

Multiple Sclerosis Lesion Segmentation Using Dictionary Learning and Sparse Coding

Nick Weiss¹, Daniel Rueckert², and Anil Rao²

¹ University of Lübeck, Lübeck, Germany

² Imperial College London, London, UK

Abstract. The segmentation of lesions in the brain during the development of Multiple Sclerosis is part of the diagnostic assessment for this disease and gives information on its current severity. This laborious process is still carried out in a manual or semiautomatic fashion by clinicians because published automatic approaches have not been universal enough to be widely employed in clinical practice. Thus Multiple Sclerosis lesion segmentation remains an open problem. In this paper we present a new unsupervised approach addressing this problem with dictionary learning and sparse coding methods. We show its general applicability to the problem of lesion segmentation by evaluating our approach on synthetic and clinical image data and comparing it to state-of-the-art methods. Furthermore the potential of using dictionary learning and sparse coding for such segmentation tasks is investigated and various possibilities for further experiments are discussed.

1 Introduction

Multiple Sclerosis (MS) is an autoimmune demyelinating disease occurring in the central nervous system (CNS). It is chronic, inflammatory and currently incurable. The underlying cause for the spontaneous degeneration of the myelin and subsequently the axons is still unknown and the lesions can appear at various locations within the brain causing a wide range of symptoms such as numbness, weakness, visual impairment or loss of balance. Approved medications and therapies present a symptomatic treatment to decrease the severity, occurrence and duration of certain symptoms [2].

Magnetic resonance imaging (MRI) significantly contributes to the evaluation of new therapies during clinical trials, as it is very sensitive to most of the lesions appearing in the white matter (WM) of the brain [7]. Lesions are visible as hyperintense areas in T2-weighted (T2w) and often hypointense in T1-weighted (T1w) MR images. Counting these white matter lesions (WML) and determining their total lesion load (TLL) are key criteria for quantifying the progression of MS and the current diagnosis criteria for MS (McDonald Criteria) [12]. In clinical practice the detection and segmentation of WML is still done in a manual or semiautomatic fashion by most clinicians. This is a time consuming task that suffers from a large intra- and interexpert variability [9]. Thus an automatic approach is highly desirable.

Over the last two decades several automatic methods have been proposed. The conclusion of two recent reviews [10,5] was that MS lesion and segmentation remains an open problem. No automatic method has been widely employed in clinical practice as they are too specific to deal with the heterogeneity in location and texture of lesions, and the differences in MRI acquisition. While supervised methods rely on images with previously segmented lesions, unsupervised methods do not require labeled data [5]. Unsupervised methods often try to model the intensity distribution of the healthy brain tissues, namely WM, grey matter (GM) and cerebrospinal fluid (CSF). Voxels that cannot be explained by this model are called outliers and labeled as lesions. Van Leemput et al. [19] developed a well-known method based on such a model.

We create an unsupervised approach that also segments the lesions as outliers with respect to healthy brain tissue. It introduces dictionary learning and sparse coding for the segmentation of MS lesions, which is new to our knowledge. Many areas of image processing have already benefited from this methodology [17,3]. The principal idea is to learn a dictionary primarily from healthy brain image tissue and then try to sparsely reconstruct image patches using the dictionary. Image patches containing lesions have a higher reconstruction error and thresholding this reconstruction error for every voxel then provides a lesion segmentation. In this paper we describe our new approach and evaluate it by using synthetic and clinical brain images with a ground truth segmentation. Finally we compare the results with other methods and outline the potential of this new approach.

2 Material and Methods

As stated in Fig. 1 our method can be divided into three different parts. The first part is preprocessing and includes brain extraction, image patch extraction and patch normalization. The second part focuses on learning a dictionary from these image patches and trying to reconstruct them using sparse coding. The last part shows how we provide a relative reconstruction error map and get the threshold for the final WML segmentation. In the remaining subsections we give information about the synthetic image data, clinical image data, the evaluation and the implementation.

2.1 Preprocessing

Since we want to segment lesions from the brain tissue, we require a brain mask that excludes the skull and all non-brain tissue. For the synthetic image data used in this paper a brain mask is provided along with the images. For the clinical image data, a brain mask is created using the brain extraction tool (BET) [13] and a bias correction is performed [18].

The m image patches lying within the brain mask ($m \approx 1.5 \cdot 10^6$ for a voxel size of 1 mm^3) are extracted and realigned to one-dimensional vectors $x_1, \dots, x_m \in \mathbb{R}^k$ with $k = 27$ for a three-dimensional image patch of size $3 \times 3 \times 3$. Each patch is

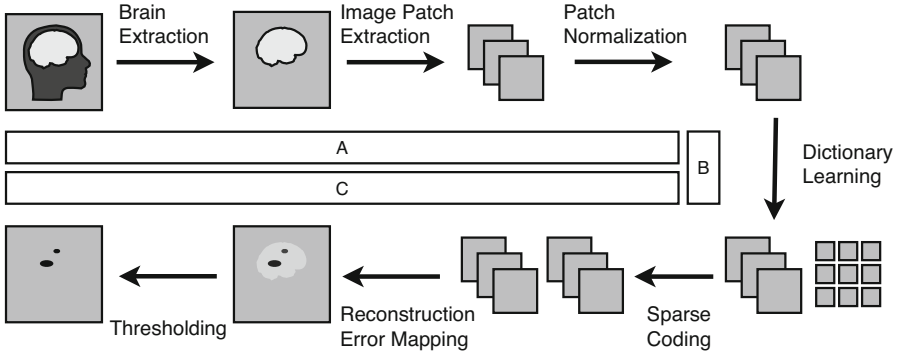


Fig. 1. A brief overview of the presented method. (A) Preprocessing including brain extraction, patch extraction and patch normalization. (B) Dictionary learning and sparse coding reconstruction of the image patches. (C) Mapping of the reconstruction error and finding a threshold for the final segmentation.

then divided by the value of the highest L_1 -norm of all patches and its square root is taken so that each patch fulfills $\|x_i\|_2^2 \leq 1$. To combine image patches of different modalities the vectors can easily be concatenated after the described patch normalization. A multiplication by the weight $\sqrt{1/z}$ where z is the number of available image sequences (e.g. T1/T2/FLAIR/etc.) ensures that the condition $\|x_i\|_2^2 \leq 1$ with $x_i \in \mathbb{R}^{z \cdot k}$ is still satisfied.

2.2 Dictionary Learning and Sparse Coding

A dictionary $D \in C = \{D \in \mathbb{R}^{k \times l} \text{ s.t. } \forall j : \|d_j\|_2^2 \leq 1\}$ with l atoms d_j is learned from these image patches so that it solves the optimization problem

$$\min_D \sum_{i=1}^m \frac{1}{2} \|x_i - D\alpha_i\|_2^2 \text{ s.t. } \|\alpha_i\|_1 \leq \lambda_1. \quad (1)$$

Dictionary learning searches for a basis D that satisfies $x_i \approx D\alpha_i$ for most image patches. Although the image patches with lesions are included in this learning process they do not impair the dictionary's ability to only represent the healthy brain image patches well. The reason for this is that just a small percentage of the patches include lesions.

The L_1 -norm constraint in (1) introduces sparsity to its solution so that only a few atoms of the dictionary should be used to represent an image patch [11]. It has been shown that learning a dictionary like this leads to a very precise image reconstruction [3]. If we try to reconstruct the image patches in a second optimization step

$$\min_{\alpha_i} \|x_i - D\alpha_i\|_2^2 \text{ s.t. } \|\alpha_i\|_1 \leq \lambda_2 \quad (2)$$

we will obtain a reconstruction error for each image patch depending on the sparsity constraint. An appropriately chosen parameter λ_2 shows that healthy

brain tissue can be easily reconstructed with a few atoms while the image patches containing lesions produce a higher error using the same amount of atoms.

2.3 Reconstruction Error and Thresholding

The relative reconstruction error of each image patch

$$err(x_i, \alpha_i) = \frac{\|x_i - D\alpha_i\|_2}{\|x_i\|_2} \quad (3)$$

can now be obtained and mapped at the position of the centered voxel within the patch. The result is an error map throughout the whole brain.

Finally a threshold is applied to obtain the segmentation of the WML from the combined error map. We do this by creating a smoothed histogram of the error map for a randomly chosen subject, and search for a point where the histogram derivative magnitude is close to zero that separates the lesion and non-lesion voxels for this subject. This derivative magnitude threshold is applied to the histograms of all subjects in the data set. Since we only use one subject during this process, this may still be considered a largely unsupervised procedure.

2.4 Synthetic Image Data

As suggested by Garcia-Lorenzo et al. [5] the freely available BrainWeb image data is a good first evaluation step to show a proof of concept and to test the method's robustness towards noise and intensity inhomogeneity [1]. Furthermore it allows us to compare our results to others as many authors evaluate their methods with this data.

It is possible to create T1w, T2w images with different levels of noise (3%, 5%, 7%, 9%) and intensity inhomogeneity (0%, 40%) in an online MRI simulator. The ground truth segmentation is available for all these WML as well as the brain segmentation. The images are simulated with an isotropic resolution of 1 mm^3 and have a size of $181 \times 217 \times 181$.

2.5 Clinical Image Data

The evaluation with synthetic image data is limited since we only have one phantom and images with simulated lesions are much easier to segment [5]. Furthermore we could not consider other MRI sequences like fluid attenuated inversion recovery (FLAIR) which is known to produce images predestinated for lesion segmentation.

We therefore further validate our framework using clinical data provided by the MS lesion segmentation challenge which was introduced as a workshop at MICCAI 2008 and still available [15]. In this paper, we focus on the 20 available training cases, which we downsample from the original resolution of 0.5 mm^3 isotropic to 1 mm^3 isotropic. The resulting images are then of size $256 \times 256 \times 256$

voxels. The data comes from the Children’s Hospital Boston (CHB) and the University of North Carolina (UNC).

For the synthetic data we used both T1w and T2w images, since using the T2w alone cannot discriminate between CSF and lesions. Inclusion of the T1w patches gives the CSF its own footprint, thereby reducing the number of false positives segmented as lesion. In the clinical data, we use just the FLAIR sequence since the CSF appears as hypointense and the lesions as hyperintense in these images, so T1w images are not required.

2.6 Evaluation

The most common validation measure in the context of segmentation is the Dice Similarity Coefficient $DSC = \frac{2 \times TP}{FP + FN + 2 \times TP}$ with the number of true positive (TP), false positive (FP) and false negative (FN) voxels. The DSC rewards a method for its ability to detect lesions and to reject healthy tissue. It is used to evaluate the synthetic image data as it is used by the other authors.

For the clinical image data, we also calculate the true positive rate $TPR = \frac{TP}{TP + FN}$, and the positive predictive value $PPV = \frac{TP}{TP + FP}$. This allows us to compare our results with those of Geremia et al. [8].

2.7 Implementation

Our approach was implemented using MATLAB and the SParse Modeling Software (SPAMS) [11]. SPAMS was used to solve the optimization problems (1) and (2). Different parameters have been tested for this method. A good result is provided by an image patch size of $3 \times 3 \times 3$ and a dictionary D with $l = 100$ atoms. The following sparsity constraints were also determined empirically: $\lambda_1 = 0.9$ and $\lambda_2 = 0.8$. The whole segmentation took approximate 5 minutes and was carried out using an Intel Core 2 Duo processor at 2.4 GHz with 4 GB of RAM.

3 Results and Discussion

The proposed method is evaluated first using the synthetic BrainWeb image data, using different levels of noise and intensity inhomogeneity. We compared the results with four other unsupervised methods whose authors provided the DSC for all the different cases and are considered as state-of-the-art methods for segmentation of MS lesions [19,6,4,16].

Our method gives DSC measures that are highly competitive with those of the other approaches. The mean DSC is 71 for 0 % intensity non-uniformity (INU) and 63 for 40 % INU. Though at the lower noise levels (3 %, 5 %) and independent of the intensity non-uniformity (0 %, 40 %) the results are lower with up to 16 percentage points (pps) compared to the best methods. The intensity non-uniformity of 40 % slightly decreased the DSC for most of the methods. It is noticeable that our approach is robust to the presence of noise across all experiments, despite there being no explicit noise-reducing step. The dictionary

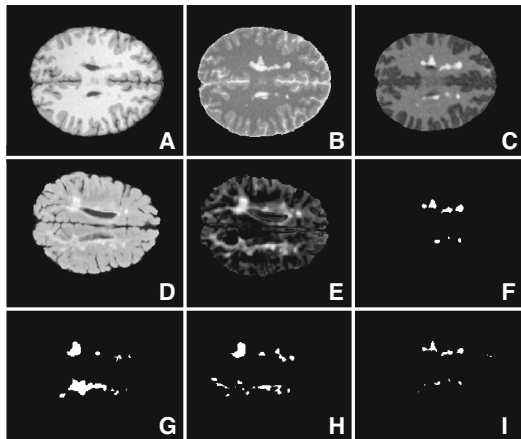


Fig. 2. Synthetic T1w (A) and T2w (B) images with lesions and their error map (C) after reconstruction. The final lesion segmentation (F) of the synthetic data compared with the ground truth (I). Clinical FLAIR image (D), its error map (E) after reconstruction, final lesion segmentation (H) and ground truth (G).

obviously adapts itself as it is learned with noisy image patches. This is a clear advantage of our method compared to the others.

As expected the clinical data is much more challenging (see figure 2 and Table 1). The TLL varies heavily from 105 mm^3 (UNC05) to 22542 mm^3 (CHB1) across all subjects while the TLL for the synthetic image data is constant with 3512 mm^3 . The created brain masks also have a big impact on the results although we try to provide a good individual brain mask for each subject by varying the parameters of BET and using morphological operators. Different threshold values are provided for the synthetic data, the UNC and the CHB data as the raters show a high inter-expert variability and seem to have slightly different definitions of lesions. Across the whole clinical data, the proposed method achieved a mean (standard deviation) TPR of 33% (19) and a PPV of 37% (20). This is higher than the TPR of 21% (14) and PPV of 30% (17) provided by Souplet [14], another unsupervised approach and the MS lesion segmentation challenge winner in 2008. The method by Geremia [8] reaches a TPR of 40% (18) and a PPV of 40% (20) which is slightly higher than our results but within the reach. Other than ours their method is supervised, which therefore requires training data. This data can be expensive to obtain. However we can further extend our method to include some supervision which may improve our results. Importantly, the TPR and PPV values for the subject with the largest lesion load, CHB01, are competitive with those of the rival methods. This demonstrates that, even in the presence of larger lesion loads, the sparsity constraints in the dictionary learning and reconstruction phases enable the identification of lesions as outliers with respect to healthy tissue. Indeed, even extremely high lesion loads $\approx 60000 \text{ mm}^3$ only represent less than 5% of the total brain volume, so their detection should still be possible.

Table 1. TPR/PPV/DSC results (%) on clinical brain image data. Our method is compared to two state-of-the-art rival methods (2008: Souplet [14], 2010: Geremia [8]). Note that DSC measures were not provided by the authors for those methods.

	Souplet		Geremia		Our Method				Souplet		Geremia		Our Method		
Patient	TPR	PPV	TPR	PPV	TPR	PPV	Dice	Patient	TPR	PPV	TPR	PPV	TPR	PPV	Dice
UNC01	1	1	2	1	33	29	31	CHB01	22	41	49	64	60	58	59
UNC02	37	39	48	36	54	51	53	CHB02	18	29	44	63	27	45	34
UNC03	12	16	24	35	64	27	38	CHB03	17	21	22	57	24	56	34
UNC04	38	54	54	38	40	51	45	CHB04	12	55	31	78	27	66	38
UNC05	38	8	56	19	25	10	16	CHB05	22	42	40	52	29	33	31
UNC06	57	9	15	8	13	55	20	CHB06	13	46	32	52	10	36	16
UNC07	27	18	76	16	44	23	30	CHB07	13	39	40	54	14	48	22
UNC08	27	20	52	32	43	13	20	CHB08	13	55	46	65	21	73	32
UNC09	16	43	67	36	69	6	11	CHB09	3	18	23	28	5	22	8
UNC10	22	28	53	34	43	23	30	CHB10	5	18	23	39	15	12	13

4 Conclusions

We have evaluated our method using dictionary learning for MS lesion segmentation with synthetic and clinical image data. The results were competitive with state-of-the-art methods and displayed the robustness of our method towards noise. There are many potential ways of extending the presented framework. One idea is to use the combined intensities from different MRI sequences to learn separate dictionaries for WM, GM and CSF. An outlier or lesion is then identified if it cannot be well reconstructed from any of the dictionaries. Alternatively, we could use the reconstruction coefficients α across the dictionaries to identify lesions. We could also introduce supervision into our method by learning a dictionary specific to lesions in addition to dictionaries for WM, GM and CSF. Patch size is another important factor to consider: While we have empirically found that a 3x3x3 patch works best, it would be interesting to incorporate larger patches into the framework, which may potentially capture more texture information and further improve results.

References

1. Cocosco, C., Kollokian, V., Kwan, K., Pike, G.B.: BrainWeb: Online Interface to a 3D MRI Simulated Brain Database - Abstract - Europe PubMed Central. Neuroimage 5, 425 (1997)
2. Compston, A., Coles, A.: Multiple sclerosis. The Lancet 372(9648), 1502–1517 (2008)
3. Elad, M.: Sparse and Redundant Representations. From Theory to Applications in Signal and Image Processing. Springer (2010)
4. Forbes, F., Doyle, S., Garcia-Lorenzo, D., Barillot, C., Dojat, M.: A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation. In: 13th International Conference on Artificial Intelligence and Statistics, AISTATS 2010, vol. 9, pp. 225–232 (2010)
5. Garcia-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L., Collins, D.L.: Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Medical Image Analysis 17(1), 1–18 (2013)

6. García-Lorenzo, D., Lecoœur, J., Arnold, D.L., Collins, D.L., Barillot, C.: Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 584–591. Springer, Heidelberg (2009)
7. Ge, Y.: Multiple sclerosis: the role of MR imaging. *AJNR. American Journal of Neuroradiology* 27(6), 1165–1176 (2006)
8. Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel MR images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 111–118. Springer, Heidelberg (2010)
9. Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, G.J., Plummer, D.L., Tofts, P.S., McDonald, W.I., Miller, D.H.: Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. *Magnetic Resonance Imaging* 14(5), 495–505 (1996)
10. Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J.C., Quiles, A., Valls, L., Ramió-Torrentà, L., Rovira, À.: Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Information Sciences* 186(1), 164–185 (2012)
11. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 689–696 (2009)
12. Polman, C.H., et al.: Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology* 69(2), 292–302 (2011)
13. Smith, S.M.: Fast robust automated brain extraction. *Human Brain Mapping* 17(3), 143–155 (2002)
14. Souplet, J.C., Lebrun, C., Ayache, N., Malandain, G.: An automatic segmentation of T2-FLAIR multiple sclerosis lesions. In: Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge, pp. 1–11 (2008)
15. Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., Warfield, S.: 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. *MIDAS Journal*, 1–5 (2008)
16. Tomas-Fernandez, X., Warfield, S.K.: Population intensity outliers or a new model for brain WM abnormalities. In: Proceedings of 9th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1543–1546 (2012)
17. Tošić, I., Frossard, P.: Dictionary learning. *IEEE Signal Processing Magazine* 28(2), 27–38 (2011)
18. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging* 29(6), 1310–1320 (2010)
19. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging* 20(8), 677–688 (2001)