes most of the sulcal CSF voxels [41]. Visually, the dataset contains two groups with different contrasts, 6 MP-RAGE-type and 12 SPGR-type contrasts. Since any example-based method relies on “similar looking” patches from an atlas, the atlas should be carefully chosen so that  $a_1$  and  $s_1$  have similar tissue contrast.

Two subjects, one with SPGR-type and one with MRAGE-type contrast, were randomly chosen as atlases for the two groups. Gaussian blurred tissue label-masks ( $\sigma = 3$  mm) were used as priors. The remaining subjects were segmented using the appropriate corresponding atlases. Dice coefficients comparing the four methods, TOADS [10], LoAd [6], nonlocal-patch-based method [24], and S3DL on the tissue classes are shown in Fig. 9. Note that the nonlocal method, as provided by the website, only produces subcortical GM and ventricle segmentations. We did not compare with Freesurfer, since Freesurfer has previously been evaluated on the IBSR data (average Dice coefficients of 0.584 and standard deviation of 0.057) [50]. S3DL outperforms the other three methods significantly in the ventricles, ( $p < 0.01$ ), using the Wilcoxon signed-rank test. Similarly, S3DL produces significantly larger Dice coefficient for subcortical GM than TOADS and LoAd ( $p = 0.0002$  and  $0.03$ , respectively), but is similar to the nonlocal method ( $p = 0.10$ ). The median Dice for ventricles and subcortical GM for TOADS, LoAd, nonlocal method, and S3DL are 0.705, 0.587, 0.721, 0.790 and

0.621, 0.746, 0.754, 0.772, respectively. S3DL is significantly better than LoAd on WM and GM segmentation ( $p < 0.01$ ), but gives similar segmentation to TOADS ( $p = 0.07$  for GM and  $p = 0.06$  for WM). The median Dice for GM and WM for TOADS, LoAd, and S3DL are 0.943, 0.859, 0.942 and 0.890, 0.718, 0.881, respectively.

### D. Experiments on NPH Data

We applied S3DL on ten subjects with NPH having enlarged ventricles. Fig. 10 shows one subject with five automated segmentations and the manual delineation of the ventricles. Visually, S3DL produces most similar segmentation to the manual. Freesurfer segments most of the ventricles as non-WM hypointensities, while LoAd segments part of the ventricles are sulcal CSF. The atlas is chosen to be a healthy volunteer of age 50 years. We empirically selected the `relaxation_factor` parameter of LoAd for each subject separately so that best ventricle segmentation was obtained. The relaxation factor adaptively smooths (or relaxes) the tissue memberships in a patient-specific manner, similar to our adaptive atlases. The Markov random field prior (`mrf_beta`) parameter of LoAd was chosen to be 0.25. They are also segmented with Freesurfer [12] using the `-bigventricles` switch. Manual segmentations were available only for the ventricles.

To demonstrate the effect of adaptive priors, Fig. 11(a) shows the ventricle memberships of one subject for the first three iterations. The first iteration does not include the adaptive priors; therefore, the membership is low ( $< 0.5$ ) deep inside the ventricles. Note that the membership is nonzero inside the ventricles of the subject, even where the atlas allows only WM tissue. This is due to the fact that patches have both intensity and priors as

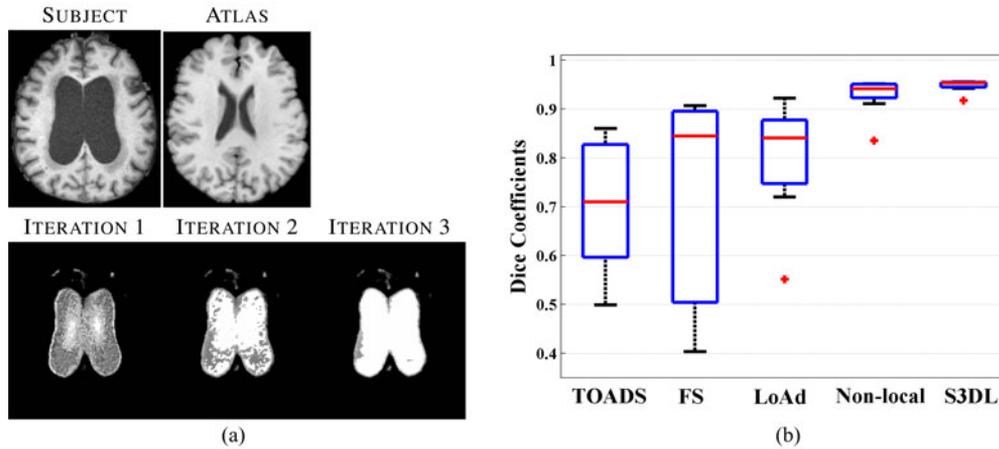


Fig. 11. (a) Ventricle memberships are shown for the first three iterations of the algorithm, along with the subject and a representative slice from the atlas. (b) Dice coefficients of between manual ventricle segmentation and four automatic methods are shown for ten subjects with NPH.

features; but unlike a mixture model, the priors are not multiplicative. Therefore, a zero prior does not necessarily indicate a zero membership.

Quantitative improvement is shown in Fig. 11(b) where Dice coefficients between manually segmented ventricles and automatic segmentations are plotted for the three methods. S3DL produces the largest Dice coefficient (median 0.954) across all subjects with very little variance, and it is significantly ( $p < 0.01$ ) larger than TOADS (median Dice 0.710), Freesurfer (median Dice 0.843), LoAd (median Dice 0.840), and the nonlocal method (median Dice 0.942). Median ventricle volume of the manual segmentations is 145.5 cc, while it is 78.5, 124.1, 118.8, 136.5, and 172.0 cc for TOADS, Freesurfer, LoAd, nonlocal method, and S3DL, respectively. The p-values are  $6 \times 10^{-4}$ , 0.005, 0.001, 0.94, and 0.97, indicating that TOADS, Freesurfer, and LoAd produce significantly lower volumes than the manual segmentation, while the nonlocal method and S3DL do not have any statistically significant VD.

#### IV. DISCUSSION AND CONCLUSION

We have presented a patch-based sparse dictionary learning method for binary or multiclass MR brain image segmentation. The use of patches over single voxel intensities provides improved discrimination of anatomical structures. Contrary to previous patch-based segmentation methods, we use adaptive priors to localize different tissues with similar intensities as well as capture wide variabilities in anatomy between a subject and an atlas. We do not require any deformable registration of the subject to the atlas.

Algorithms for the automatic detection of cerebral WM lesions are an important tool for understanding the progression of MS. Since lesions are often hyperintense in FLAIR images and hypointense in  $T_1$ -w MPRAGE images, most methods for lesion segmentation [38], [51]–[53] estimate lesions as outliers of joint intensity distributions. On the other hand, example-based methods estimate a regression from intensity features of a training MPRAGE and FLAIR pair of images to

their manual segmentation label, which is often binary [40], [47], [54], [55]. Recently, patch-based lesion detection methods have been shown to out-perform a parametric model-based method [56], [57].

Model-based lesion segmentation methods, such as LesionTOADS, often find a single threshold from the joint intensity distribution of MR images to identify lesions. If the intensity of a voxel is below a certain intensity threshold in FLAIR, it can never be segmented as a lesion. In contrast, example-based methods look for similar patches (incorporating the neighborhood information via 3-D patches as well) from the dictionary and fuse the labels as delineated by human raters. Thus, even if there is a patch with intensities below such a threshold (obtained from a model-based method), it can have high lesion membership if there are enough similar examples in the dictionary.

Although S3DL showed improved accuracy over LesionTOADS and OASIS on most segmentation scores, the LTPR was worse. A caveat of the LTPR measurement is that even if there are additional lesions in the segmentation (as sometimes is the case in overestimation), it is possible to have the best score of 1. We believe that the improved LTPR of LesionTOADS over S3DL is because they tend to overestimate lesions. On the other hand, LFPR represents the fraction of detected lesions that are not present in the truth. A caveat of this metric is that if a method underestimates lesions, it is possible to get the best score of 0. Also, LFPR does not provide an idea of the volume of false positives.

A current limitation of the lesion segmentation method is that for a new dataset from a different scanner or sequence, a few atlases need to be manually delineated. Then the optimal membership threshold and the number of atlases need to be reestimated. The number of training atlases can be small though, as we used four in our experiments. However, we have shown that improvement in segmentation from one to four atlases is fairly small, with median Dice coefficient increasing only by 0.03 in case of lesion segmentation.

We have currently optimized the brain segmentation algorithm for five classes. Segmentations of other neuroanatomical

structures such as the hippocampus, amygdala, or specific gyri are certainly possible using the method. However, additional atlases, and an extended set of features would be required given the wide variability of the human anatomy and the ambiguous boundaries of such structures.

The results on NPH data demonstrate the advantage of prior adaptation. The difference in anatomy between the atlas and the subject is widely variable; therefore, the fixed priors cannot capture the range. However, the S3DL priors are adaptively modified at each iteration to account for the wide variability. This allows S3DL to produce accurate ventricle segmentations even though such geometry is not present in the atlas.

Our approach learns a new “subject specific” dictionary for every subject. To improve runtime of the algorithm, a dictionary can be learnt on one image and then used to segment others. However, this is possible only for a set of images acquired with the same imaging sequence and at the same scanner and site. Therefore, if the noise level on the images is similar, we expect that one trained dictionary might be suitable across a class of subject data.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. P. Calabresi for providing access to the MS imaging data and Ms. J. L. Cuzzocreo for her help with lesion delineations.

#### REFERENCES

- [1] S. M. Resnick, D. L. Pham, M. A. Kraut, A. B. Zonderman, and C. Davatzikos, “Longitudinal magnetic resonance imaging studies of older adults: A shrinking brain,” *J. Neurosci.*, vol. 23, no. 8, pp. 3295–3301, 2003.
- [2] N. Shiee, P. L. Bazin, K. M. Zackowski, S. K. Farrell, D. M. Harrison, S. D. Newsome, J. N. Ratchford, B. S. Caffo, P. A. Calabresi, D. L. Pham, and D. S. Reich, “Revisiting brain atrophy and its relationship to disability in multiple sclerosis,” *PLoS One*, vol. 7, no. 5, p. e37049, 2012.
- [3] J. Ashburner and K. J. Friston, “Unified segmentation,” *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.
- [4] Y. Zhang, M. Brady, and S. Smith, “Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm,” *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [5] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, “Automated model-based tissue classification of MR images of the brain,” *IEEE Trans. Med. Imag.*, vol. 18, no. 10, pp. 897–908, Oct. 1999.
- [6] M. J. Cardoso, M. J. Clarkson, G. R. Ridgway, M. Modat, N. C. Fox, and S. Ourselin, “LoAd: A locally adaptive cortical segmentation algorithm,” *NeuroImage*, vol. 56, no. 3, pp. 1386–1397, 2011.
- [7] S. Roy, A. Carass, P.-L. Bazin, and J. L. Prince, “A Rician mixture model classification algorithm for magnetic resonance images,” in *Proc. Int. Symp. Biomed. Imag.*, 2009, pp. 406–409.
- [8] D. L. Pham, “Spatial models for fuzzy clustering,” *Comput. Vis. Image Understand.*, vol. 84, no. 2, pp. 285–297, 2001.
- [9] S. Roy, H. Agarwal, A. Carass, Y. Bai, D. L. Pham, and J. L. Prince, “Fuzzy c-means with variable compactness,” in *Proc. Int. Symp. Biomed. Imag.*, 2008, pp. 452–455.
- [10] P. L. Bazin and D. L. Pham, “Homeomorphic brain image segmentation with topological and statistical atlases,” *Med. Image Anal.*, vol. 12, no. 5, pp. 616–625, 2008.
- [11] N. Shiee, P. Bazin, A. Ozturk, D. S. Reich, P. A. Calabresi, and D. L. Pham, “A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions,” *NeuroImage*, vol. 49, no. 2, pp. 1524–1535, 2009.
- [12] A. M. Dale, B. Fischl, and M. I. Sereno, “Cortical surface-based analysis I: Segmentation and surface reconstruction,” *NeuroImage*, vol. 9, no. 2, pp. 179–194, 1999.
- [13] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, “Image analogies,” in *Proc. Conf. Comput. Graph. Interactive Tech.*, 2001, pp. 327–340.
- [14] S. Roy, A. Carass, and J. L. Prince, “A compressed sensing approach for MR tissue contrast synthesis,” in *Proc. 22nd Int. Conf. Inf. Proc. Med. Imag.*, 2011, pp. 371–383.
- [15] S. Roy, A. Jog, A. Carass, and J. L. Prince, “Atlas-based intensity transformation of brain MR images,” in *Multimodal Brain Image Analysis*. New York, NY, USA: Springer-Verlag, 2013, pp. 51–62.
- [16] S. Roy, A. Carass, and J. L. Prince, “Synthesizing MR contrast and resolution through a patch matching technique,” *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 7623, p. 76230j, 2010.
- [17] A. Jog, S. Roy, A. Carass, and J. L. Prince, “Magnetic resonance image synthesis through patch regression,” in *Proc. Int. Symp. Biomed. Imag.*, 2013, pp. 350–353.
- [18] A. Buades, B. Coll, and J. M. Morel, “A non-local algorithm for image denoising,” in *Proc. Int. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 2, pp. 60–65.
- [19] S. Roy, A. Carass, and J. Prince, “Magnetic resonance image example-based contrast synthesis,” *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2348–2363, Dec. 2013.
- [20] T. Cao, C. Zach, S. Modla, D. Powell, K. Czymbek, and M. Niethammer, “Registration for correlative microscopy using image analogies,” in *Proc. Workshop Biomed. Image Registration*, 2012, pp. 296–306.
- [21] T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, D. Rueckert, and The Alzheimer’s Disease Neuroimaging Initiative, “Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling,” *NeuroImage*, vol. 76, no. 1, pp. 11–23, 2013.
- [22] P. Coupé, S. F. Eskildsen, J. V. Manjon, V. S. Fonov, D. L. Collins, and the Alzheimer’s disease Neuroimaging Initiative, “Simultaneous segmentation and grading of anatomical structures for patient’s classification: Application to Alzheimer’s disease,” *NeuroImage*, vol. 59, no. 4, pp. 3736–3747, 2012.
- [23] H. Wang, J. W. Suh, S. R. Das, J. Pluta, C. Craige, and P. A. Yushkevich, “Multiatlas segmentation with joint label fusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 611–623, Jun. 2012.
- [24] P. Coupé, J. V. Manjon, J. Pruessner, M. Robles, V. Fonov, and D. L. Collins, “Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation,” *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.
- [25] L. Wang, F. Shi, Y. Gao, G. Li, J. H. Gilmore, W. Lin, and D. Shen, “Integration of sparse multimodality representation and anatomical constraint for iso-intense infant brain MR image segmentation,” *NeuroImage*, vol. 89, no. 1, pp. 152–164, 2014.
- [26] O. M. Benkarim, P. Radeva, and L. Igual, “Label consistent multiclass discriminative dictionary learning for MRI segmentation,” in *Articulated Motion Deformable Objects*. New York, NY, USA: Springer-Verlag, 2014, pp. 138–147.
- [27] D. L. Collins, C. Holmes, T. Peters, and A. Evans, “Automatic 3-D model-based neuroanatomical segmentation,” *Human Brain Mapping*, vol. 3, no. 3, pp. 190–208, 1995.
- [28] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, “Automatic anatomical brain MRI segmentation combining label propagation and decision fusion,” *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.
- [29] D. L. Collins and J. C. Pruessner, “Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion,” *NeuroImage*, vol. 52, no. 4, pp. 1355–1366, 2010.
- [30] S. K. Warfield, K. H. Zou, and W. M. Wells, “Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation,” *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, Jul. 2004.
- [31] M. Sabuncu, B. Yeo, K. V. Leemput, B. Fischl, and P. Golland, “A generative model for image segmentation based on label fusion,” *IEEE Trans. Med. Imag.*, vol. 29, no. 10, pp. 1714–1729, Oct. 2010.
- [32] H. Wang, J. W. Suh, J. Pluta, M. Altinay, and P. Yushkevich, “Optimal weights for multi-atlas label fusion,” in *Proc. 22nd Int. Conf. Inf. Proc. Med. Imaging*, 2011, pp. 73–84.
- [33] S. Roy, A. Carass, J. L. Prince, and D. L. Pham, “Subject specific sparse dictionary learning for atlas-based brain MRI segmentation,” in *Proc. MICCAI Workshop Mach. Learn. Med. Imag.*, 2014, vol. 8679, pp. 248–255.
- [34] M. Jenkinson and S. Smith, “A global optimization method for robust affine registration of brain images,” *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, 2001.

- [35] D. L. Pham and J. L. Prince, "An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities," *Pattern Recog. Lett.*, vol. 20, no. 1, pp. 57–68, 1999.
- [36] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [37] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [38] N. Shiee, P.-L. Bazin, J. Cuzzocreo, A. Blitz, and D. Pham, "Segmentation of brain images using adaptive atlases with application to ventriculomegaly," in *Proc. 22nd Int. Conf. Inf. Process. Med. Imag.*, 2011, pp. 1–12.
- [39] C. A. Cocosco, V. Kollokian, R. K. S. Kwan, and A. C. Evans, "BrainWeb: Online interface to a 3-D MRI simulated brain database," *NeuroImage*, vol. 5, no. 4, p. S425, 1997.
- [40] E. M. Sweeney, R. T. Shinohara, N. Shiee, F. J. Mateen, A. A. Chudgar, J. L. Cuzzocreo, P. A. Calabresi, D. L. Pham, D. S. Reich, and C. M. Crainiceanu, "OASIS is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in MRI," *NeuroImage Clin.*, vol. 2, pp. 402–413, 2013.
- [41] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 153–163, Feb. 2012.
- [42] J. C. Bezdek, "A Convergence theorem for the fuzzy ISO-DATA clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980.
- [43] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: Theory and rationale for its development: The international consortium for brain mapping," *NeuroImage*, vol. 2, no. 2, pp. 89–101, 1995.
- [44] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.
- [45] A. Carass, J. Cuzzocreo, M. B. Wheeler, P.-L. Bazin, S. M. Resnick, and J. L. Prince, "Simple paradigm for extra-cerebral tissue removal: Algorithm and analysis," *NeuroImage*, vol. 56, no. 4, pp. 1982–1992, 2011.
- [46] L. R. Dice, "Measure of the amount of ecological association between species," *Ecology*, vol. 26, pp. 297–302, 1945.
- [47] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache, "Spatial decision forests for MS lesion segmentation in multichannel magnetic resonance images," *NeuroImage*, vol. 57, no. 2, pp. 378–390, 2011.
- [48] M. Styner, J. Lee, B. Chin, M. S. Chin, O. Commowick, H.-H. Tran, V. Jewells, and S. Warfield, "3D Segmentation in the clinic: A grand challenge II: MS lesion segmentation," presented at the The MIDAS Journal-MS Lesion Segmentation MICCAI 2008 Workshop, pp. 1–6, 2008. [Online]. Available: <https://www.nitrc.org/projects/msseg>
- [49] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.
- [50] L. D. Eggert, J. Sommer, A. Jansen, T. Kircher, and C. Konrad, "Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain," *PLoS One*, vol. 7, no. 9, p. e45081, 2012.
- [51] J. Souplet, C. Lebrun, N. Ayache, and G. Malandain, "An automatic segmentation of T2-FLAIR multiple sclerosis lesions," presented at the Midas Journal-MS Lesion Segmentation, MICCAI Workshop, New York, NY, USA, 2008.
- [52] J. Lecoecur, J. Ferré, and C. Barillot, "Optimized supervised segmentation of MS lesions from multispectral MRIs," presented at the workshop on medical image analysis on multiple sclerosis, London, U.K., 2009.
- [53] L. S. Ait-Ali, S. Prima, P. Hellier, B. Carsin, G. Edan, and C. Barillot, "STREM: A robust multidimensional parametric method to segment MS lesions in MRI," in *Proc. Med. Image Comput. Comput. Asst. Intervention*, 2005, pp. 409–416.
- [54] Z. Lao, D. Shen, D. Liu, A. F. Jawad, E. R. Melhem, L. J. Launer, R. N. Bryan, and C. Davatzikos, "Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine," *Acad. Radiol.*, vol. 15, no. 3, pp. 300–313, 2008.
- [55] A. P. Zijdenbos, R. Forghani, and A. C. Evans, "Automatic "pipeline" analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis," *IEEE Trans. Med. Imag.*, vol. 21, no. 10, pp. 1280–1291, Oct. 2002.
- [56] S. Roy, A. Carass, Q. He, A. Jog, J. L. Cuzzocreo, D. S. Reich, J. L. Prince, and D. L. Pham, "Example-based lesion segmentation," *Proc. SPIE Med. Imag.*, vol. 9034, p. 90341Y, 2014.
- [57] N. Weiss, D. Rueckert, and A. Rao, "Multiple sclerosis lesion segmentation using dictionary learning and sparse coding," in *Proc. Med. Image Comput. Comput. Asst. Intervention*, 2013, vol. 8149, pp. 735–742.

Authors' photographs and biographies not available at the time of publication.