MSSEG Challenge Proceedings: Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure

Olivier Commowick, Frédéric Cervenansky, Roxana Ameli

MSSEG Challenge Proceedings: Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure

Editors

Olivier Commowick, Frédéric Cervenansky, Roxana Ameli

http://portal.fli-iam.irisa.fr/msseg-challenge

Preface

This proceedings book gathers methodological papers of segmentation methods evaluated at the first MICCAI Challenge on Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure. This challenge took place as part of an effort of the OFSEP[1] (French registry on multiple sclerosis aiming at gathering, for research purposes, imaging data, clinical data and biological samples from the French population of multiple sclerosis subjects) and FLI[2] (France Life Imaging, devoted to setup a national distributed e-infrastructure to manage and process medical imaging data). These joint efforts are directed towards automatic segmentation of MRI scans of MS patients to help clinicians in their daily practice. This challenge took place at the MICCAI 2016 conference, on October 21st 2016.

More precisely, the goals of this challenge were multiple. It first aimed at evaluating state-of-the-art and advanced segmentation methods from the participants on a database following a standard protocol[3]. For this, both lesion detection (how many lesions are detected) and lesion segmentation (how precise the lesions are delineated) were evaluated on a multi-centric database (38 patients from four different centers, imaged on 1.5 or 3T scanners, each patient being manually annotated by seven experts from three different French centers, located in Bordeaux, Lyon and Rennes).

This challenge was also the occasion to perform this advanced evaluation on a common infrastructure, provided by FLI. As such, challengers were asked to provide their pipeline as a Docker container image. After integration in the VIP platform[4], the challengers pipelines were then evaluated independently by the challenge organization team, the testing data and evaluation results being queried and stored in a Shanoir database[5]. This infrastructure enabled a fair comparison of the algorithms in terms of running time comparison and ensuring all algorithms were run with the same parameters for each patient (which is required for a truly automatic segmentation). These proceedings do not include results of the evaluation, rather the evaluated methods descriptions. Evaluation results are available on the challenge website[6] from the day of the challenge.

As a conclusion note, the organizers of the challenge are welcoming new pipelines to be evaluated after the challenge itself. Interested teams may go on the challenge website to register their new method and evaluate it on our data.

Acknowledgments

This challenge workshop was partially supported by the France Life Imaging national project in France (ANR-11-INBS-006), and by the OFSEP national

[1] OFSEP: http://www.ofsep.org
[2] FLI: https://project.inria.fr/fli/en/
[3] OFSEP acquisition protocol: http://www.ncbi.nlm.nih.gov/pubmed/25660217
[4] VIP: http://vip.creatis.insa-lyon.fr
[5] Shanoir: http://shanoir.org
[6] MSSEG challenge: http://portal.fli-iam.irisa.fr/msseg-challenge

project in France (ANR-10-COHO-002). This challenge would also not have been possible without the great work of FLI, OFSEP and VIP engineers, namely: Sorina Camarasu-Pop, Pascal Girard, Florent Leray, Julien Louis, Simon Loury, Aneta Morawin, Mathieu Simon and Yao Yao.

Organization

The challenge was jointly organized by France Life Imaging (for methodological and technical resources) and the OFSEP French national cohort in MS (for data provision and medical expertise). The organization was based on three specific boards: a scientific, technical and clinical committees.

Scientific committee

Christian Barillot	CNRS, Rennes, France
Olivier Commowick	Inria, Rennes, France
Charles Guttman	Brigham and Women's Hospital, Boston, MA, USA
Martin Styner	University of North Carolina, Chapel Hill, NC, USA
Simon Warfield	Children's Hospital, Boston, MA, USA

Technical committee

Frédéric Cervenansky	Université Claude Bernard, Lyon, France
Tristan Glatard	CNRS, Lyon, France
Audrey Istace	OFSEP, Lyon, France
Michael Kain	FLI-IAM, Inria, Rennes, France
Baptiste Laurent	FLI-IAM, Inserm, Brest, France

Medical committee

Roxana Ameli	University Hospital of Lyon, France
François Cotton	University Hospital of Lyon, France
Jean-Christophe Ferré	University Hospital of Rennes, France
Thomas Tourdias	University Hospital of Bordeaux, France

Table of Contents

Multiple Sclerosis lesion segmentation using an automated multimodal Graph Cut

Jeremy Beaumont Olivier Commowick Christian Barillot

VisAGeS U746 INSERM / INRIA, IRISA UMR CNRS 6074, Rennes, France

Abstract. In this paper, we present an algorithm for Multiple Sclerosis (MS) lesion segmentation. Our method is fully automated and includes three main steps: 1. the computation of a rough total lesion load in order to optimize the parameter set of the following step; 2. the detection of lesions by graph cut initialized with a robust Expectation-Maximization (EM) algorithm; 3. the application of rules to remove false positives and to adjust the contour of the detected lesions. Our algorithm will be tested on the FLI 2016 MSSEG challenge data.

Keywords: Graph Cut, Expectation-Maximization, Multiple sclerosis, Tissue classification

1 Introduction

Multiple Sclerosis (MS) is a chronic demyelinating disease that affects the central nervous system. Brain lesions detection plays an important role in Multiple Sclerosis (MS) studies, as it is used to evaluate patient disease and its future evolution. Currently, lesions are detected by manual or semi-automatic segmentation methods, which are very time consuming and which show a high inter and intra-raters variability [8]. This issue can be solved with fully automated MS lesion segmentation methods. Here, we present one, based on the combination of graph cut and robust EM tissues segmentation using multiple sequences of Magnetic Resonance Imaging (MRI). Our process is applied to the FLI 2016 MSSEG challenge data.

2 Challenge data and evaluation criteria

2.1 Data and pre-processing

15 MS patients data sets are available to allow challengers to optimize their segmentation algorithms. These data sets contain pre-processed and unprocessed data, available for challengers who would rather do their own pre-processing on the data, the ground truth and the seven manual segmentations used to compute it.

The challenge data sets include T1-w, T1-w Gadolinium, T2-w, PD and FLAIR sequences. It will not be described further in this paper, more details

can be found on the challenge website[1]. The pre-processed data are denoised with the NL-means algorithm [5], rigidly registered [3] towards the FLAIR images, brain extracted using the volBrain platform [10] and bias corrected using the N4 algorithm [11]. As data are brain extracted, brain masks are provided in the pre-processed data sets. We decided to use the pre-processed data. Therefore, the method we described below will focus only on the MS lesions segmentation itself.

2.2 Evaluation criteria for the MS-SEG challenge

The qualitative evaluation of the proposed segmentation algorithm is made through two categories of evaluation metrics: lesion detection (are the lesions well detected independently of the contour quality?) and segmentation precision (are the lesion contours close to those of the ground truth?). For the MS-SEG challenge, the different segmentation workflow results will be compared to a ground truth for each MS patient using several evaluation metrics and will be ranked using two of them[2]:

F1 score: A metric used to assess the capacity of an algorithm to detect lesions. The F1 score is a combination of the lesion sensitivity (SensL), i.e. the proportion of detected lesions in the ground truth, and the lesion positive predictive value (PPVL), i.e. the proportion of true positive lesions inside the result of segmentation algorithms.

Dice score: A well known overlap metric used to assess the capacity of an algorithm to be accurate in lesion delineation.

The ground truth is computed with the Logarithmic Opinion Pool Based STAPLE (LOP STAPLE) method [1], using seven independent manual segmentations for each patient.

3 MS lesions segmentation workflow

3.1 Lesions detection using graph cut

The segmentation algorithm relies on a graph cut approach previously presented in [6] and [2]. 3 MR sequences and the brain mask are required for this algorithm. We choose to use T1-w, T2-w and FLAIR sequences. We do not use PD as it generally shows less MS lesion contrast than T2-w and FLAIR.

Graph cut principle: MS patient images are used to generate a graph which will be exploited to segment in an optimal way MS lesions from both contour and regional information. This graph is initialized in a manner that each of its

[1] `https://portal.fli-iam.irisa.fr/msseg-challenge/data`

[2] More details are provided on the challenge website: `https://portal.fli-iam.irisa.fr/msseg-challenge/evaluation`

nodes corresponds to a voxel and is connected to two others nodes representing the object class for MS lesions and the background class for normal appearing brain tissues (NABT). These two nodes are respectively called terminal source and sink. The image nodes are connected with their spatial neighbors by n-links, whose values are computed using a spectral gradient [6] and depend on the similarity of the two considered voxels. The t-links connect nodes in the image to their corresponding terminal source and sink nodes and represent how voxels fit into given models of the object and background. The simplest way to estimate object and background models is to use seeds chosen by a user. However, user interactions are prohibited if we want to develop an automated algorithm. This is why we compute the seeds in images with a 3-class multivariate Gaussian Mixture Model (GMM), where each class is equivalent to a brain tissue: White Matter (WM), Grey Matter (GM) and Cerebrospinal Fluid (CSF). MS lesions are considered as the outliers of this model.

Seeds computation: To be robust to outliers, the 3-class multivariate GMM is estimated using an Expectation Maximisation (EM) algorithm [7] which optimizes a trimmed likelihood. This EM algorithm has a parameter, h, representing the portion of voxels that are removed from the estimation. Its value needs to be adjusted to reject MS lesions as well as other outliers like veins or skull stripping errors from the estimation of the NABT model.

The obtained parameters of the GMM are used to compute a Mahalanobis distance [4] between each voxel of the images and each class of the NABT model. From this distance, a p-value, used to represent the probability not to fit into each of the 3 classes, can be computed. For each voxel i, we keep the lowest p-value among the three classes, denoted p_i. Sinks should have a high value when their corresponding voxels are close to the NABT model. The sinks t-links weights W_{bi} are then computed as:

$$W_{bi} = 1 - p_i \tag{1}$$

All voxels that do not fit in the NABT model have a high p-value, therefore, we wish to differentiate MS lesions from other outliers (vessels, skull stripping errors ...) using a priori knowledge about lesion intensities. MS lesions are usually hyperintense compared to WM in T2-w and FLAIR images. A fuzzy logic approach has been chosen to model this expert's knowledge. Instead of defining a binary threshold for hyper intensity, a fuzzy weight, computed for each sequence from the two parameters *slope begining* S_b and *slope end* S_e, is characterized (see [6] for more details). The final sources weights W_{oi} are computed by taking the minimum value between the p-value and the fuzzy weights W_{T2} and W_{flair}:

$$W_{oi} = \min(p_i, W_{T2}, W_{flair}) \tag{2}$$

Parameters definition: The presented algorithm works with three parameters: h, S_b and S_e. In order to obtain the best segmentation results, we optimize

these parameters with the provided training data set. We note that the parameter h depends on the proportion of outliers in an image, and as such is directly linked with the Total Lesion Load (TLL) of MS patients. Therefore, the algorithm parameters have to be adapted to the MS patient TLL, which has to be estimated before performing the segmentation. We define two parameter sets: one for mild lesion load (TLL $< 25\ cm^3$), and the other one for severe lesion load (TLL $\geq 25\ cm^3$). These sets are presented in Table 1. A rough TLL estimation is automatically computed with the following steps:

1. Non-linear registration of an atlas on the T1-w image. This atlas contains CSF, GM and WM probability maps plus a brain mask without the cerebellum and the brainstem.
2. Masking the T2-w and FLAIR images to keep only the WM in the two hemispheres (the amount of lesions is usually lower in the cerebellum and the brainstem and can be removed of the rough TLL estimation), using the atlas WM probability map and brain mask without the cerebellum and the brainstem.
3. Segmentation of the T2-w and FLAIR masked images with the K-means algorithm [9]. T2-w and FLAIR images are segmented respectively in 4 and 3 classes. This segmentation is performed to extract MS lesions, regrouped in one class in each image, from WM.
4. Intersection of the T2-w and FLAIR MS lesions classes.
5. Computation of the volume from the resulting image, which corresponds to an approximation of the TLL.

	h	S_b	S_e
Mild Lesion Load	0.1	2.0	4.0
Severe Lesion Load	0.4	3.0	4.0

Table 1: Values of our segmentation algorithm parameters, optimized on the training data.

3.2 Post-processing

After the detection of candidate lesions, some false positives still remain. To remove these artifacts, we add a post-process, made of the following steps:

1. lesions which have a size lower than 3 mm^3 are removed
2. lesions touching the brain mask border are removed, as they are probably false positives due to vessels or skull stripping errors
3. lesions not sufficiently located in WM are removed, as MS lesions are typically located there

4. lesions which do not touch a mask computed from MS patient T2-w and FLAIR sequences are removed. Lesions are considered as hyper intense in these two modalities, so it is possible to build a mask of "probable lesions", i.e. regions where lesions may appear and out of which no lesion may be seen. This mask is built by automatically thresholding the T2-w and FLAIR images and intersecting those masks. Our segmentation method generates several false positives in the brainstem, therefore, the mask used in this post-processing step also excludes this region.
5. lesions delineations are improved using the mask of "probable lesions" computed in the previous step

4 Results

4.1 Sample results

Table 2 presents an evaluation of the whole segmentation algorithm with a post-processing step on the training data. An example of segmentation result is shown in Figure 1.

	Dice scores	SensL	PPVL	F1 scores
Mild Lesion Load	0.4703	0.4124	0.5698	0.4441
Severe Lesion Load	0.7219	0.2775	0.4605	0.3061
Mean	0.5709	0.3584	0.5260	0.3889

Table 2: Evaluation of our segmentation algorithm on the training data

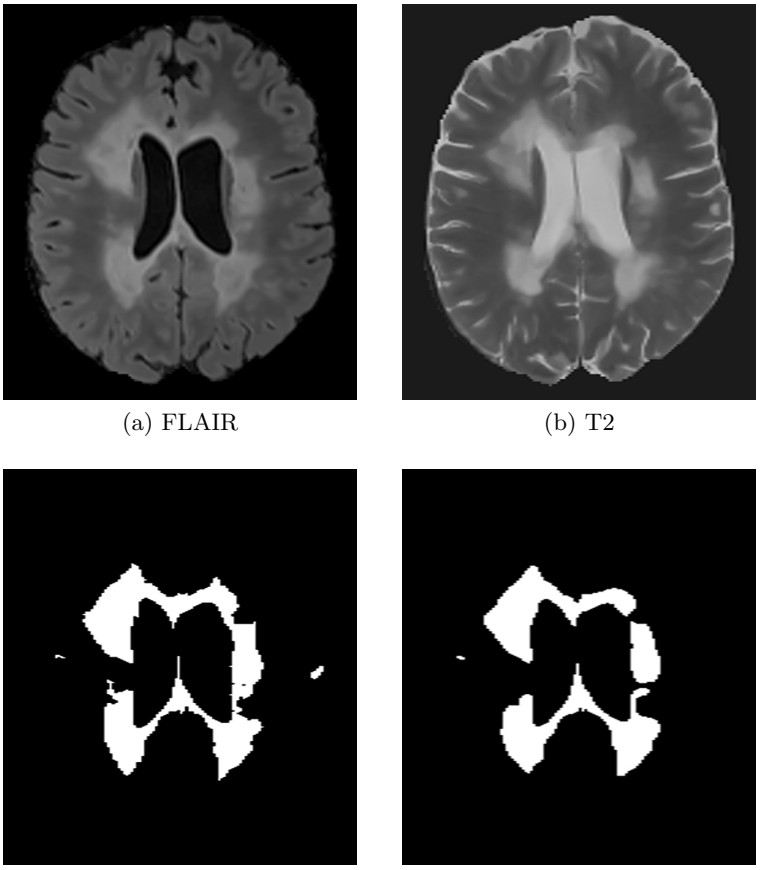
4.2 Implementation and Computation Times

The algorithm benefices of a multi-threaded implementation, based on ITK and available in open-source software Anima[3]. The total computation time to process each segmentation of the data set on a computer with an Intel(R) Xeon(R) CPU E5-2660 v3 @ 2.60GHz (8 cores) is approximately 10 minutes.

5 Conclusion

A fully automated MS lesion segmentation method using a graph cut initialized with a robust EM algorithm was presented. The results of the segmentation depend on the algorithm input parameters, which are directly linked with the MS patient TLL. The TLL is difficult to estimate and an error could result in a bad choice of these parameters, which may influence the segmentation workflow leading to worse results. Consequently, the automation of the presented method

[3] `https://github.com/Inria-Visages/Anima-Public`



(a) FLAIR (b) T2

(c) Ground truth (d) Automatic segmentation

Fig. 1: Automatic segmentation of the data set 01016SACH

is a complicated task where the initialization of the parameters is very important to reach satisfactory results. These results are improved with a post-processing step in order to reduce the number of false positives to be as close as possible to the ground truth.

References

1. Akhondi-Asl, A., Hoyte, L., Lockhart, M.E., Warfield, S.K.: A logarithmic opinion pool based STAPLE algorithm for the fusion of segmentations with associated reliability weights. IEEE transactions on medical imaging 33(10), 1997–2009 (Oct 2014)

2. Catanese, L., Commowick, O., Barillot, C.: Automatic graph cut segmentation of multiple sclerosis lesions. In: ISBI challenge on longitudinal Multiple Sclerosis lesion segmentation (Apr 2015)
3. Commowick, O., Wiest-Daesslé, N., Prima, S.: Block-matching strategies for rigid registration of multimodal medical images. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). pp. 700–703 (May 2012)
4. Commowick, O., Fillard, P., Clatz, O., Warfield, S.K.: Detection of DTI White Matter Abnormalities in Multiple Sclerosis Patients. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008, pp. 975–982. No. 5241 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Sep 2008), dOI: 10.1007/978-3-540-85988-8_116
5. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. IEEE Transactions on Medical Imaging 27(4), 425–441 (Apr 2008)
6. Garcia-Lorenzo, D., Lecoeur, J., Douglas, A., Louis Collins, D.: Multiple Sclerosis Lesion Segmentation Using an Automatic Multimodal Graph Cuts. In: MICCAI. LNCS, vol. 5762, pp. 584–591 (2009)
7. Garcia-Lorenzo, D., Prima, S., Arnold, D.L.: Trimmed-Likelihood Estimation for Focal Lesions and Tissue Segmentation in Multisequence MRI for Multiple Sclerosis. In: IEEE Trans. Med. Imag. vol. 30, no. 8, pp. 1455–1467. IEEE (2011)
8. 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 (Jan 1996)
9. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28(1), 100–108 (1979)
10. Manjon, J.V., Coupé, P.: volBrain: An online MRI brain volumetry system. In: Organization for Human Brain Mapping'15. Honolulu, United States (Jun 2015)
11. 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 (Jun 2010)

Automatic Multiple Sclerosis lesion segmentation from Intensity-Normalized multi-channel MRI

Jeremy Beaumont Olivier Commowick Christian Barillot

VisAGeS U746 INSERM / INRIA, IRISA UMR CNRS 6074, Rennes, France

Abstract. In the context of the FLI MICCAI 2016 MSSEG challenge for lesion segmentation, we present a fully automated algorithm for Multiple Sclerosis (MS) lesion segmentation. Our method is composed of three main steps. First, the MS patient images are registered and intensity normalized. Then, the lesion segmentation is done using a voxel-wise comparison of multi-channel Magnetic Resonance Images (MRI) against a set of controls. Finally, the segmentation is refined by applying several lesion appearance rules.

Keywords: Multiple Sclerosis, Intensity Normalization, Statistics, MRI

1 Introduction

Multiple Sclerosis (MS) is a auto-immune brain degenerative disease causing irreversible patient handicap and which is still not well understood. Lesion detection is a major step to evaluate the patient disease status and its future evolution. Manual and semi-automatic segmentation methods are very time consuming and can show high inter and intra-rater variability [5]. To solve this issue, we present a fully automated method for MS lesions segmentation based on the combination of intensity standardization and voxel-wise comparison of multi-channel Magnetic Resonance Images (MRI) of the patient and control subjects. Our process is applied to the FLI 2016 MSSEG challenge data.

2 Challenge data and evaluation criteria

2.1 Data and pre-processing

To allow challengers to optimize their segmentation algorithms, MS-SEG challenge organizers gave access to 15 MS patient data sets. Each data set contains pre-processed and unprocessed data, available for challengers who wish to perform their own pre-processing on the data, along with the ground truth and the seven manual segmentations used to compute it.

The challenge data sets include T1-w, T1-w Gadolinium, T2-w, PD and FLAIR sequences[1]. Pre-processed data are provided in order to reduce the dependency of the segmentation results on pre-processing performance. The pre-processed data are denoised with the NL-means algorithm [4], rigidly registered [2] towards the FLAIR images, brain extracted using the volBrain platform [8] and bias corrected using the N4 algorithm [12]. As brain extraction was performed, brain masks are provided in the pre-processed data sets. We decided to use the pre-processed data. For this reason, the method we describe below will focus only on the MS lesions segmentation itself.

2.2 Evaluation criteria for the MS-SEG challenge

The quality of the proposed segmentation algorithm may be assessed through two categories of evaluation metrics: lesion detection (are the lesions well detected independently of the contour quality?) and segmentation precision (are the lesion contours close to those of the ground truth?). For the MS-SEG challenge, the organizers will compare the results of the different segmentation workflows to a ground truth for each MS patient and will use several evaluation metrics, out of which two will be used for the ranking of the challengers algorithms[2]:

F1 score: This metric is used to evaluate the quality of an algorithm in terms of lesions detection. It corresponds to the combination of the lesion sensitivity (SensL), i.e. the proportion of detected lesions in the ground truth, and the lesion positive predictive value (PPVL), i.e. the proportion of true positive lesions inside the result of segmentation algorithms.

Dice score: This well known overlap metric is used to evaluate the quality of an algorithm in terms of segmentation precision.

The ground truth is computed with the Logarithmic Opinion Pool based STAPLE (LOP STAPLE) method [1], using seven independent manual segmentations for each patient.

3 MS lesions segmentation workflow

3.1 Intensity Standardization

Our segmentation workflow is based on the voxel-wise comparison of MS patient images against a set of controls. However, intensity profile of conventional MRI has a high inter-subject and inter-scanner variability. To solve this issue, Karpate et al. [7] proposed the estimation of a correction factor which is used to make corresponding anatomical tissues take on the same intensity profile.

[1] The challenge data sets will not be described further in this paper, more details can be found on the challenge website: `https://portal.fli-iam.irisa.fr/msseg-challenge/data`

[2] More details are provided on the challenge website: `https://portal.fli-iam.irisa.fr/msseg-challenge/evaluation`

Image intensities of a healthy brain can be modeled by a 3-class Gaussian Mixture Model (GMM), where each Gaussian represents one of the brain tissues: White Matter (WM), Gray Matter (GM) and Cerebrospinal Fluid (CSF). MS lesions are considered as outliers of this model. Estimating the three classes parameters is rendered difficult because of MS lesion outlier intensities. We therefore estimate them with a modification of the Maximum Likelihood Estimator (MLE) proposed by Notsu et al. [9], more robust to outliers. This estimation is based on the γ-loss function for the Normal distribution, used to maximize the MLE in the form of γ divergence, and is casted to yield an Expectation Maximisation (EM) algorithm [7].

Once the parameters are estimated, we obtain the means and covariances of tissues for the source and target images. These values are used to define a linear correction function which can be solved by linear regression. The results of the linear regression are then exploited to normalize the intensity profiles of the images.

3.2 MS lesions detection

MRI registration and intensity normalization make lesion segmentation possible through a comparison of vector of intensities between the patient and control subjects. Patient images are registered on the set of controls using a linear registration method, based on the use of a block-matching algorithm as presented in [2,10], and a non linear registration method, based on the estimation of a dense non linear transformation between the images as presented in [11]. The methodology used by Karpate et al. [6] to compare the multi-channel vectors of intensities between MS patient and a group of controls is based on the computation of statistical differences through the Mahalanobis distance [3]. These vectors of intensities are built from the available images and can therefore use any combination of them, like FLAIR, or T2-w and FLAIR, or DP, T2-w and FLAIR, making this a parameter of the algorithm.

3.3 Refinement of the segmentation

The intensities of pixels corresponding to brain tissues can vary in function of the brain region where they are located. Indeed, pixels which belong to the white matter brainstem and cerebellum are usually more intense than pixels which belong to the white matter hemispheres. This phenomenon induces the detection of several false positives in the brainstem and the cerebellum. Therefore, intensity standardization and MS lesions detection are computed on one hand in the two hemispheres and on the other hand in the brainstem and the cerebellum.

3.4 Post-processing

This comparison based segmentation algorithm may generate false positives for several reasons (registration errors, presence of noise in the images ...). Therefore, we add a post-processing step to our segmentation workflow in order to reduce the number of false positives. The post-processing is made of four steps:

1. lesions which have a size lower than 3 mm^3 are removed
2. lesions touching the brain mask border are removed, as they are probably false positives due to vessels or skull stripping errors
3. lesions not sufficiently located in WM are removed, as MS lesions are typically located there
4. lesions which do not touch a mask computed from MS patient T2-w and FLAIR sequences are removed. Lesions are considered as hyper intense in these two modalities, so it is possible to build a mask of "probable lesions", i.e. regions where lesions may appear and out of which no lesion may be seen. This mask is built by automatically thresholding the T2-w and FLAIR images and intersecting those masks. Our segmentation method generates several false positives in the brainstem, therefore, the mask used in this post-processing step also excludes this region.
5. lesions delineations are improved using the mask of "probable lesions" computed in the previous step

4 Results

4.1 MRI sequences used for MS lesions detection

Our algorithm can work with only one MR sequence or with several modalities. We have tested our segmentation workflow with T2-w and FLAIR sequences, discarding T1-w and PD images, as they generally show less MS lesion contrast than T2-w and FLAIR. Table 1 presents an evaluation of our segmentation (without post-processing algorithm) on the 15 training data. We worked on two possibilities to combine T2-w and FLAIR images for MS lesions segmentation:

Intersection: Here, we consider that the T2-w and FLAIR MS patient images are registered and intensity normalized on a set of controls. Vector images are created to combine T2-w and FLAIR sequences, on one hand, from MS patient images, and on the other hand, from control subject images. These vector images are then compared with the method described in section 3.2. Intersection of T2-w and FLAIR is interesting as it generates theoretically less false lesions detection as T2-w or FLAIR segmentation alone. However, this way to combine images has a drawback: in some cases, it is possible that a lesion intensity profile is close to the one of a tissue in a modality, which may reduce the Mahalanobis distance and induce the non-detection of a lesion even if its intensity profile is far from tissues intensity profiles in the other modalities.

Union: Here, we consider that our algorithm has already been used to segment MS patient lesions, first with T2-w sequences, and secondly with FLAIR sequences. Then, both segmentations are added. This induces, in theory, that more lesions are detected than in T2-w and in FLAIR segmentations alone. Yet, this type of combination has an inconvenient: false lesions detections from the both T2-w and FLAIR segmentations are kept, thereby adding some noise in the final segmentation.

	Dice scores	SensL	PPVL	F1 scores
T2-w	0.2610	0.3564	0.0305	0.0540
FLAIR	0.4231	0.5519	0.1021	0.1650
T2-w $\bigcup$ FLAIR	0.2978	0.5554	0.0373	0.0685
T2-w $\bigcap$ FLAIR	0.2476	0.5122	0.0205	0.0390

Table 1: Evaluation of our segmentation algorithm on the training data without post-processing.

4.2 Sample results

Table 2 presents an evaluation of the whole segmentation algorithm with a post-processing step on the training data. We choose to perform segmentation twice, using different MRI sequences each time, in order to remove some false positives and to improve lesions delineation. One process uses FLAIR image when the other uses the intersection of T2-w and FLAIR images as theses modalities are the ones which provide the best results. An example of segmentation result is shown in Figure 1.

	Dice scores	SensL	PPVL	F1 scores
T2-w	0.4010	0.3771	0.1811	0.2102
FLAIR	0.5053	0.5309	0.1803	0.2552
T2-w $\bigcup$ FLAIR	0.4244	0.4932	0.1530	0.2135
T2-w $\bigcap$ FLAIR	0.4396	0.5270	0.1507	0.2166
FLAIR and T2-w $\bigcap$ FLAIR	0.5663	0.4560	0.2871	0.3290

Table 2: Evaluation of our segmentation algorithm with a post-processing step.

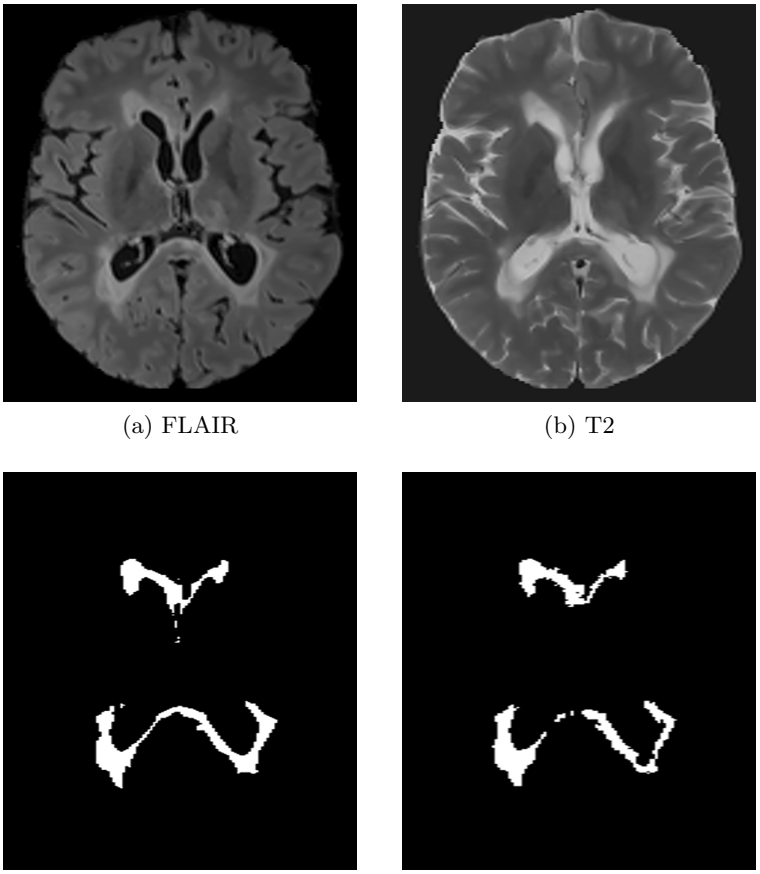
4.3 Implementation and Computation Times

The pipeline presented for the FLI2016 MICCAI MSSEG challenge used the combination of FLAIR and T2 $\bigcap$ FLAIR modalities to perform the segmentation as this process is the one which provides the best results (see Table 2). The algorithm implementation is multi-threaded, based on ITK and available in Anima[3]. The total computation time to process each segmentation of the data set on a computer with an Intel(R) Xeon(R) CPU E5-2660 v3 @ 2.60GHz (8 cores) is approximately 8 minutes.

5 Conclusion

We presented a fully automated MS lesion segmentation method based on intensity normalization and voxel-wise comparison. MS lesion segmentation is a

[3] `https://github.com/Inria-Visages/Anima-Public`



(a) FLAIR (b) T2

(c) Ground truth (d) Automatic segmentation

Fig. 1: Automatic segmentation of the data set 01016SACH

complicated task as MS lesion definition is inter-expert dependent. There is a high variability in the detection of lesions even between ground truth and manual segmentations (dice scores and F1 scores for training data vary respectively from 0.26 to 0.88 and 0.13 to 1). A few MR sequences also have a low initial resolution. This may influence the segmentation workflow and lead to worse results. Consequently, the choice of the optimal modalities used to compute a segmentation and the definition of efficient post-processing steps can be a complicated task.

References

1. Akhondi-Asl, A., Hoyte, L., Lockhart, M.E., Warfield, S.K.: A logarithmic opinion pool based STAPLE algorithm for the fusion of segmentations with associated

reliability weights. IEEE transactions on medical imaging 33(10), 1997–2009 (Oct 2014)
2. Commowick, O., Wiest-Daesslé, N., Prima, S.: Block-matching strategies for rigid registration of multimodal medical images. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). pp. 700–703 (May 2012)
3. Commowick, O., Fillard, P., Clatz, O., Warfield, S.K.: Detection of DTI White Matter Abnormalities in Multiple Sclerosis Patients. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008, pp. 975–982. No. 5241 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Sep 2008), dOI: 10.1007/978-3-540-85988-8_116
4. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. IEEE Transactions on Medical Imaging 27(4), 425–441 (Apr 2008)
5. 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 (Jan 1996)
6. Karpate, Y., Commowick, O., Barillot, C.: Robust Detection of Multiple Sclerosis Lesions from Intensity-Normalized Multi-Channel MRI (Feb 2015)
7. Karpate, Y., Commowick, O., Barillot, C., Edan, G.: Longitudinal Intensity Normalization in Multiple Sclerosis Patients. In: Linguraru, M.G., Laura, C.O., Shekhar, R., Wesarg, S., Ballester, M.A.G., Drechsler, K., Sato, Y., Erdt, M. (eds.) Clinical Image-Based Procedures. Translational Research in Medical Imaging, pp. 118–125. No. 8680 in Lecture Notes in Computer Science, Springer International Publishing (Sep 2014), dOI: 10.1007/978-3-319-13909-8_15
8. Manjon, J.V., Coupé, P.: volBrain: An online MRI brain volumetry system. In: Organization for Human Brain Mapping'15. Honolulu, United States (Jun 2015)
9. Notsu, A., Komori, O., Eguchi, S.: Spontaneous Clustering via Minimum Gamma-Divergence. Neural Computation 26(2), 421–448 (Feb 2014)
10. Ourselin, S., Roche, A., Prima, S., Ayache, N.: Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000, pp. 557–566. No. 1935 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (Oct 2000), dOI: 10.1007/978-3-540-40899-4_57
11. Suarez, R.O., Commowick, O., Prabhu, S.P., Warfield, S.K.: Automated delineation of white matter fiber tracts with a multiple region-of-interest approach. NeuroImage 59(4), 3690–3700 (Feb 2012)
12. 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 (Jun 2010)

Automatic multiple sclerosis lesion segmentation with P-LOCUS

Senan Doyle[1], Florence Forbes[2], and Michel Dojat[3]

[1] Pixyl Medical, Grenoble, France
[2] Inria Grenoble Rhône-Alpes, Grenoble, France
[3] INSERM U1216, GIN, Grenoble, France

Abstract. P-LOCUS provides automatic quantitative neuroimaging biomarker extraction tools to aid diagnosis, prognosis and follow-up in multiple sclerosis studies. The software performs accurate and precise segmentation of multiple sclerosis lesions in a multi-stage process. In the first step, a weighted Gaussian tissue model is used to perform a robust segmentation. The algorithm avails of complementary information from multiple MR sequences, and includes additional estimated weight variables to account for the relative importance of each voxel. These estimated weights are used to define candidate lesion voxels that are not well described by a normal tissue model. In the second step, the candidate lesion regions are used to populate the weighted Gaussian model and guide convergence to an optimal solution. The segmentation is unsupervised, removing the need for a training dataset, and providing independence from specific scanner type and MRI scanner protocol.

1 Introduction

MS brain lesion segmentation is important for diagnosis, prognosis, and patient follow-up. Typically, this task is performed manually by a medical expert, however automatic methods are sought to alleviate the tedious, time consuming and subjective nature of manual delineation. Automatic methods are motivated by the demand for large-scale multi-center clinical research studies that require precise, repeatable and cost-efficient analysis. Automatic brain image segmentation remains a challenging task for a number of reasons, including the presence of artefacts, and the heterogeneity of MRI scanner protocol.

Automated or semi-automated MS brain lesion detection methods can be classified according to their use of multiple sequences, *a priori* knowledge about the structure of normal brain, and the specific tissue segmentation model. In most approaches, normal brain tissue prior probability maps are used to help identify lesion as an outlier.

Existing methods frequently avail of complementary information from multiple sequences. For example, lesion voxels may appear hyperintense in one modality and normal in another. This is implicitly used by neurologists when examining data. In an statistical framework, complementary information from different sequences can help to better discriminate data generated by different probabilistic

distributions in a multi-dimensional space. Intensity distributions are commonly modeled as multi-dimensional Gaussian distributions. This provides a way to combine the multiple sequences in a single segmentation task but with all the sequences having equal importance. Given that the information content and discriminative power to detect lesions varies between different MR sequences, we adopt a weighted data model, originally proposed by Forbes *et al* [4], that allows for the identification of atypical lesion voxels and the subsequent inclusion of the lesion class as an additional model component.

2 Weighted Model

We consider a finite set V of N voxels on a regular 3D grid. The intensities observed at each voxel are denoted by $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. Each $\mathbf{y}_i = \{y_{i1}, \ldots, y_{iM}\}$ is itself a vector of M intensity values corresponding to M different MR sequences. The goal is to assign each voxel i to one of K classes considering the observed features data $\mathbf{y}$. For brain tissue segmentation, we consider in general 3 tissues plus some possible additional classes to account for lesions in pathological data. We denote the hidden classes by $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$, and the set in which $\mathbf{z}$ takes its values by $\mathcal{Z}$. Typically, the $\mathbf{z}_i$'s take their values in $\{1 \ldots K\}$. We consider non-negative weights $\omega = \{\omega_i, i \in V\}$ in a state space denoted by $\mathcal{W}$ and with $\omega_i = \{\omega_{i1}, \ldots, \omega_{iM}\}$. In our general setting the weights are sequence and voxel-specific.

The segmentation task is recast into a missing data framework in which $\mathbf{y}$ are observations and $\mathbf{z}$ are missing variables. Their joint distribution $p(\mathbf{y}, \mathbf{z}|\omega; \psi)$ is governed by the weights $\omega \in \mathcal{W}$ and parameters $\psi \in \underline{\Psi}$, which are both unknown and need to be estimated. A prior distribution $p(\omega)$ is defined on the weights, considered additional missing variables. Denoting the parameters by $\psi = \{\beta, \phi\}$, we assume that the joint distribution $p(\mathbf{y}, \mathbf{z}, \omega; \psi)$ is a MRF with the following energy function:

$$H(\mathbf{y}, \mathbf{z}, \omega; \psi) = H_{\mathbf{Z}}(\mathbf{z}; \beta) + H_W(\omega) + \sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi)$$

where the energy term $H_W(\omega)$ involving only ω does not depend on ψ and the $g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi)$s are probability density functions of $\mathbf{y}_i$.

For the data term $\sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi)$ in (1), we consider M-dimensional Gaussian distributions with diagonal covariance matrices. For each class k, ${}^t(\mu_{k1}, \ldots, \mu_{kM})$ is the mean vector and $\{s_{k1}, \ldots, s_{kM}\}$ the covariance matrix components. When $\mathbf{z}_i = k$, then $\mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, s_{\mathbf{z}_i m})$ represents the Gaussian distribution with mean μ_{km} and variance s_{km}. The whole set of Gaussian parameters is denoted by $\phi = \{\mu_{km}, s_{km}, k = 1 \ldots K, m = 1 \ldots M\}$. Our data term is then defined by setting

$$g(\mathbf{y}_i|\mathbf{z}_i, \omega_i; \phi) = \prod_{m=1}^{M} \mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, \frac{s_{\mathbf{z}_i m}}{\omega_{im}}) \ ,$$

which is proportional to $\prod_{m=1}^{M} \mathcal{G}(y_{im}; \mu_{\mathbf{z}_i m}, s_{\mathbf{z}_i m})^{\omega_{im}}$. Intuitively, the impact of a larger ω_{im} is to give more importance to the intensity value y_{im} in the model. A weight of one recovers the standard multivariate Gaussian case.

The missing data term $H_{\mathbf{Z}}(\mathbf{z}; \beta)$ in (1) is set to a standard Potts model, with external field ξ and spatial interaction parameter η, and whose energy is

$$H_{\mathbf{Z}}(\mathbf{z}; \beta) = \sum_{i \in V} (\xi_{i\mathbf{z}_i} + \sum_{j \in \mathcal{N}(i)} \eta \, \langle \mathbf{z}_i, \mathbf{z}_j \rangle),$$

where $\mathcal{N}(i)$ denotes the voxels neighboring i and $\langle \mathbf{z}_i, \mathbf{z}_j \rangle$ is 1 when $\mathbf{z}_i = \mathbf{z}_j$ and 0 otherwise. Parameter $\beta = \{\xi, \eta\}$ with $\xi = \{^t(\xi_{i1} \dots \xi_{iK}), i \in V\}$ being a set of real-valued K-dimensional vectors and η a real positive value.

The weights are assumed independent from parameters ψ and independent across modalities. The simplest choice is to define a prior $p(\omega) = \prod_{m=1}^{M} \prod_{i \in V} p(\omega_{im})$ where each $p(\omega_{im})$ is a Gamma distribution with hyperparameters α_{im} (shape) and γ_{im} (inverse scale). Thus

$$H_W(\omega) = \sum_{m=1}^{M} \sum_{i \in V} ((\alpha_{im} - 1) \log \omega_{im} - \gamma_{im} \, \omega_{im}).$$

In practice, the set of hyperparameters is fixed so that the modes of each prior $p(\omega_{im})$ are located at some *expert weights* $\{\omega_{im}^{exp}, m = 1 \dots M, i \in V\}$ accounting for some external knowledge, if available. Formally, we set $\alpha_{im} = \gamma_{im} \, \omega_{im}^{exp} + 1$ to achieve this. The expert weights can be chosen according to the specific task. For example, when voxels with typical lesion intensities are not numerous enough to attract a model component, increasing the expert weight for some of them will help in biasing the model toward the identification of a lesion class.

A solution to the model is found using the Expectation-Maximization (EM) framework [2] combined with a variational approximation for tractability in the presence of Markov dependencies. In particular, the mean field principle provides a deterministic way to deal with intractable MRF models[1].

3 Method

Of the four possible input sequences available, the method uses only the unprocessed T1-weighted and Flair. The images are masked, co-registered and corrected for inhomogeneities using the N4 algorithm.

The segmentation process consists of two stages, as detailed in [3]. In the first step, we set $K = 3$, considering only the three normal tissue classes (with all ω_{im}^{exp} and γ_{im} set to 1). The ξ parameters in the MRF prior are set to $\xi_{ik} = \log f_{ik}$ where f_{ik} is the normalized value given by a normal tissue atlas. The interaction parameter η is estimated using a stochastic gradient descent method as specified in [1]. The estimated weights for the Flair sequence are thresholded

at a value of one to identify outlier regions corresponding to candidate lesion regions. This candidate region is refined using additional intensity, location and size constraints, as in [7, 5, 6]. Retained lesions are hyperintense in Flair, confined to white matter and greater than $5mm^3$.

In the second step, the candidate region is used to specify the parameters of the weight distribution in a $K = 4$ segmentation setting. We set γ_{im} according to: $\gamma_{im} = \gamma_{\mathcal{L}}$ for all $i \in \mathcal{L}$ and $\gamma_{im} = \gamma_{\bar{\mathcal{L}}}$ for all $i \notin \mathcal{L}$, where $\gamma_{\mathcal{L}}$ and $\gamma_{\bar{\mathcal{L}}}$ are values to be specified. We set $\gamma_{\bar{\mathcal{L}}} = 1000$ to express our a priori trust in the estimation of the normal brain tissue classes from the preliminary first step, and set $\gamma_{\mathcal{L}} = 10$ to allow some flexibility in the weight estimation.

The expert weight is fixed to $\omega_{\mathcal{L}} = 2$ and $\omega_{\bar{\mathcal{L}}} = 1$. Large values of $\omega_{\mathcal{L}}$ make the lesion class more representative and handle the possibility of very small lesions, while a small $\omega_{\bar{\mathcal{L}}}$ ensures that the weighting of a large candidate lesion region does not affect the estimation of other classes. A post-processing step removes artefacts based on spatial location.

4 Conclusion

The adaptive weighting model facilitates accurate and robust MS lesion segmentation from T1-weighted and Flair sequences. The advantage of this approach is that the weights, and therefore the 'outliers' are obtained in a multi-sequence framework that provides a more robust estimation of normal tissue parameters. The method is independent of MRI scanner, and does not require training data.

References

1. G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pat. Rec.*, 36(1):131–144, 2003.
2. A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39:1–38, 1977.
3. F. Forbes, S. Doyle, D. Garcia-Lorenzo, C. Barillot, and M. Dojat. A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation. In *AISTATS, Sardignia, May 13-15*, 2010.
4. Florence Forbes, Senan Doyle, Daniel García-Lorenzo, Christian Barillot, and Michel Dojat. Adaptive weighted fusion of multiple mr sequences for brain lesion segmentation. In *ISBI*, pages 69–72. IEEE, 2010.
5. O. Freifeld, H. Greenspan, and J. Goldberger. Lesion detection in noisy MR brain images using constrained GMM and active contours. In *IEEE ISBI*, pages 596–599, 2007.
6. D. Garcia-Lorenzo, L. Lecoeur, D.L. Arnold, D. L. Collins, and C. Barillot. Multiple Sclerosis lesion segmentation using an automatic multimodal graph cuts. In *MICCAI*, pages 584–591, 2009.
7. K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE trans. Med. Ima.*, 20(8):677–688, 2001.

MS Lesion Segmentation Using FLAIR MRI Only

Jesse Knight, April Khademi

Image Analysis in Medicine Lab, School of Engineering, University of Guelph
jknigh04@uoguelph.ca

Abstract. There have been many efforts to automate segmentation of MS lesions in brain MRI, since human delineation is time consuming and error prone. However, most existing methods require multiple coregistered MR sequences, tissue priors or parametric models, and are rarely validated on multi-scanner image databases. In this work, a fast, FLAIR-only lesion segmentation algorithm is proposed, that does not use tissue priors or parametric models. The method uses an edge-based model of partial volume averaging to estimate fuzzy membership profiles of tissue classes. Results are further refined using an upstream image standardization pipeline, and downstream post processing. Lesion segmentation performance is measured on 15 volumes from three different scanners, demonstrating the robustness of the approach.

1 Introduction

Multiple Sclerosis (MS) is a demyelinating autoimmune disease affecting the central nervous system. MR imaging of MS lesions plays an integral role in disease diagnosis, monitoring, and research [1]. Since manual segmentation of MS lesions is subjective, unreliable, and time consuming, there has been significant interest in automating this task [2]. Existing approaches have performed classification based on voxel values from multiple coregistered MR sequences, neighborhood voxel values, normalized coordinates, and tissue prior images [2].

Despite the large number of proposed methods, very few have been validated on images from different scanners, which reflects the challenging variability in clinical images [2]. For instance, the works by Souplet et al. [3], Garcia-Lorenzo et al. [4], and Wang et al. [5] fit parametric models to the graylevel distributions of each tissue class, and use images from a single scanner for validation. However, it has been shown that graylevel distributions are highly specific to the acquiring scanner [6]. Similarly, in the works by Wu et al. [7], Steenwijk et al. [8], and Samaille et al. [9], K-Nearest Neighbour classification is used with features from multiple MR sequences; yet without normalization of graylevels across scanners, these approaches are unlikely to generalize. There are also additional challenges associated with using multiple MRI modalities or tissue priors, including the need for accurate registration, additional scan time, limited retrospective image availability, and increased segmentation model complexity.

Here we present an edge-based segmentation method, based on previous work [6, 10] and apply it to MS lesion segmentation. Our method overcomes many of the challenges outlined above, since it does not rely on parametric graylevel distribution models, it employs only the most lesion-discriminative MRI sequence, Fluid Attenuation Inversion Recover (FLAIR), and it does not require registration to any additional templates or tissue priors. We also demonstrate reliability of the method on images from three different MR scanners.

2 Methods

2.1 Image Data & Manual Segmentations

This work uses the 15 raw FLAIR images provided in the unprocessed training dataset of the MICCAI 2016 MSSEG Challenge; the brain is isolated in each image, however, using the brain masks provided in the pre-processed dataset. The binary consensus segmentations also provided, being the fusion of 7 expert manual tracings using the LOP-STAPLE method [11], are used as the ground truth for performance evaluation.

2.2 Edge-Based Lesion Segmentation

The proposed image processing pipeline is summarized in Fig. 1. The method takes a single FLAIR image, denoted $Y(x)$ in 3D space, $x = \{x_1, x_2, x_3\} \in \mathbb{Z}^3$. The pipeline consists of image preprocessing, an edge-based fuzzy classification model, thresholding, and post processing. Output lesion masks are then compared to the consensus segmentations using quantitative performance measures.

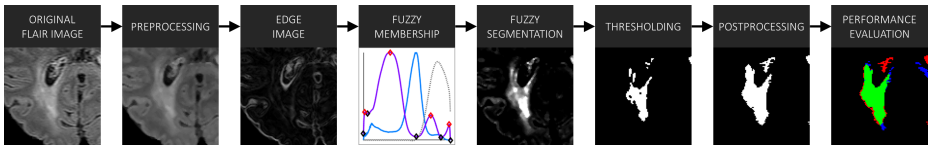


Fig. 1. Image processing pipeine

Preprocessing Preprocessing is employed to correct intensity inhomogeneity and reduce random noise in the source FLAIR images. First, intensity inhomogeneity (bias field) is corrected using the Segment feature in the SPM12 toolbox [12]. While this tool also produces tissue probability images, these are not needed for lesion segmentation and are discarded. A 3D Gaussian low pass filter, with $\sigma = 0.5$ mm, is then used to minimize image noise; it was found that this filter outperformed median, anisotropic diffusion, and bilateral filters.

Fuzzy Classification Next, we perform fuzzy classification using an edge-based model [6, 10]. This model assumes that the graylevels of each tissue class (CSF,

Brain, Lesion) are distributed along a unique range, and that graylevels in between these ranges represent voxels subject to partial volume averaging (PVA, a mixture of two tissue types). For a PVA-affected voxel, comprising a mixture of classes c_j and c_k, with graylevels y_j and y_k, respectively, and where α_{jk} indicates the mixing proportion of class c_j, the graylevel $y_{jk}(x)$ is modelled as [13],

$$y_{jk}(x) = \alpha_{jk}(x) \times y_j(x) + (1 - \alpha_{jk}(x)) \times y_k(x)$$

We take the gradient of this equation to solve for $\alpha_{jk}(x)$, in order to quantify the exact amount of each tissue present in each voxel. However, we observe that the solution in the spatial domain is intractable, since the integration bounds x_0 and x_N are undefined [10],

$$\alpha'_{jk}(x) = \frac{y'_{jk}(x)}{y_j - y_k} \quad \rightarrow \quad \alpha_{jk}(x) = \frac{1}{y_j - y_k} \int_{x_0}^{x_N} y'_{jk}(x)dx$$

To circumvent this, all voxels in the image volume are considered simultaneously in the graylevel (global) domain: $\alpha'(x) \rightarrow \alpha'(y)$, where the integration becomes feasible. The global estimate of $\alpha'(y)$ is computed using the conditional expectation operator, given an initial spatial estimate of edge strength, which is proportional to $\alpha'(x)$.

First, the gradient magnitude image, $G(x)$, is computed using the centered difference kernel, $D = [-1, 0, +1]$, applied in all three dimensions.

$$\begin{aligned} G(x) &= |\nabla_x Y(x)| \\ &= \left[\left(\frac{dY(x)}{dx_1} \right)^2 + \left(\frac{dY(x)}{dx_2} \right)^2 + \left(\frac{dY(x)}{dx_3} \right)^2 \right]^{\frac{1}{2}} \end{aligned}$$

For a robust estimate of the spatial domain edge strength, histogram equalization is performed to ensure that edge magnitudes are distributed consistently and without outliers, yielding $\alpha'(x)$,

$$\alpha'(x) = CDF_G(G(x))$$

Next, the distribution of edge strength in the graylevel domain, $\alpha'(y)$, is estimated as the expected value of edge in the image for each graylevel. The expectation is computed robustly using a 2-bin Gaussian kernel density estimate with bins $p(\alpha'(x) = 1 \mid y)$ and $p(\alpha'(x) = 0 \mid y)$, and $\sigma = 1/6$ (see [10] for derivation),

$$\begin{aligned} \alpha'(y) &= \mathbb{E}\{\alpha'(x) \mid y\} \\ &= \sum_{a \in 0,1} a \cdot p(\alpha'(x) = a \mid y) \\ &= p(\alpha'(x) = 1 \mid y) \end{aligned}$$

Postprocessing of $\alpha'(y)$ consists of smoothing, and trimming to remove the top 0.02% of brain voxels. This reliably yields a distribution like that shown in Fig.

2, top, dark purple, in which local minima (black diamonds) represent graylevels corresponding to pure tissue classes minimal edge information and maxima (red diamonds) represent class transitions edgy regions. Before integration to solve for $\alpha(y)$, we refine: $\alpha'(y) \rightarrow \tilde{\alpha}'(y)$ to yield a more robust estimate of pure tissue and PVA ranges: extrema are identified using peak detection, and locally normalized to {0,1}, and the magnitude is squared, as shown in Fig. 2, top, light purple. This $\tilde{\alpha}'(y)$ is then used to find the partial volume fraction $\alpha(y)$

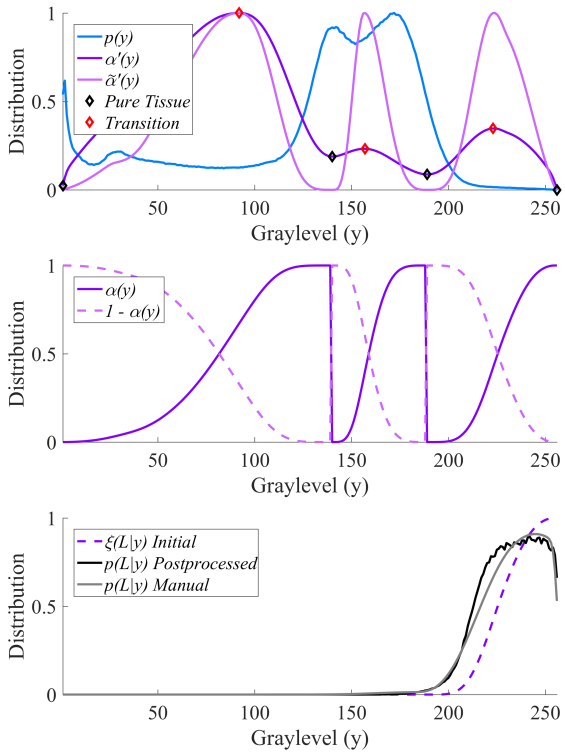


Fig. 2. Graylevel distributions for one FLAIR volume; top - histogram, edge distribution, and extrema; middle - partial volume fraction from piecewise integration of the edge distribution; bottom - distributions of lesion: initial fuzzy estimate, post-processed, and manual estimate.

for each pair of adjacent tissue classes, since the desired integration can now be performed between pure tissue graylevels y_j, y_k (local minima),

$$\alpha_{jk}(y) = \frac{\sum_{y_k}^{y} \tilde{\alpha}'(y)dy}{\sum_{y_k}^{y_i} \tilde{\alpha}'(y)dy}, \quad y \in [y_k, y_i]$$

Finally, the partial volume fraction profiles for each pair of adjacent tissue classes, $\alpha_{jk}(y)$ (Fig. 2, middle) are combined in smooth piecewise segments to give the fuzzy membership function for each tissue class, $\xi(c_k \mid y)$ (e.g. Fig. 2, bottom),

$$\xi(c_k \mid y) = \begin{cases} 0, & y \in [0, y_i) \\ \alpha_{jk}(y), & y \in [y_i, y_k) \\ 1, & y = y_k \\ 1 - \alpha_{jk}(y), & y \in (y_k, y_\ell] \\ 0, & y \in (y_\ell, y_{max}] \end{cases}$$

This smooth characterization of $\xi(c_k \mid y)$ differs from previous works [6, 10], where graylevel ranges corresponding to pure tissues and PVA were estimated using piecewise *linear* functions. This new formulation ensures that graylevels with minimum edge correspond to pure classes, and that maximum fuzzy membership slope (class transition) occurs at the maximum edge content $\alpha(y)$ between pure tissue classes, reflecting the edge-based estimate of PVA distribution.

Post Processing The initial fuzzy classification is thresholded at a level $\tau \in \{0, 1\}$ to give a binary segmentation image, denoted A. Three false positive reduction (FPR) strategies are then appended to the pipeline to refine A based on prior knowledge. The FPR exclusion criteria include:

- Minimum lesion volume (using 6-connectedness, MLV, mm^3)
- Minimum distance from the brain edge (3D, DBE, mm)
- Minimum distance from the brain midline (2D, DBM, mm)

False negatives are also reduced using region growing (RG), in an attempt to mimic the experts' inclusion of marginally hyperintense white matter (also called dirty-appearing white matter) which surrounds lesion cores. During each of N_{RG} region growing iterations, each neighboring voxel at location x is added to the segmentation A if the graylevel is acceptably close to an adaptive threshold: the q^{th} quantile of the graylevels in the original segmentation:

$$\textbf{if}: \left| y(x^*) - Q_{Y(x \in A)}(q) \right| < \Psi, \quad \textbf{then}: A \leftarrow A \cup x^*$$

The fuzzy threshold τ, FPR constraints, and RG parameters were all optimized using a semi-guided simplex gradient ascent with respect to group DSC.

2.3 Performance Analysis

Performance of the algorithm for lesion segmentation is evaluated using the overlap metrics recommended by the MSSEG Challenge. These compare automated segmentation image $A(x)$ to the ground truth segmentation image $T(x)$:

- Dice Similarity Coefficient: $2\frac{A \cap T}{A+T}$
- Positive Predictive Value: $\frac{A \cap T}{A}$
- True Positive Rate: $\frac{A \cap T}{T}$

Additionally, volume agreement with expert lesion load (LL) is illustrated using a Bland-Altman plot, and analysis of covariance (ANCOVA) is used to test for significant impacts of scanner and LL on DSC.

3 Results

This section presents segmentation performance data, following initial fuzzy segmentation, and postprocessing using the optimized parameters, as well as factors influencing performance Performance. The mean DSC for all 15 training cases was 0.60 (Table 1), but for the 10 cases with LL < 5 ml, it increased to 0.70. Accordingly, analysis of covariance revealed that DSC was significantly correlated with LL, independent of scanner ($p = 0.013$), while performance was not significantly impacted by scanner when controlling for LL ($p = 0.354$).

Table 1. Performance metrics by case

LL (ml)	Case #	DSC	PPV	TPR
72	1	0.81	0.80	0.82
49	7	0.78	0.84	0.74
43	3	0.78	0.82	0.75
39	13	0.61	0.83	0.49
30	5	0.66	0.68	0.66
29	15	0.71	0.85	0.62
22	2	0.63	0.72	0.56
13	10	0.64	0.99	0.48
12	11	0.69	0.88	0.57
6	14	0.68	0.92	0.54
5	9	0.26	1.00	0.15
4	8	0.40	0.92	0.26
2	12	0.32	0.25	0.44
2	6	0.34	0.98	0.21
Mean		0.60	0.80	0.53
Mean (LL>5)		0.70	0.83	0.62

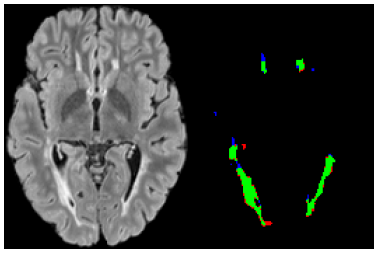


Fig. 3. Sample segmentation; green: true positives, blue: false negatives, red: false positives

An example segmentation is shown in Fig. 3, where it can be seen that false positives arise mostly along the lesion borders, while false negatives are attributable to small, ambiguous lesions near the gray matter. The Bland Altman plots (Fig. 4), demonstrate a potential bias of the method towards underestimation of LL (mean differ-ence of 3.6 ml), and a small positive correlation between LL and undersegmentation (slope of 0.97). This result is corroborated by higher mean PPV (0.80) than TPR (0.53), over all cases, indicating higher specificity than sensitivity.

4 Discussion

The edge-based segmentation algorithm performs well on the training data provided, given the large intensity inhomogeneity and low resolution in the MRI volumes. We observed that the ground truth distribution of lesion versus graylevel,

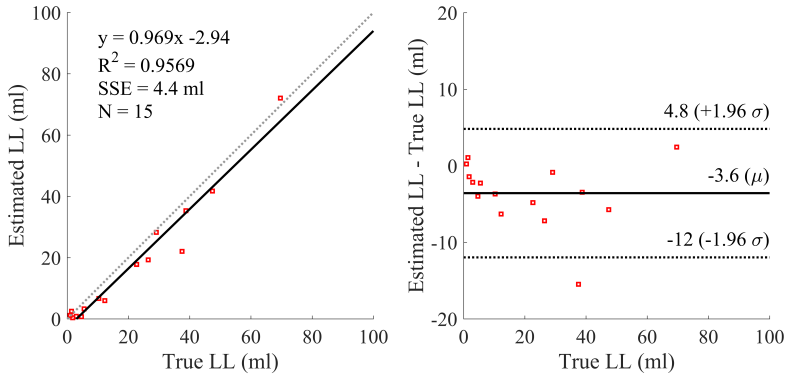


Fig. 4. Bland-Altman plots showing volume agreement with expert segmentations

$p(L \mid y)$ (Fig. 2, bottom), using the binary consensus segmentations, rarely reaches a value of 1, and has long, gradual slopes towards the maximum. This likely arises from the ambiguous segmentation of so-called dirty appearing white matter, and it means that even optimal thresholding results would incur significant false positive and false negative fractions. Pre- and post-processing techniques both help to reduce these effects, and it is worth noting that our preprocessing pipeline resulted in better separability of classes by graylevel than the preprocessed images provided for the competition.

As with most MS lesion segmentation algorithms, the proposed method is challenged by small lesion loads. However, performance was not impacted by scanner or the associated voxel resolution, demonstrating the robustness of this efficient nonparametric approach to modeling FLAIR image variability. Many of the false positives occur either in the gray matter (due to bias field) or as CSF pulsation artifacts, so future works could investigate the use of tissue priors, improved inhomogeneity correction, or multiscale estimation of edge content. However, in hoping to avoid expensive and error-prone image registrations in our approach, we are investigating primarily multiscale or context-based edge characterization, and an iterative bias-field correction with lesion masking.

5 Conclusion

In conclusion, we present a fast, robust segmentation algorithm capable of delineating MS lesions in FLAIR MRI alone. The method achieves good performance (DSC = 0.60) on a challenging database of 15 images from 3 different scanners, and even better performance (DSC = 0.70) on lesion loads greater than 5 ml. Unlike many segmentation tools, the algorithm does not rely on parametric models of tissue class graylevel distributions, and consequently demonstrates no change in performance with different scanners. Additionally, the method requires only one MRI sequence, FLAIR, and no additional tissue priors or training data, thereby avoiding additional MR scan time, limited image availability, and the need for expensive and imperfect registration.

References

1. Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F.D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A.J., Waubant, E., Weinshenker, B., Wolinsky, J.S.: Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of Neurology **69**(2) (2011) 292–302
2. 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) (2013) 1–18
3. Souplet, J., Lebrun, C., Ayache, N., Malandain, G.: An Automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions. MICCAI Grand Challenge Workshop: Multiple Sclerosis Lesion Segmentation Challenge (2008) 1–11
4. García-Lorenzo, D., Prima, S., Arnold, D.L., Collins, D.L., Barillot, C.: Trimmed-likelihood estimation for focal lesions and tissue segmentation in multisequence MRI for multiple sclerosis. IEEE Transactions on Medical Imaging **30**(8) (2011) 1455–67
5. Wang, R., Li, C., Wang, J., Wei, X., Li, Y., Zhu, Y., Zhang, S.: Automatic segmentation and volumetric quantification of white matter hyperintensities on fluid-attenuated inversion recovery images using the extreme value distribution. Neuroradiology **57**(3) (2015) 307–320
6. Khademi, A., Venetsanopoulos, A., Moody, A.R.: Robust white matter lesion segmentation in FLAIR MRI. IEEE Transactions on Bio-Medical Engineering **59**(3) (2012) 860–871
7. Wu, Y., Warfield, S.K., Tan, I.L., Wells, W.M., Meier, D.S., van Schijndel, R.A., Barkhof, F., Guttmann, C.R.G.: Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. NeuroImage **32**(3) (2006) 1205–1215
8. Steenwijk, M.D., Pouwels, P.J.W., Daams, M., van Dalen, J.W., Caan, M.W.A., Richard, E., Barkhof, F., Vrenken, H.: Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). NeuroImage. Clinical **3** (2013) 462–9
9. Samaille, T., Fillon, L., Cuingnet, R., Jouvent, E., Chabriat, H., Dormont, D., Colliot, O., Chupin, M.: Contrast-based fully automatic segmentation of white matter hyperintensities: method and validation. PloS one **7**(11) (2012) e48953
10. Khademi, A., Venetsanopoulos, A., Moody, A.R.: Generalized method for partial volume estimation and tissue segmentation in cerebral magnetic resonance images. Journal of Medical Imaging **1**(1) (2014) 14002
11. Akhondi-Asl, A., Hoyte, L., Lockhart, M.E., Warfield, S.K.: A logarithmic opinion pool based STAPLE algorithm for the fusion of segmentations with associated reliability weights. IEEE Transactions on Medical Imaging **33**(10) (2014) 1997–2009
12. Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage **26**(3) (2005) 839–851
13. Santago, P., Gage, H.D.: Quantification of MR brain images by mixture density and partial volume modeling. IEEE Transactions on Medical Imaging **12**(3) (1993) 566–574

Automatic Multiple Sclerosis Lesion Segmentation Using Hybrid Artificial Neural Networks

Amirreza Mahbod, Chunliang Wang, Örjan Smedby

STH Medical Imaging and Visualization Group
KTH Royal Institute of Technology
Stockholm, Sweden

Abstract. Multiple sclerosis (MS) is a demyelinating disease which could cause severe motor and cognitive deterioration. Segmenting MS lesions could be highly beneficial for diagnosing, evaluating and monitoring the disease progression. To do so, manual segmentation, performed by experts, is often performed in hospitals and clinical environments. Although manual segmentation is accurate, it is time consuming, expensive and might not be reliable.
The aim of this work was to propose an automatic method for MS lesion segmentation and evaluate it using brain images available within the MICCAI MS segmentation challenge.
The proposed method employs supervised artificial neural network based algorithm, exploiting intensity-based and spatial-based features as the input of the network. This method achieved relatively accurate results with acceptable training and testing time for training datasets.

Keywords: Multiple Sclerosis Segmentation, Artificial Neural networks, Machine Learning, MRI

1 Introduction

Multiple Sclerosis (MS) lesion is a demyelinating disease which appears in white matter (WM) region of the brain and causes various motor and cognitive impairments [1]. One of the most accurate imaging tools for diagnosing MS is magnetic resonance imaging (MRI), which provides not only qualitative but also quantitative evidences for assessing disease development and treatment efficacy.

Segmentation of MS lesions in MR scans is a crucial step of quantitative disease analysis. The most common method for that is manual segmentation which is also regarded as the "gold standard" method for distinguishing lesions and other brain tissues. Although this method is accurate, it is a tedious and complex procedure that is not practical for analyzing large amount of MRI datasets

in clinical practice. Moreover the manual approach also suffers from bad reproducibility due to large intra- and inter-observer variation [2,3].

Difficulties with manual segmentation led to development of various segmentation methods with different level of accuracy and complexity [1]. Among these methods, machine learning-based segmentation algorithms are popular choices for MS lesion segmentation and in recent MS segmentation studies, they have been proved to be superior to other conventional methods [1,4].

Recently, a machine learning method for brain segmentation using shape context [5] has been developed and evaluated within the MRBrainS13 challenge [6], providing rather accurate results with relatively short training time. The aim of this paper is to test a similar segmentation framework for MS lesion segmentation. The proposed algorithm is based on artificial neural networks (ANN) which exploits conventional spatial-based and intensity-based features. The implemented method is applied on the MICCAI MS segmentation challenge [7] and the results are compared to the segmentation performed by clinical experts.

2 Method

The goal of MS lesion segmentation is to divide the image into two meaningful, homogeneous and non overlapping regions which represent the lesion and its background. The proposed segmentation method for this paper consists of several parts which are shown in Fig. 1. While the main structure of the algorithm is similar to [5] , it slightly differs regarding the classifier and feature extraction parts. In the following, the details of different parts of the algorithm are discussed in detail.

2.1 Image Data and Ground Truth

In this project, data from the MICCAI challenge for MS segmentation [7] were used which contained totally 53 datasets from 4 different sites and 4 different MRI scanners (GE Discovery 3T, Philips Ingenia 3T, Siemens Aera 1.5T and Siemens Verio 3T), in total 15 1.5T and 38 3T MR brain exams. Each exam consisted of 4 series of MRI images which are 3D FLAIR, 3D T1-weighted, 3D T1-weighted GADO an 2D PD-/T2-weighted scans. Both preprocessed and unprocessed datasets were provided by the challenge organizer. Preprocessed scans were denoised, registered, brain extracted and bias field corrected with state-of-the-art preprocessing algorithms. The Philips Ingenia, Siemens Aera and Siemens Verio MRI volume contains $144 \times 512 \times 512$, $128 \times 224 \times 256$, $261 \times 336 \times 336$ voxels respectively. From all datasets, 15 cases were provided with manual segmentation (ground truth) for training purpose.The raw and segmented images for the remaining 38 datasets were kept away from the participant for evaluating the performance of each proposed methods and also testing on infrastructure which was provided by organizer. It also provided fair comparison between different participants in terms of running time and consistent parameters for all datasets. For each dataset, 7 manual segmented scans, performed by different experts and

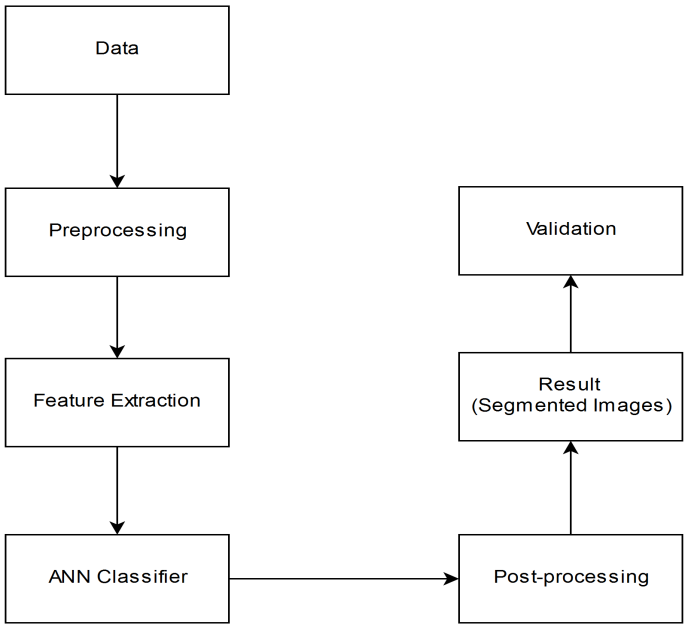


Fig. 1: Generic flow chart of proposed method for MS lesion segmentation.

also consensus manual segmentation were provided which contained two classes (lesion and background). However, only consensus manual segmented scans were used as ground truth in this paper.

2.2 Preprocessing

The aim of preprocessing is to prepare the data in suitable way to be fed to the classifier. Since datasets were already preprocessed, only a few preprocessing steps were applied on each volume which are described in the following.

- **Histogram matching**: In this step the histograms of each group of datasets (from different scanners) were matched to the first dataset in that group. This step is important for preventing the network from being confused by different histogram shapes. Histogram matching was performed using the Insight Segmentation and Registration Toolkit (ITK) [8].
- **Bias field correction**: Intensity inhomogeneity artifact was removed using ITK [8] . However, this step is not of high importance for the datasets used in this project since they were already corrected.
- **Removing extremely low intensity values from training and testing dataset**: Outlier intensity might have an adverse impact on the segmentation results. Therefore, the 4th percentile of each dataset was calculated as the lower boundary, and all voxels below this calculated value were clipped off and mapped to the derived boundary.

- **Normalizing data** : Since intensities from different channels within a patient and also intensities of the same channel for different patients could have large variability, normalization is a critical step to be applied on data. Therefore, in this work all images were normalized with zero mean and standard deviation of one.

2.3 Feature Extraction

As the input of networks, spatial-based and intensity-based features were extracted. These features are based on [9] and are described as follows:

- Intensity of 3D FLAIR channel
- The intensity after convolution with Gaussian kernels with $\sigma = 1, 2, 3\ mm^2$
- The gradient magnitude of the intensity after convolution with Gaussian kernels with $\sigma = 1, 2, 3\ mm^2$
- The Laplacian of the intensity after convolution with Gaussian kernels with $\sigma = 1, 2, 3\ mm^2$
- Spatial information of all voxels (x, y, z) which were divided by the length, width, and height of the brain respectively.

2.4 Classification

Artificial Neural Networks were uses as the main learning algorithm of the proposed algorithm. While Support Vector Machine (SVM), Self Orginizing Map (SOM) and multi layer perceptron (MLP) were implemented in the frame of this work, the testing results on training datasets from MLP were more accurate with acceptable training time. Thus, this network was chosen as the main ANN classifier in this paper.

The proposed network consisted of one hidden layer with 100 hidden nodes. The learning rule for updating weights was the generalized delta rule which calculated the weight updates from output to input layer (error back-propagation)[10]. In the training phase, 50000 voxels were randomly selected from training datasets. Out of all training samples, around 90 % of samples were randomly selected within the brain region and the resting samples from MS lesions. The training label includes 2 classes, namely MS lesions and background. The ANN training was set to stop after 1000 iterations.

The output of the network contained the information of the voxel classes which was used for final segmentation as described in the next section.

2.5 Image Segmentation and Post-processing

For segmenting the test dataset, all voxels are passed into the ANN, which in turn outputs the probabilities of the voxel belonging to the 2 classes. The voxel is assigned to the class that gives the highest probability. The only post-processing step that was applied on classifier results was setting everything outside the

brain mask as background since the brain masks were provided for all datasets by the organizer. However, using other conventional post-processing methods such as using morphological filters and using prior knowledge that MS lesions could be only located in WM could be beneficial to increase the accuracy of the results [3].

2.6 Accuracy Analysis

Several similarity indexes could be used in order to validate the results. In this paper, segmentation results were compared to the ground truth based on Dice similarity index, which indicates how well the two masks overlap with each other. The Dice coefficient can be calculated as:

$$DI\% = \frac{2|A_i \bigcap B_i|}{|A_i| + |B_i|} \tag{1}$$

where i stand for tissue types (e.g. MS-lesion, background), A stands for manual segmentation (ground truth) and B for MRI segmentation algorithm. It has the range from 0 (the worst result) to 100 (the best result if A and B are completely identical).

3 Results

Fig. 2 shows the results after segmentation with proposed algorithms for two sample slices from the first datasets for visual inspection. The first column shows sample raw images which were used for feature extraction as input of the network while second column shows the results of proposed segmentation method and the last column shows the misclassified voxels.

The primitive evaluation results according to Dice score are listed in Table 1 for the training datasets which were provided by manual segmentation.

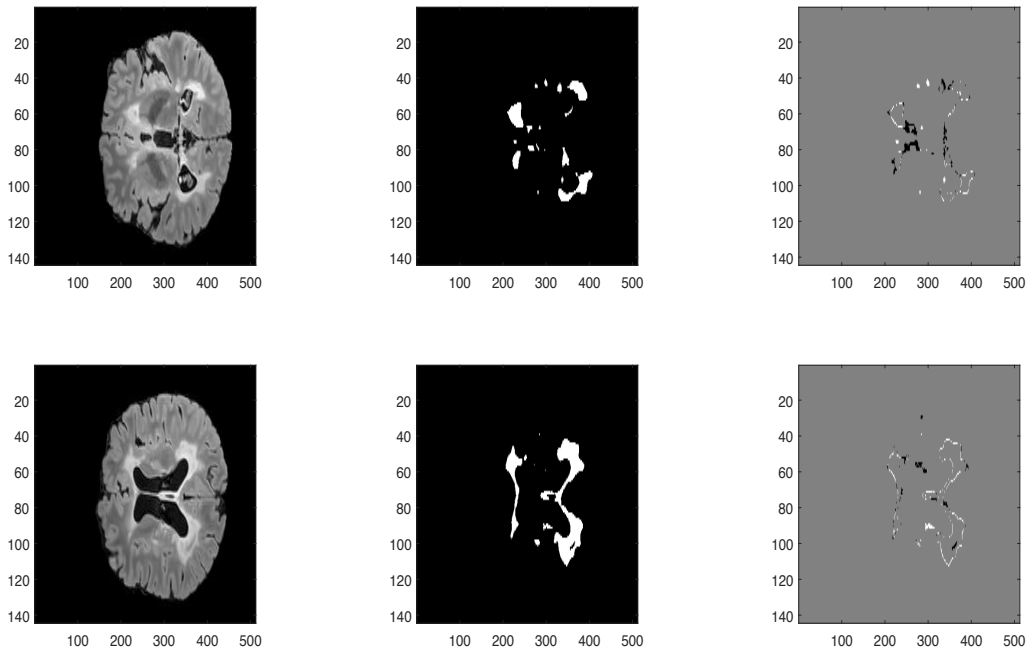


Fig. 2: Examples of segmentation results for two sample slices (312, 337) of first dataset. First column shows the raw 3D FLAIR scans, second column shows the segmentation result from proposed method and the last column shows the misclassified voxels

Table 1: Primitive evaluation based on Dice score.

Test dataset	**Magnetic Field Strength (T)**	**Total Number of lesion voxels**	**Dice Score(%)**
1	3	253175	84.01
2	3	142052	79.07
3	3	79772	61.94
4	3	3344	19.50
5	3	106697	64.34
6	1.5	1348	10.20
7	1.5	35433	67.59
8	1.5	2229	29.42
9	1.5	3491	30.66
10	1.5	9166	63.24
11	3	26680	37.34
12	3	3673	20.45
13	3	96861	70.80
14	3	14219	47.84
15	3	68247	66.60

All experiments were performed on a desktop computer with NVIDIA GTX 980 4GB, 32 GB installed memory and an Intel Xeon E5-2630 2.40 GHz CPU. The main implementation of the algorithm was done with MATLAB (version 2016a).

4 Discussion

The main difference between the proposed method and our previous work for brain segmentation [5] is that the shape context features were eliminated in the current pipeline. In our preliminary test, adding these features improves the segmentation accuracy slightly for the majority of cases (Dice coefficient goes up by 3.5 % on average). However, due to the long processing time for creating these features (20 minutes per case) and difficulty to integrate this part to the evaluation platform, these features were removed from the pipeline. Except this part, the proposed framework is almost identical to the one proposed in [5].

As can be seen from Fig. 2, for two sample slices most of the segmentation errors within the brain occurred in the border of MS lesion and the background, where inter-observer variation of the manual delineation could also be high.

As shown in Table. 1, the accuracy for the 3T scans are better compared to the 1.5T scans, in general. This may be due to lower resolution and tissue contrast of the 1.5T scans which make the segmentation task more challenging in particular at the border of lesion areas. Moreover, the level of accuracy also depends on the size of the lesions. For the datasets with large amount of MS lesions, the accuracy is considerably better compared to the datasets with small MS lesions. This is expected as the numbers of voxels affected by partial volume effects are relative smaller in the cases where the lesions are large.

From our experiment, the network has a tendency to produce relative higher amount of false positive errors than false negative errors. In order to minimize this effect, 90 percent of the training samples were selected from the background and only 10 percent were selected from the lesion area. Boot-strapping is an alternative solution to deal with this problem and could possibly lead to better performance.

5 Conclusion

In this paper, a fully automatic method for MS lesion segmentation has been proposed. The proposed method incorporates the spatial-based and intensity-based features in the machine learning framework to segment the images to the lesion and background. Results of the segmentation are relatively accurate with relative short training and testing time. Further investigation is needed for developing the current method to improve the accuracy.

References

1. M. Styner, J. Lee, B. Chin, M. Chin, O. Commowick, H. Tran, S. Markovic-Plese, V. Jewells, and S. Warfield, "3D segmentation in the clinic: A grand challenge II: MS lesion segmentation," *MIDAS Journal*, vol. 2008, pp. 1–6, 2008.
2. I. D. T, B. Goossens, and W. Philips, "MRI Segmentation of the Human Brain : Challenges , Methods , and Applications," vol. 2015, 2015.
3. S. Damangir, "Segmentation of White Matter Lesions Using Multispectral MRI and Cascade of Support Vector Machines with Active Learning." KTH, School of Computer Science and Communication (CSC), 2011.
4. D. García-Lorenzo, S. Francis, S. Narayanan, D. L. Arnold, and D. L. Collins, "Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging," *Medical image analysis*, vol. 17, no. 1, pp. 1–18, 2013.
5. A. Mahbod, C. Wang, M. Chowdhury, and Ö. Smedby, "Brain Segmentation Using Artificial Neural Networks with Shape Context," *Under Preparation*, 2016.
6. A. Mendrik, K. Vincken, H. Kuijf, M. Breeuwer, W. Bouvy, J. de Bresser, A. Alansary, M. de Bruijne, A. Carass, A. El-Baz, A. Jog, R. Katyal, A. Khan, F. van der Lijn, Q. Mahmood, R. Mukherjee, A. van Opbroek, S. Paneri, S. Pereira, M. Persson, M. Rajchl, D. Sarikaya, O. Smedby, C. Silva, H. Vrooman, S. Vyas, C. Wang, L. Zhao, G. J. Biessels, and M. Viergever, "MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans," *Computat Intellig Neuroscience*, 2015.
7. Website, "MS segmentation challenge using a data management and processing infrastructure." [Online]. Available: https://portal.fli-iam.irisa.fr/msseg-challenge/overview
8. T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, "Engineering and algorithm design for an image processing API: a technical report on ITK-the insight toolkit," *Studies in health technology and informatics*, pp. 586–592, 2002.
9. A. V. Opbroek, F. V. D. Lijn, M. D. Bruijne, A. van Opbroek, F. van der Lijn, and M. de Bruijne, "Automated brain-tissue segmentation by multi-feature SVM classification," *Bigr.Nl*, 2013.
10. S. Marsland, *Machine Learning: an Algorithmic Perspective*. CRC press, 2015.

Nabla-net: a deep dag-like convolutional architecture for biomedical image segmentation: application to white-matter lesion segmentation in multiple sclerosis

Richard McKinley[1], Tom Gundersen[2], Franca Wagner[1], Andrew Chan[3], Roland Wiest[1], Mauricio Reyes[4]

[1] Department of Diagnostic and Interventional Neuroradiology, Inselspital, Bern University of Bern, Switzerland
[2] Red Hat AB, Oslo, Norway
[3] University Clinic for Neurology, Inselspital, University of Bern, Switzerland
[4] Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland

Abstract. Biomedical image segmentation requires both voxel-level information and global context. We present a deep convolutional architecture which combines a fully-convolutional network for local features and an encoder-decoder network in which convolutional layers and maxpooling compute high-level features, which are then upsampled to the resoultion of the initial image using further convolutional layers and tied unpooling. We apply the method to segmenting multiple sclerosis lesions.

Introduction

Medical image segmentation is a fundamental problem in biomedical image computing. Manual segmentation of medical images is both time-consuming and prone to substantial inter-rater error: automated techniques offer the potential for robust, repeatable and fast image segmentation.

In the past few years, neural networks have returned to the fore as the most promising technique for supervised machine learning.(LeCun, Bengio, and Hinton 2015) In particular, convolutional neural networks have dominated the field of image recognition. More recently, techniques for image recognition have been reworked to object location and segmentation. One recurring theme is that segmentation requires a combination of low-level and high-level features, with several techniques and architectures having been suggested for upscaling and incorporating high-level with low-level features. (Hariharan et al. 2015; Long, Shelhamer, and Darrell 2015; Ronneberger, Fischer, and Brox 2015; Brosch et al. 2016)

In this paper, we introduce an architecture, called nabla nets, for image segmentation, with application in the medical image segmentation domain. Nabla net

is a dag-like deep neural network architecture, combining a fully-convolutional pathway learning low-level features and an encoder-decoder network learning high-level features. We describe the general features of nabla-net, its application to the segmentation of multiple sclerosis lesions.

Method

Summary

The fundamental basis of nabla net is a deep encoder/decoder network. Such a network comprises a series of encoder layers, each of which reduces feature maps with higher spatial dimensions to feature maps with lower spatial dimension, followed by a series of decoder layers, which expand features with low spatial dimensions to features with high spatial dimensions. Concretely, for the problem of segmenting MS lesions, the encoder-decoder pathway of the nabla net applied to a 256 * 256 image would compute in the first encoder layer 64 256*256 feature maps, which are then reduced to 64 128*128 feature maps by maxpooling. This is repeated in encoder layer two, yielding 64 64*64 feature maps, and then one further time, yielding 64 32*32 feature maps.

These feature maps are then upscaled by decoder layers, yielding subsequently 64 64*64 feature maps, 64 256*256 feature maps, and finally 64 256*256 feature maps. The upscaling is performed by using "tied unpooling", as described below.

Such an encoder-decoder network can be trained to reconstruct the input image: in this case, the network would be called a "convolutional autoencoder". Such networks have been popular in the past as a method of unsupervised pre-training for image classification, but have largely fallen out of favour, as better methods of assigning network weights have emerged. In this case, we instead train the network to reproduce the MS lesion segmentation: results of training a pure encoder-decoder network on MS lesion data is shown in Figure 1.

The output of the pure encoder-decoder network provides a good localisation of the lesions, but is unable to provide crisp boundaries: for that reason, the final prediction of a nabla net is produced by combining the 64*256*256 layer arising from the first layer encoder layer with the 64*256*256 output of the final decoder layer. These 128 feature maps are then processed by a final fully convolutional layer, before the final prediction of the lesion map is made. This ensures that a combination of low-level and high-level features are available for the prediction.

Techniques used

To produce feature maps, we make use of maxpooling (Jarrett et al. 2009), with non-overlapping pool size 2, in which the feature map is scaled by a factor of two

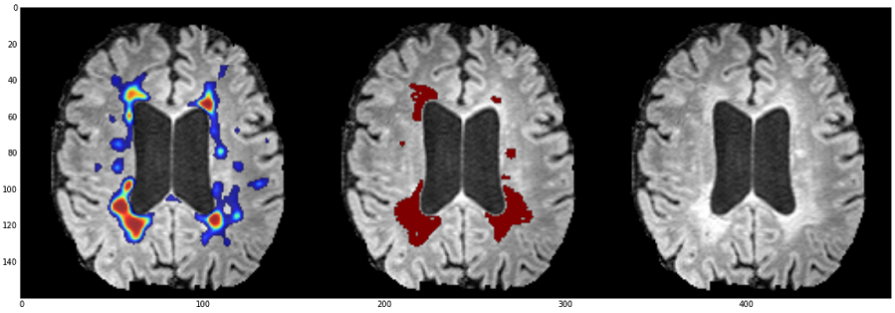


Fig. 1. Example raw output of the pure encoder-decoder network, showing good localisation of the lesions but poor delineation

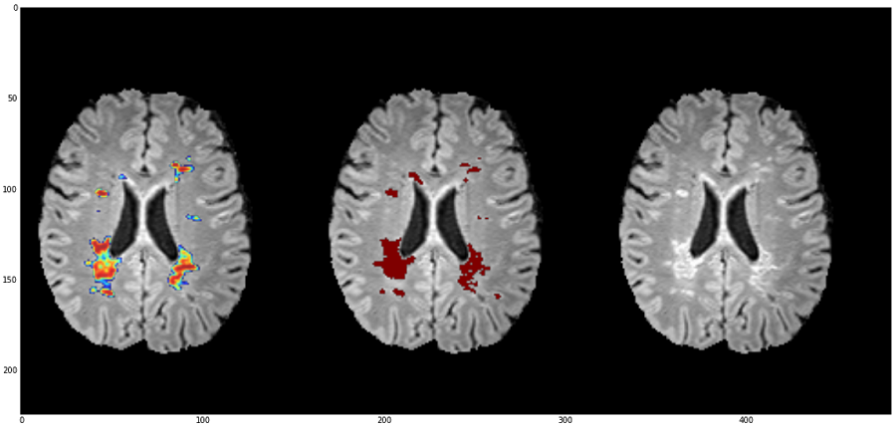


Fig. 2. Example raw output of Nabla net trained on axial FLAIR slices, showing improved lesion outline detection

in both dimensions, by replacing 2 by 2 patches with the maximum intensity in that patch. To scale up the feature maps, we use "tied unpooling" (Zeiler, Taylor, and Fergus 2011; Badrinarayanan, Kendall, and Cipolla 2015), in which feature maps are upscaled by filling only the positions of the maximums of the corresponding maxpooling layer. We use make use of batch normalization (Ioffe and Szegedy 2015) to accelerate the learning process. The architecture of the nabla net makes no use of fully-connected layers, and as such it can be applied to any image with dimensions divisible by eight, this being enforced by the number of maxpooling steps.

The Nabla net Architecture

As stated above, the nabla net is built from encode and decode layers. Each encode layer has the following structure:

Layer	Name	Dimension of output
1	Input	(p, q, r)
2	Zero padding	(64, q+2, r+2)
3	3 by 3 convolutional	(64, q, r)
4	Batch Normalization	(64, q, r)
5	ReLu	(64, q,r)
6	Zero padding	(64, q+2, r+2)
7	3 by 3 convolutional	(64, q, r)
8	Batch Normalization	(64, q, r)
9	ReLu	(64, q,r)

Each decode layer has the following structure

Layer	Name	Dimension of output
1	Input	(p, q, r)
2	Zero padding	(64, q+2, r+2)
3	3 by 3 convolutional	(64, q, r)
4	Batch Normalization	(64, q, r)
5	Zero padding	(64, q+2, r+2)
6	3 by 3 convolutional	(64, q, r)
7	Batch Normalization	(64, q, r)

The whole network has the following structure

Layer	Name	Dimension of output	Comments
1	Input	(5, n*8, m*8)	Five slices,

Layer	Name	Dimension of output	Comments
			dimensions (n*8, m*8)
2	Encode 1	(64, n*8, m*8)	
3	Maxpool 1	(64, n*4, m*4)	
4	Encode 2	(64, n*4, m*4)	
5	Maxpool 2	(64, n*2, m*2)	
6	Encode 3	(64, n*2, m*2)	
7	Maxpool 3	(64, n, m)	
8	Encode 4	(64,n,m)	
9	Decode 4	(64,n,m)	
10	Unpool 3	(64, n*2, m*2)	Tied to Maxpool 3
11	Decode 3	(64, n*2, m*2)	
12	Unpool 2	(64, n*4, m*4)	Tied to Maxpool 2
13	Decode 2	(64, n*4, m*4)	
14	Unpool 1	(64, n*8, m*8)	Tied to Maxpool 1
15	Merge	(128 , n*8, m*8)	Concatenate the outputs of layers 2 and 14
16	Encode Final	(64, n*8, m*8)	
17	1*1 Convolutional	(1,n*8, m*8)	
18	Sigmoid	(1, n*8, m*8)	Loss = binary crossentropy

The network was built using the Keras framework (https://github.com/fchollet/keras), and trained using Theano (http://deeplearning.net/software/theano) as a backend, using an Adadelta optimizer.

A version of Nabla net, trained on skull-stripped FLAIR images from the Inselspital, was applied to the MSSEG training cases, and achieved a mean Dice score of 0.67, and a standard deviation of 0.11. For the challenge, as requested by the organisers, we retrained the system from scratch using only the MSSEG training data, as described below.

Nabla nets for the MSSEG challenge

We trained three copies of nabla net, one for each of the three directions { axial, saggital, coronal }. The networks were trained on the unproicessed FLAIR data, resampled to isotropic 1mm voxel spacing. Each training case comprised the data from five consecutive { axial, saggital, coronal } slices, with such a set of slices being included as training data if the middle slice contained voxels within the brain mask. Ground truth for such a set of slices was given by the lesion mask of the central slice: thus, the lesion mask of a slice was predicted from the slice itself and the four slices surrounding it.

There was a substantial data imbalance between lesion and non-lesion pixels in the training data, owing not only to the size of the lesions, but also to the

presence of pixels outside of the brain mask in the image. Since sampling the training data was not an option (as nabla net operates on whole slices), instead the individual voxels were weighted according to their importance to the model. Rather than simply weighting all lesion voxels (which typically has the effect of moving the optimal decision boundary, but not improving the classifier) we instead calculated, for each case, the 25th percentile of the scaled intensity within the lesion mask, and weighted the loss function from voxels above that intensity ten times more than other voxels.

Training the models

The MSSEG training data is derived from multicentric, multi-scanner data, with data from three different scanner types being available. For each of the three models trained, one example from each scanner type was randomly selected, and the data from those cases used as validation data, to monitor for overfitting.

The models were trained with the Adadelta method. Models were trained until no improvement in loss was seen in the validation set over five epochs (early stopping) and the model with best performance (binary crossentropy) on the validation set was selected for the final model.

Applying to new cases

Given a new case, the lesion mask was predicted, using only the processed FLAIR maps, as follows.

Given a case from the MSSEG training data, the pipeline for lesion segmentation was as follows: resample the processed FLAIR volume to 1mm isovoxels. Apply the saggital, coronal and axial models to the resampled volume (padding the image to ensure that the slice dimensions are divisible by eight). The final lesion heatmap is given by averaging the heatmaps arising from the three models. This heatmap is then resampled to the native voxel spacing of the original volume. An initial segmentation is made by setting all voxels with posterior > 0.5 as lesion voxels, and all voxels with posterior < 0.1 as non lesion: the final segmentation is derived by using a random walk segmentation (beta = 10) as implemented in the python scikit-image package (http://scikit-image.org/).

Processing a single case, on a laptop with an 8Gb NVIDIA GTX980M GPU, took an average of 210s. Processing on the VIP system, which does not have access to GPU accelleration, was substantially slower, on the order of two hours per case.

References

Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. 2015. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." *Cvpr 2015*, 5. doi:10.1103/PhysRevX.5.041024.

Brosch, T, L Tang, Y Yoo, D Li, A Traboulsee, and R Tam. 2016. "Deep 3D Convolutional Encoder Networks with Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation." *IEEE Transactions on Medical Imaging* PP (99): 1. doi:10.1109/TMI.2016.2528821.

Hariharan, Bharath, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. 2015. "Hypercolumns for object segmentation and fine-grained localization." In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June-2015:447–56. doi:10.1109/CVPR.2015.7298642.

Ioffe, Sergey, and Christian Szegedy. 2015. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." *ArXiv:1502.03167*, 1–11. doi:10.1007/s13398-014-0173-7.2.

Jarrett, Kevin, Koray Kavukcuoglu, Marc'Aurelio Ranzato, and Yann LeCun. 2009. "What is the best multi-stage architecture for object recognition?" In *Proceedings of the IEEE International Conference on Computer Vision*, 2146–53. doi:10.1109/ICCV.2009.5459469.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep learning." *Nature* 521 (7553): 436–44. doi:10.1038/nature14539.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. "Fully convolutional networks for semantic segmentation." In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June-2015:3431–40. doi:10.1109/CVPR.2015.7298965.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–41. doi:10.1007/978-3-319-24574-4_28.

Zeiler, Matthew D., Graham W. Taylor, and Rob Fergus. 2011. "Adaptive deconvolutional networks for mid and high level feature learning." In *Proceedings of the IEEE International Conference on Computer Vision*, 2018–25. doi:10.1109/ICCV.2011.6126474.

Prediction of MS Lesions using Random Forests

John Muschelli[1], Elizabeth Sweeney[2], Jacob Maronge[3], Ciprian Crainiceanu[1]

[1]Johns Hopkins University, Baltimore MD USA
[2]Rice University, Houston TX USA
[3]University of Wisconsin, Madison WI USA

Abstract. We describe an analysis pipeline for processing the 2016 MSSEG MICCAI challenge data, where the goal is to identify lesions from patients with multiple sclerosis (MS). From this data, we then create a series of imaging predictors derived from all the imaging sequences (T1-weighted, T2-weighted, FLAIR, and proton density) to build a random forest classifier for MS lesions.

Keywords: random forests, segmentation, multiple sclerosis

1 Introduction

Patients with multiple sclerosis (MS) have white matter lesions in their brains, which magnetic resonance imaging (MRI) scans is sensitive to. Manual segmentation of the MRI scan by an expert reader is the gold standard for volume estimation and localization, but is time-consuming and has within- and across-reader variability. We propose an automated segmentation approach using a random forest algorithm with features extracted from multiple sequences of MRI scans.

2 Methods

2.1 Data

The overall training data consisted of 15 scans from 3 different scanners from the MICCAI 2016 Multiple Sclerosis Segmentation (MSSEG) Grand Challenge (`https://portal.fli-iam.irisa.fr/msseg-challenge/overview`). The data consisted of 4 imaging sequences: 1) a T1-weighted image, 2) a T2-weighted image, 3) Fluid attentuated inversion recovery (FLAIR) image, and 4) and proton density (PD) image. A T1-weighted image was taken before and after injection of gadolinium, denoted as T1-pre and T1-post, respectively. Therefore, each person had 5 images used for prediction. Lesions were manually segmented on the FLAIR image by 7 expert readers and a consensus was derived using the logarithmic opinion pool based STAPLE (LOP-STAPLE) method (Akhondi-Asl et al. 2014).

We used 8 of the participant's data to build a model. The remaining 7 participant scans were used as a test set. For each participant in the test set, the

model predicted a probability of lesions on a voxel level, which was thresholded using a cutoff derived from the training data, resulting in a binary image mask of predicted lesions.

2.2 Preprocessing

All operations were performed using the R statistical language (R Core Team 2015). The fslr package (Muschelli et al. 2015) was used, which relies on FSL (Jenkinson et al. 2012), as well as the ANTsR package (`https://github.com/stnava/ANTsR`), which is an implementation of the Advanced Normalization Tools (ANTs) software ported to R.

The pipeline consisted of only the unprocessed images from the challenge, using the T1-weighted, T2-weighted, proton density, and fluid attenuated inversion recovery (FLAIR) sequences. The unprocessed images were corrected using the N4 inhomogeneity correction (N. J. Tustison et al. 2010). For each imaging sequence, we estimated a brain mask using a simple label fusion approach using 15 templates from the multi-atlas skull stripping software (MASS) (Doshi et al. 2013). Each template was registered to the target image using an affine transformation followed by symmetric normalization (SyN) (B. B. Avants et al. 2008). The template brain mask was transformed into the native space and voxels classified as brain if over 50% of the template masks categorized that voxel as brain. Brain-extracted images were corrected again using the N4 correction to remove any additional inhomogeneity only within the brain.

Brain-extracted and inhomogeneity-corrected images were registered to the FLAIR image of the study using a rigid-body transformation and a Lanczos windowed sinc interpolation.

We normalized the intensity of each scan by subtracting the image by the 20% trimmed mean and dividing the image by the 20% trimmed standard deviation estimated from the intensity of values within the FLAIR brain mask. All imaging predictors were calculated using the intensity-normalized images in the FLAIR space. All predictors were computed on all imaging sequences, unless otherwise specified.

2.3 Imaging Predictors

We derived a set of imaging predictors from each scan. We will describe each predictor together with the rationale for their use. These features make up the potential set of predictors/features for image segmentation.

Normalized Intensity The first predictor is the normalized intensity value in standard deviation units denoted by $x(v)$. This is the main predictor used in visual inspection. For example, hyperintensities on the FLAIR, PD, and T2 images as well as hypointensities of the T1-pre are typically indicative of lesions.

Local Moment Information For each voxel, we extracted a neighborhood, denoted $N(v)$, of all adjacent voxels along the 3 dimensions together with the voxel itself. If $x_k(v)$ denotes the voxel intensity for voxel neighbor k, where $k = 1, \ldots, N(v) = 27$, then the local mean intensity is defined as:

$$\bar{x}(v) = \frac{1}{N(v)} \sum_{k \in N(v)} x_k(v). \tag{1}$$

Based on similar ideas we have also calculated statistics based on higher order moments and define the local standard deviation (SD), skew, and kurtosis as:

$$\begin{aligned}
\mathrm{SD}(v) &= \sqrt{\frac{1}{N(v)} \sum_{k \in N(v)} \{x_k(v) - \bar{x}(v)\}^2} \\
\mathrm{Skew}(v) &= \frac{\frac{1}{N(v)} \sum_{k \in N(v)} \{x_k(v) - \bar{x}(v)\}^3}{\left[\frac{1}{N(v)} \sum_{k \in N(v)} \{x_k(v) - \bar{x}(v)\}^2\right]^{3/2}} \\
\mathrm{Kurtosis}(v) &= \frac{\frac{1}{N(v)} \sum_{k \in N(v)} \{x_k(v) - \bar{x}(v)\}^4}{\left[\frac{1}{N(v)} \sum_{k \in N(v)} \{x_k(v) - \bar{x}(v)\}^2\right]^2}
\end{aligned}$$

We did not divide by $\{N(v) - 1\}$ in standard deviation and skew formula and did not subtract by 3 for kurtosis. As $N(v)$ is the same at every voxel, these simplified choices will have no effect on modeling or prediction.

The higher order moments can provide information about how homogeneous the intensities in the neighborhood are and where edges may be located.

First-pass Segmentation A major advantage of our approach is that it can use the results of other segmentation algorithms as covariates in our model. Consider, for example, Atropos (Brian B Avants et al. 2011), a previously published, open source, general segmentation tool based on Markov random fields for image segmentation. We used Atropos to conduct a 4-tissue class segmentation of the FLAIR image and combined the top 2 probability classes into one class. The probability of the resulting class was then used a predictor, denoted by Atropos(v). Although Atropos has been shown to perform well in other studies for tissue-class segmentation (Brian B Avants et al. 2011, Menze et al. (2015)), the Atropos segmentation did not perform adequately in MS lesion segmentation. However, using the Atropos segmentation probabilities as predictors can be done seamlessly in our approach. Similarly, the results of any other segmentation

approach can be incorporated in our approach and the relative performance of methods can be compared.

Contralateral Difference Images As lesions at a specific location at a specific visit are commonly constrained to one side of the brain, the contralateral side tends to have different intensities values. In contrast, for non-lesion voxels, the contralateral voxels tend to have similar intensity values due to the quasi-symmetry of the brain. To take advantage of this property, we right-left flipped the image, and computed a difference image

$$f(v) = x(v) - x(v^*), \tag{2}$$

where v^* is the contralateral voxel of v.

Global Head Information We also created 3 images by smoothing the original image using large Gaussian kernels ($\sigma = 5mm^3, 10mm^3, 20mm^3$) to account for potential heterogeneity in intensity. These smooth images we denoted by $s_5(v)$, $s_{10}(v)$ and $s_{20}(v)$, respectively.

Standardized-to-template Intensity We have also incorporated predictors that contrast the voxel intensities with those of an average brain obtained from the training scans. More precisely, we registered the T1-weighted image, previously registered the FLAIR image, using affine transformations followed by SyN (B. B. Avants et al. 2008) to this emplate. We transformed each intensity-normalized sequence into template space using this transformation. For each sequence, we created a voxel-wise mean image M and voxel-wise standard deviation S image across the registered images in template space for the training scan.

For all scans in the training and test sets, we created a standardized voxel intensity with respect to this population, z_{template}, using the following equation:

$$z_{\text{template}}(v) = \frac{x(v) - M(v)}{S(v)},$$

where this operation is performed voxel-wise, as denoted by v. This image (z_{template}) was then warped back into the original space to align this predictor with the other predictors using the inverse from the SyN registration step.

Quantile Images Although we believe the z-score normalization creates a standardized intensity value for comparison across participants, differences in distributions still exist. As we commonly are trying to identify high or low-valued voxels in most imaging sequences, we wish to focus on the quantile a voxel falls into in the distribution of all voxels within that brain imaging sequence. Therefore for each sequence s for each participant i, we estimated an empirical

cumulative distribution function (CDF), $F_{i,s}$, from all voxels within the brain mask. Using this, we estimated the quantile of each voxel intensity as follows:

$$q_{i,s}(v) = F_{i,s}^{-1}(x(v))$$

Applying this procedure to all voxels results in a quantile image, which we denote $q(v)$. We believe this may make images more comparable. For example, the fact a voxel has a value of 2 in the FLAIR image may not be as important to discern lesion status compared to the fact it is in the top 10% of the FLAIR intensities over the brain.

2.4 Model Estimation

The full training data for model/algorithm creation consisted of all predictors in the training scans for all voxels within each participant's brain mask. We subsampled $100,000$ voxels from each participant in the training data to fit models, denoted as fitting voxels. On these fitting voxels, we fit random forest classifier (Breiman 2001) with the `randomForest` package (Liaw and Wiener 2002) using the default parameters and 1000 trees.

2.5 Method Evaluation

For each voxel not within the fitting voxels, we estimated a probability that voxel was lesion. Using these probabilities and the true label for lesion, we optimized a probability cutoff using the Dice Similarity Index (DSI) (Dice 1945), an overlap measure.

For each participant in the test set, we estimated the probability a voxel was lesion versus non-lesion. Using the cutoff derived from the non-fitting voxels in the training data, we thresholded the probability to create a binary mask of predicted lesions. This mask was compared to the gold standard for that participant using the DSI.

Bibliography

Akhondi-Asl, Alireza, Lennox Hoyte, Mark E Lockhart, and Simon K Warfield. 2014. "A Logarithmic Opinion Pool Based STAPLE Algorithm for the Fusion of Segmentations with Associated Reliability Weights." *IEEE Transactions on Medical Imaging* 33 (10). IEEE: 1997–2009.

Avants, B. B., C. L. Epstein, M. Grossman, and J. C. Gee. 2008. "Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain." *Medical Image*

Analysis, Special issue on the third international workshop on biomedical image registration - WBIR 2006, 12 (1): 26–41. doi:10.1016/j.media.2007.06.004.

Avants, Brian B, Nicholas J Tustison, Jue Wu, Philip A Cook, and James C Gee. 2011. "An Open Source Multivariate Framework for N-Tissue Segmentation with Evaluation on Public Data." *Neuroinformatics* 9 (4). Springer: 381–400.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1). Springer: 5–32.

Dice, Lee R. 1945. "Measures of the Amount of Ecologic Association Between Species." *Ecology* 26 (3): 297–302. `http://www.jstor.org/stable/1932409`.

Doshi, Jimit, Guray Erus, Yangming Ou, Bilwaj Gaonkar, and Christos Davatzikos. 2013. "Multi-Atlas Skull-Stripping." *Academic Radiology* 20 (12). Elsevier: 1566–76.

Jenkinson, Mark, Christian F. Beckmann, Timothy E. J. Behrens, Mark W. Woolrich, and Stephen M. Smith. 2012. "FSL." *NeuroImage* 62 (2): 782–90. doi:10.1016/j.neuroimage.2011.09.015.

Liaw, Andy, and Matthew Wiener. 2002. "Classification and Regression by RandomForest." *R News* 2 (3): 18–22. `http://CRAN.R-project.org/doc/Rnews/`.

Menze, Bjoern H, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, et al. 2015. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)." *Medical Imaging, IEEE Transactions on* 34 (10). IEEE: 1993–2024.

Muschelli, John, Elizabeth Sweeney, Martin Lindquist, and Ciprian Crainiceanu. 2015. "fslr: Connecting the FSL Software with R." *R Journal* 7 (1). R Foundation for Statistical Computing: 163–75.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. `https://www.R-project.org/`.

Tustison, Nicholas J., Brian B. Avants, Philip A. Cook, Yuanjie Zheng, Alexander Egan, Paul A. Yushkevich, and James C. Gee. 2010. "N4ITK: Improved N3 Bias Correction." *IEEE Transactions on Medical Imaging* 29 (6): 1310–20. doi:10.1109/TMI.2010.2046908.

Unsupervised multiple sclerosis lesion detection and segmentation using rules and level sets

Eloy Roura, Mariano Cabezas, Sergi Valverde, Sandra González-Villà, Joaquim Salvi, Arnau Oliver, and Xavier Lladó

Department of Computer Architecture and Technology, University of Girona. Spain.
{eroura,mcabezas,svalverde,sgonzalez,qsalvi,aoliver,llado}@atc.udg.edu

Abstract. Lesion segmentation plays an important role in the diagnosis and follow-up of multiple sclerosis (MS). This task is very time-consuming and subject to intra-inter rater variability. In this paper, we present a tool for automated MS lesion segmentation using only T1w and FLAIR images. Our approach is based on two main steps, initial brain tissue segmentation according to the gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) performed in T1w images, followed by a second step where the lesions are segmented as outliers to normal apparent GM brain tissue on the FLAIR image. After a false-positive reduction step is applied on the detected lesions, their segmentation is refined using a level sets strategy.

1 Introduction

Magnetic resonance imaging (MRI) plays an important role in medical image analysis for both clinical and research studies. Inflammatory demyelinating diseases such as multiple sclerosis (MS) [10] presents plaques (lesions) of demyelination typically observed in conventional MRI. Detecting those lesions is a crucial task for MS diagnosis as stated in the 2010 revision of the McDonald criteria [12]. Thus, a fully automatic tool that can segment the lesions would prevent user variability and reduce the time consumption considerably. In the literature, there is not yet a standard tool feasible for daily clinical practice [9], although many attempts have been proposed thus far [5,7,15,17,18,20]. Automatic detection of MS lesions is a challenging problem [6,9] that is hampered by factors such as diversity among devices, MRI acquisition protocols and case of studies.

In this work, we present a tool that follows the principles of a recently released tool for MS lesion detection [3,13], that was configured and tested for both 1.5T and 3T images. This algorithm is based on two main steps, initial brain tissue segmentation according to gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) in T1w images, followed by a second step where lesions are segmented as outliers to normal apparent GM brain tissue on the fluid attenuated inversion recovery (FLAIR) image. To extend this tool for processing 3T brain volumes, we have changed the bias normalization and tissue segmentation steps. Moreover, we have modified the lesion segmentation process to include three iterations and a level sets approach. These changes allow the reduction of false-positive (FP) detection, maintaining a good true-positive (TP) rate.

2 MS lesion segmentation pipeline

2.1 Preprocessing

Since our pipeline is optimised for a certain preprocessing setup, we use our own preprocessing steps instead of segmenting the preprocessed dataset. This preprocessing includes the following steps:

Skull stripping and tissue segmentation This step is applied by means of the SPM tissue segmentation algorithm [1]. As the result is a probability map, we performed a maximum likelihood between the three main tissue classes: WM, GM and CSF. To obtain the brain mask we apply a probability threshold set at 0.5 [2,14].

Image denoising This step is applied in order to enhance and restore the MRI image. We smooth the image histogram by using the 3D Matlab implementation of the anisotropic diffusion filter of Perona and Malik [11].

Bias correction This step is applied to correct the inhomogeneities. For this purpose, we use a Matlab implementation of the bias field correction proposed by Larsen et al. [8], which implements the well-known non-parametric, non-uniform intensity normalization (N3) method [16] as a Bayesian modeling method.

Intra-subject registration Finally, we used SPM spatial co-registration, estimate and re-slice [14] to correct spacing and misalignments.

2.2 Lesion segmentation

Lesion detection Following the unsupervised strategy presented in Roura et al. [13] (see Figure 1), we look for the hyperintensity regions as the outliers in the GM tissue of the FLAIR image with three iterations. First of all, we need to distinguish among the three main brain tissues. This is already obtained when performing the skull stripping process via the SPM tissue segmentation on T1w images. Afterwards, the hyperintensities are detected by a thresholding in FLAIR images and refinement step which is performed three times, where the first iteration takes into account very large lesions, the second segments large and bright lesions and the third is performed at a lower threshold to look for small lesions. The outliers are computed using a threshold defined as:

$$Thr = \mu + \alpha\sigma \tag{1}$$

where μ and σ are the mean intensity and the standard deviation of the GM histogram (computed by the full width at half maximum), respectively. Candidate lesions are adjusted by the parameter α. As stated in Roura et al. [13], this alpha parameter has a strong impact on the results obtained.

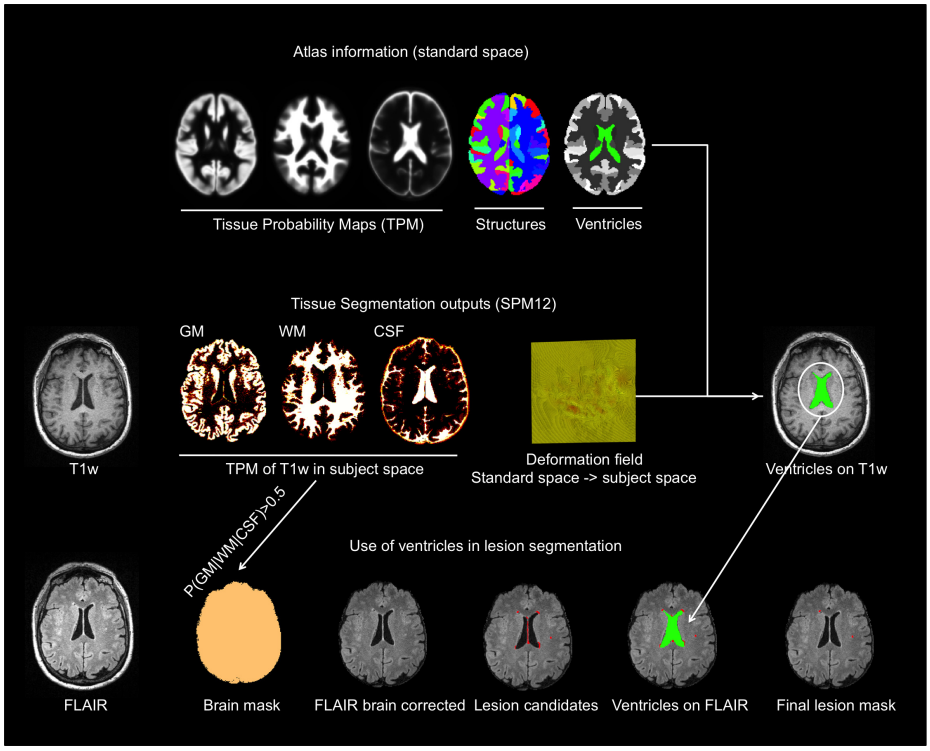


Fig. 1. Scheme of the whole pipeline. The first line represents the atlas information used when segmenting the tissues and the ventricles of the T1w image in the second row. The bottom row shows all the steps over the FLAIR image when segmenting the white matter lesions.

Originally, three post-processing parameters were used to reduce the false positive (FP) lesions after applying this threshold: lesion size (λ_{size}), percentage of lesion voxels belonging to either WM or GM, and percentage of neighbor lesion voxels belonging to WM (λ_{nb}). However, due to the tissue segmentation we discarded the second rule. Moreover, since our first iteration is now focused on detecting large hyper-intense lesions, we only apply the size rule during this first iteration; whereas in the other two following steps, we follow the strategy presented in Roura et al. [13]. After optimisation, the parameters were set as defined in Table 1.

Detection refinement with structure information In the second and third iterations, all the lesions attached to the ventricles' region and presenting elongated shapes are discarded in order to avoid periventricular inflammatory regions. In Figure 1 there is an example of this scenario, where we show the ventricles and some elongated candidate regions that have been removed. We have used the maximum probability tissue labels derived from the MICCAI 2012

Table 1. Parameters for the three iterations of the lesion detection process after optimisation with the training database.

	α	λ_{size}	λ_{nb}
First iteration	2	7 ml	-
Second iteration	2.5	0.035 ml	0.30
Third iteration	2	0.020 ml	0.45

Grand Challenge and Workshop on Multi-Atlas Labeling. The data was labeled by Neuromorphometrics, Inc. using MRI from the OASIS project. This atlas belongs to the standard space used by SMP12 tissue segmentation; therefore, one can easily obtain the deformation fields to this space from any subject space. Since we already performed this process at the early steps, we are able to apply the inverse deformation fields with a nearest-neighbor interpolation to the labels belonging to the ventricles. This procedure allows to pullback any of the brain structures of this atlas (see Figure 1). Once the ventricles have been brought to the subject space, we obtain a smoothness region by applying morphological operations (dilation and erosion).

Lesion segmentation refinement using level sets Finally, the remaining lesion detections are used as the seed regions of the active contours segmentation to improve the segmentation results. We formulate a level set method following the principle of [4] of an active contour model detecting objects with undefined edges, i.e. with a low or unappreciable gradient at the boundaries of the objects. Let $\Phi(x,t)$ be the level set function and $\Phi(x,t) = 0$ be the interface, which bounds the open region Ω, i.e. $\Phi(x,t) = 0 = \delta\Omega$. Therefore, the Φ properties remain as follows:

$$\Phi(x,t) > 0 \text{ for } x \in \Omega^+ \text{ (inside the object} \rightarrow \text{lesion)}$$
$$\Phi(x,t) < 0 \text{ for } x \in \Omega^- \text{ (outside the object} \rightarrow \text{normal tissue)}$$
$$\Phi(x,t) = 0 \text{ for } x \in \delta\Omega$$

We initialize the surface $S \subset \Omega$ as the contour of the initial lesion mask. Therefore, the inside(S) is denoted as $s1 \in \Omega^+$ bounded by S, and the outside(S) as $s2 \in \Omega^-$, belonging to the ventricles. The surface S evolves by minimizing the following energy based-segmentation:

$$\Phi = \Phi + \delta t * \lambda_1 K_s * \lambda_2 F_i \tag{2}$$

where δt is the time step for each iteration, λ_1 is the weight for the smooth term K_s denoting the curvature of the surface [19], and λ_2 is the weight for the image force F_i computed as follows:

$$F_i = -(I - \mu_{s2})^2 + (I - \mu_{s1})^2 \tag{3}$$

3 Discussion

We presented an unsupervised pipeline that uses the unprocessed T1-w and FLAIR images to detect and segment MS. This pipeline integrates pre-processing, detection and false positive reduction in terms of both detection and segmentation without needing any training dataset. In order to do this, three detection iterations are applied in order to detect different lesion topologies by varying the post-processing parameters. Finally, a level sets approach is applied to improve segmentation results.

Acknowledgments. This work has been partially supported by "La Fundació la Marató de TV3", by Retos de Investigación TIN2014-55710-R, and by MPC UdG 2016/022 grant.

References

1. Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage 26(3), 839 – 851 (2005)
2. Boesen, K., Rehm, K., Schaper, K., Stoltzner, S., Woods, R., Lüders, E., Rottenberg, D.: Quantitative comparison of four brain extraction algorithms. NeuroImage 22(3), 1255–1261 (2004)
3. Cabezas, M., Oliver, A., Roura, E., Freixenet, J., Vilanova, J., Ramió-Torrentà, L., Rovira, À., Lladó, X.: Automatic multiple sclerosis lesion detection in brain mri by flair thresholding. Computer Methods and Programs in Biomedicine 115(3), 147–161 (2014)
4. Chan, T.F., Vese, L.A.: Active contours without edges. Image Processing, IEEE Transactions on 10(2), 266–277 (2001)
5. García-Lorenzo, D., Lecoeur, J., Arnold, D., Collins, D., Barillot, C.: Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts. Lecture Notes in Computer Science 5762(PART 2), 584–591 (2009)
6. García-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)
7. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for {MS} lesion segmentation in multi-channel magnetic resonance images. NeuroImage 57(2), 378 – 390 (2011)
8. Larsen, C.T., Iglesias, J.E., Van Leemput, K.: N3 bias field correction explained as a bayesian modeling method. In: Bayesian and grAphical Models for Biomedical Imaging: First International Workshop, BAMBI 2014, Cambridge, MA, USA, September 18, 2014, Revised Selected Papers. pp. 1 – 12 (2014)
9. Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J.C., Quiles, A., Valls, L., Ramió-Torrentà, L., Rovira, A.: Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches. Information Sciences 186(1), 164–185 (2012)
10. Love, J.: Demyelinating diseases. Journal of Clinical pathology 59(11), 1151–1159 (2006)

11. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. Pattern Analysis and Machine Intelligence, IEEE Transactions on 12(7), 629–639 (Jul 1990)
12. Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F.D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A.J., Waubant, E., Weinshenker, B., Wolinsky, J.S.: Diagnostic criteria for multiple sclerosis: 2010 revisions to the mcdonald criteria. Annals of Neurology 69(2), 292–302 (2011)
13. Roura, E., Oliver, A., Cabezas, M., Valverde, S., Pareto, D., Vilanova, J., Ramió-Torrentà , L., Rovira, À., Lladó, X.: A toolbox for multiple sclerosis lesion segmentation. Neuroradiology pp. 1–13 (2015)
14. Roura, E., Oliver, A., Cabezas, M., Vilanova, J.C., Rovira, A., Ramió-Torrentà, L., Lladó, X.: MARGA: Multispectral adaptive region growing algorithm for brain extraction on axial MRI. Computer Methods and Programs in Biomedicine 113(2), 655–673 (2014)
15. Schmidt, P., Gaser, C., Arsic, M., Buck, D., F̈torschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V.J., Zimmer, C., Hemmer, B., Mühlau, M.: An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. NeuroImage 59(4), 3774–3783 (2012)
16. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging 17(1), 87–97 (1998)
17. Souplet, J., Lebrun, C., Ayache, N., Malandain, G.: An automatic segmentation of T2-FLAIR multiple sclerosis lesions. In: Multiple Sclerosis Lesion Segmentation Challenge Workshop (MICCAI-2008). pp. 1–8. New York, NY, USA, United States (2008)
18. Sushmita, D., Ponnada, A.N.: A comprehensive approach to the segmentation of multichannel three-dimensional mr brain images in multiple sclerosis. NeuroImage: Clinical 2, 184–196 (2013)
19. Weickert, J., Romeny, B.M.T.H., Viergever, M.A.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Transactions on Image Processing 7, 398–410 (1998)
20. Weiss, N., Rueckert, D., Rao, A.: Multiple sclerosis lesion segmentation using dictionary learning and sparse coding. In: Medical Image Computing and Computer-Assisted Interventional MICCAI 2013. vol. 8149, pp. 735–742 (2013)

Evaluation-Oriented Training Strategy on MS Segmentation Challenge 2016

Michel M. Santos[1], Paula R. B. Diniz[2], Abel G. Silva-Filho[1], and Wellington P. Santos[3,]

[1] Centro de Informática, Universidade Federal de Pernambuco, Pernambuco, Brazil
[2] Núcleo de Telessaúde, Universidade Federal de Pernambuco, Pernambuco, Brazil
[3] Dpto. de Eng. Biomédica, Universidade Federal de Pernambuco, Pernambuco, Brazil
michelmozinho@gmail.com

Abstract. In most of the current approaches to automatic segmentation of multiple sclerosis (MS) lesions, the segmentation methods are not optimized with respect to all relevant evaluation metrics at once. An obstacle is that the computation of relevant metrics is three-dimensional (3D). The high computational costs of 3D metrics make their use impractical as learning targets for iterative training. In a companion paper, we propose an oriented training strategy that targets cheap 2D metrics as surrogates for costly 3D metrics. We applied the evaluation-oriented training with surrogate learning targets to optimize a simple multilayer perceptron (MLP) network at the core of a segmentation pipeline. In a previous competition, our segmentation strategy achieved a performance that was comparable to state-of-the-art methods. In this paper, by the opportunity of the MS Segmentation Challenge 2016, we apply the proposed method on a larger and more diverse dataset. We expect to compare the proposed strategy to other methods concerning segmentation quality and runtime on the computational cloud provided by the challenge organizers.

1 Introduction

In most methods for multiple sclerosis (MS) lesion segmentation, specific metrics serve as evaluation criteria, but not as optimization targets. Usual evaluation metrics involve lesion count, volume, and shape [1], representing clinical criteria of disease progression [2]. On the other hand, many methods have energy, likelihood, maximum *a posteriori*, or squared error as the optimization target [3–6].

An obstacle to evaluation-oriented training is that evaluation metrics may be too costly for iterative training. The high computational cost is in part due to the three-dimensional computation of evaluation metrics. A shape-based metric, as the average surface distance, may take seconds to be computed in three dimensions [7]. Moreover, the measures based on counts of overlapping lesions between the predicted segmentation and ground truth depend on obtaining contiguous regions. To obtain contiguous lesions, the connected components method

can take seconds in 3D images as well [8]. Suppose for instance that the computation of measures spends 1 minute for each one of 10 patients. In this case, for 1000 iterations of an optimization algorithm, a single training would take almost a week (10,000 minutes). In a companion paper [9], we propose cheaper 2D metrics as surrogates for costly 3D metrics to make the oriented training feasible.

In this paper, we apply the proposed strategy to images of the MS Segmentation Challenge 2016. Compared to a previous challenge [1], the current competition is an opportunity to assess our pipeline on a larger data set involving more sequences, expert segmentations, and scanners.

2 Material and Methods

2.1 MS Segmentation Challenge 2016

Data Set and Preprocessing The MS Lesion Segmentation Challenge 2016 aims to assess automatic methods on a common infrastructure. The dataset is comprised of 53 patients (15 for training, 38 for testing). The data were acquired on 1.5T or 3T scanners, in 4 medical centers. For each patient examination, the dataset contains 7 manual segmentations and 5 magnetic resonance (MR) sequences as follows: 3D FLAIR, 3D T1-w, 3D T1-w GADO, 2D DP, and 2D T2-w. Both raw and preprocessed images are available. The available preprocessed images were subjected to the following steps: denoising with a non-local means algorithm [10], rigid registration of each image to FLAIR image [11], brain extraction via volBrain platform [12], and bias field correction through the N4 algorithm [13].

Rather than using the available preprocessed images, we employed an alternative preprocessing scheme. Our preprocessing routine yielded better values for the objective function in a preliminary evaluation. For the competition, we preprocessed the raw images according to the following steps: non-uniformity correction with N3 [14]; using mni_autoreg tool [15], intra-subject registration to FLAIR image and inter-subject registration to the MNI152 linear model [16]; brain extraction via bet2 tool [17]. The images were smoothed with a Gaussian filter. To make intensities comparable among acquisitions, the foreground intensities were transformed to have median 0 and interquartile range 1.

Evaluation Criteria In order to compute evaluation metrics, the gold standard is defined as the consensus among 7 manual segmentations through the LOP-STAPLE algorithm [18]. Algorithm assessment will occur in three categories: segmentation, lesion detection, and computational performance. The segmentation assessment involves the Dice similarity coefficient and the average surface distance. The evaluation for lesion detection considers region count metrics as the lesion-wise F1-score. Finally, the computational evaluation includes runtime and memory load as metrics.

3 Evaluation-Oriented Training through Surrogate Metrics

3.1 Segmentation Model and Optimization Method

We adopt a simple multilayer perceptron (MLP) model with a single hidden layer containing just a few neurons for fast computation of outputs. We employ particle swarm optimization (PSO) for training MLP as an alternative to backpropagation algorithm [19]. The backpropagation restricts the form of learning target. On the other hand, PSO enables to use different learning targets without being necessary to develop a specific procedure for each function to be optimized.

3.2 Computationally Cheap Learning Targets

The 3D computation of learning targets would be impractical for iterative training. To reduce the computational burden, instead of costly 3D metrics, we employed cheaper 2D surrogates. The 2D surrogates are calculated in lower dimension and averaged over a reduced number of selected slices. Thus, surrogate metrics are faster due to problem reduction, regardless of hardware-based acceleration. For the MS segmentation challenge 2016, we use a combination of three metrics as learning targets: ρ_1: *Dice similarity score*; two lesion count metrics, ρ_2: *lesion-wise true positive rate* and ρ_3: *lesion-wise positive predictive value*. On model training, we do not target surface-based metrics because algorithms will be ranked just according to ρ_1, ρ_2, and ρ_3 in the competition. Thus, the learning target for evaluation-oriented training is the 2D-Score $= (\rho_1 + \rho_2 + \rho_3)/3$. Note that the three component metrics are in the range [0, 1].

3.3 Oriented Objective Function

In evaluation-oriented training, the objective function (Algorithm 1) has as learning target the 2D-score. Moreover, the oriented objective computes an average of three 2D metrics that requires samples of slices for each patient. The objective function includes a pre-classification rule. This rule prevents unnecessary computation of MLP outputs and takes advantage of the fact that lesions appear hyper-intense in FLAIR images (Algorithm 1 at line 5). Notice that such rule requires standardized intensities. Also, note that 2D-score must be maximized, and the PSO must be configured accordingly.

Algorithm 1 Oriented Objective

```
1: function OBJECTIVE(sliceSamples)
2:     for each patient do
3:         for each slice in sliceSamples do
4:             for each voxel do
5:                 if FLAIR intensity < 0 then
6:                     assign non-lesion to voxel
7:                 else
8:                     classify voxel via MLP
9:                 end if
10:            end for
11:            compute 2D metrics
12:        end for
13:    end for
14:    compute average 2D metrics
15:    objective = 2D-score                    ▷ Section (3.2)
16: end function
```

3.4 Slice Selection

The plane, quantity, and position of the selected slices can influence the final 2D-score. We adopted a simple strategy and selected the approximate medial slices 85, 90, and 95 in the axial plane of the MNI152 space.

3.5 Model Training and Usage

The PSO-MLP training was repeated 10 times using a repeated holdout method for model selection. As input, we use FLAIR, T1, and T2 sequences that were preprocessed according to our routine. To apply a trained model to new images, we selected the MLP that yielded the best objective function value for all entries in the training set. We developed a segmentation pipeline comprised by preprocessing, segmentation (with the selected MLP) and postprocessing. In the postprocessing stage, we employed a lesion probability atlas to remove false-positives occurring in low probability regions. This lesion atlas was built from the MS challenge 2008 data set.

4 Conclusion

We described evaluation-oriented training with surrogate learning targets to train an MLP inside an MS segmentation pipeline. We report the preprocessing steps employed. The pipeline was applied to the images of the MS Segmentation Challenge 2016. In comparison to previous MS segmentation competitions, the 2016 challenge provides a larger and more diverse dataset, besides a single platform to assess the computational performance of different segmentation methods.

References

1. Styner, M., Lee, J., Chin, B., Chin, M.S., huong Tran, H., Jewells, V., Warfield, S.: 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. In: MICCAI 2008 Workshop. pp. 1–5 (2008)
2. Barkhof, F., Filippi, M., Miller, D.H., Scheltens, P., Campi, A., Polman, C.H., Comi, G., Ader, H.J., Losseff, N., Valk, J.: Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. Brain 120(11), 2059–2069 (1997)
3. Souplet, J.C., Lebrun, C., Ayache, N., Malandain, G., et al.: An automatic segmentation of T2-FLAIR multiple sclerosis lesions. In: The MIDAS Journal-MS Lesion Segmentation (MICCAI 2008 Workshop) (2008)
4. García-Lorenzo, D., Prima, S., Arnold, D.L., Collins, D.L., Barillot, C.: Trimmed-likelihood estimation for focal lesions and tissue segmentation in multisequence MRI for multiple sclerosis. Medical Imaging, IEEE Transactions on 30(8), 1455–1467 (2011)
5. Tomas-Fernandez, X., Warfield, S.: A model of population and subject (MOPS) intensities with application to multiple sclerosis lesion segmentation. Medical Imaging, IEEE Transactions on 34(6), 1349–1361 (June 2015)
6. Jesson, A., Arbel, T.: Hierarchical MRF and random forest segmentation of MS lesions and healthy tissues in brain MRI. In: The Longitudinal MS Lesion Segmentation Challenge (2015)
7. Taha, A.A., Hanbury, A.: Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC medical imaging 15(1), 29 (2015)
8. He, L., Chao, Y., Suzuki, K.: Two efficient label-equivalence-based connected-component labeling algorithms for 3-D binary images. Image Processing, IEEE Transactions on 20(8), 2122–2134 (2011)
9. Santos, M.M., Diniz, P.R., Silva-Filho, A.G., Santos, W.P.: Evaluation-oriented training via surrogate metrics for multiple sclerosis segmentation. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016 (to appear)
10. Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. IEEE transactions on medical imaging 27(4), 425–441 (2008)
11. Commowick, O., Wiest-Daesslé, N., Prima, S.: Block-matching strategies for rigid registration of multimodal medical images. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). pp. 700–703. IEEE (2012)
12. Manjon, J.V., Coupé, P.: volbrain: An online mri brain volumetry system. In: Organization for Human Brain Mapping'15 (2015)
13. 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)
14. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in mri data. IEEE transactions on medical imaging 17(1), 87–97 (1998)
15. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3d intersubject registration of mr volumetric data in standardized talairach space. Journal of computer assisted tomography 18(2), 192–205 (1994)
16. Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., et al.: A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). Philosophical

Transactions of the Royal Society of London B: Biological Sciences 356(1412), 1293–1322 (2001)
17. Jenkinson, M., Pechaud, M., Smith, S.: Bet2: Mr-based estimation of brain, skull and scalp surfaces. In: Eleventh annual meeting of the organization for human brain mapping. vol. 17, p. 167. Toronto, ON (2005)
18. Akhondi-Asl, A., Hoyte, L., Lockhart, M.E., Warfield, S.K.: A logarithmic opinion pool based staple algorithm for the fusion of segmentations with associated reliability weights. IEEE transactions on medical imaging 33(10), 1997–2009 (2014)
19. Eberhart, R.C., Shi, Y.: Computational Intelligence - Concepts to Implementations. Elsevier (2007)

MRI Robust Brain Tissue Segmentation with application to Multiple Sclerosis

Xavier Tomas-Fernandez[1] and Simon K. Warfield[1]

Children's Hospital Boston, Boston MA 02115 , USA,
Xavier.Tomas-Fernandez@childrens.harvard.edu,
http://crl.med.harvard.edu/

Abstract. We present a fully automated algorithm for the segmentation of the brain tissues (i.e. gray matter, white matter, cerebrospinal fluid and white matter hyperintense lesions) from multispectral magnetic resonance (MR) images of multiple sclerosis (MS) patients. The method performs intensity based tissue classification using a mixture model for normal brain tissues as well as MS lesions. We propose to initialize the model for MS lesions through the simultaneous estimation of a spatially varying within-the-subject intensity distribution and a spatially local intensity distribution derived from a healthy reference population.

1 Introduction

In order to provide objective assessments of segmentation performance, there is a need for an objective reference standard with associated MRI scans that exhibit the same major segmentation challenges as that of scans of patients. A database of clinical MR images, along with their segmentations, may provide the means to measure the performance of an algorithm by comparing the results against the variability of the expert segmentations. However, an objective evaluation to systematically compare different segmentation algorithms also needs an accurate reference standard.

An example of such a reference standard is the synthetic brain MRI database provided by the Montreal Neurological Institute that is a common standard for evaluating the segmentations of MS patients. The synthetic MS brain phantom available from the McConnell Brain Imaging Centre consists of T1w, T2w, and proton density MRI sequences with different acquisition parameters as well as noise and intensity inhomogeneity levels [2]. The MS brain phantom was based on the original BrainWeb healthy phantom, which had been expanded to capture three different MS lesion loads: mild (0.4 cm3), moderate (3.5 cm3), and severe (10.1 cm3). Each MS phantom was provided with its own MS lesion ground truth.

Although the BrainWeb synthetic dataset provides a reference standard, it presents several limitations. First, the BrainWeb dataset just provides one brain model, which results in a poor characterization of the anatomical variability present in the MS population. Also, although the BrainWeb dataset is based on

real MRI data, the final model is not equivalent to clinical scans in its contrast, and it produces an easier dataset to segment than real clinical scans.

To overcome these limitations, most of the lesion segmentation algorithms also evaluate their results in a dataset consisting in clinical scans. Such an approach allows for a better understanding of the performance of the evaluated algorithms when faced with real data. Unfortunately, because each segmentation algorithm is validated with different datasets, comparison between different methodologies is more difficult.

As part of the ongoing effort in providing publicly available datasets for validation of MS lesion segmentation, a new MS segmentation challenge was held during the 2016 MICCAI conference. The goals of this challenge are multiple. The first aim was to evaluate state-of-the-art advanced segmentation methods from the participants on a database following a standard protocol [1]. For this, both lesion detection (how many lesions were detected) and lesion segmentation (how precise the lesions are delineated) will be evaluated on a multi-centric database (38 patients from four different centers, imaged on 1.5 or 3T scanners, each patient being manually annotated by seven experts).

In addition to this classical evaluation, the goal of this challenge was also to provide a common infrastructure on which the algorithms will be evaluated. This infrastructure will enable a fair comparison of the algorithms in terms of running time comparison and ensuring all algorithms will be run with the same parameters for each patient (which is required for a truly automatic segmentation).

2 Method

Consider a multispectral grayscale MRI (i.e. T1w, T2w and FLAIR) formed by a finite set of N voxels. Our aim is to assign each voxel i to one of K classes (i.e. GM, WM, CSF and T2w hyperintense lesions) considering the observed intensities $\mathbf{Y} = \{\mathbf{y_1}, ..., \mathbf{y_N}\}$ with $\mathbf{y_i} \in \mathbb{R}_m$. Both observed intensities and hidden labels are considered to be random variables denoted respectively as $\mathbf{Y} = \{\mathbf{Y_1}, ..., \mathbf{Y_N}\}$ and $\mathbf{Z} = \{\mathbf{Z_1}, ..., \mathbf{Z_N}\}$. Each random variable $\mathbf{Z_i} = \mathbf{e}_k = \{z_{i1}, ..., z_{iK}\}$ is a K-dimensional vector with each component z_{ik} being 1 or 0 according whether $\mathbf{Y_i}$ did or did not arise from the k^{th} class.

It is assumed that the observed data $\mathbf{Y}$ is described by the conditional probability density function $f(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\phi_Y})$, which incorporates the image formation and noise models and depends on some parameters $\boldsymbol{\phi_Y}$. In addition, the hidden labels are assumed to be drawn according to a certain parametric probability distribution $f(\mathbf{Z}|\boldsymbol{\phi_Z})$, which depends on parameters $\boldsymbol{\phi_Z}$.

If the underlying tissue segmentation $\mathbf{Z}$ was known, estimation of the model parameters would be straightforward. However, only the image $\mathbf{Y}$ is directly observed, making it necessary to tackle this problem as one involving missing data. The Expectation-Maximization (EM) algorithm is the best candidate for model fitting. The EM algorithm finds the parameters that maximize the complete data log-likelihood by iteratively maximizing the expected value of the log-likelihood

$log(f(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\psi}))$ of the complete data $\{\mathbf{Y}, \mathbf{Z}\}$, where the expectation is based on the observed data $\mathbf{Y}$ and the estimated parameters $\boldsymbol{\psi}^{(m)}$ obtained in the previous iteration m [8].

A popular strategy to initialize this iterative process, is to use a spatially varying probability map for the different tissue classes [6]. However, because the random distribution and burden of MS lesions, prior expectations about the spatial location of tissue classes cannot be derived from a reference population of MS patients. Instead, we propose to derive a patient specific model for MS lesions through the simultaneous estimation of a spatially varying within-the-subject intensity distribution and a spatially (MOPS) local intensity distribution derived from a healthy reference population [5].

2.1 Model of Population and Subject (MOPS)

In addition to the patient scan $\mathbf{Y}$, we observe an aligned reference of R healthy subjects $\mathbf{P} = \{\mathbf{V}, \mathbf{L}\} = \{\mathbf{V}_1, ..., \mathbf{V}_R; \mathbf{L}_1, ..., \mathbf{L}_R\}$.

Segmenting the observed data $\mathbf{Y}$ implies the estimation of parameters $\boldsymbol{\psi}$.

$$\begin{aligned}
&\log L_C(\boldsymbol{\psi}) = \log\left(f\left(\mathbf{Y}, \mathbf{P}, \mathbf{Z}|\boldsymbol{\psi}\right)\right) \\
&= \log\left(\prod_{i=1}^{N}\sum_{k=1}^{K} f(\mathbf{Z}_{ik}|\boldsymbol{\psi}_k) f(\mathbf{Y}_i, \mathbf{P}_{ik}|\mathbf{Z}_{ik}, \boldsymbol{\psi}_k)\right) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log\left(\pi_k f(\mathbf{Y}_i|\mathbf{Z}_{ik}, \boldsymbol{\psi}_k) f(\mathbf{P}_{ik}|\mathbf{Y}_i, \mathbf{Z}_{ik}, \boldsymbol{\psi}_k)\right)
\end{aligned}$$

The observed population data $\mathbf{P}$ is conditionlly independent of the observed patient scan $\mathbf{Y}$ formation model parametrized by $\boldsymbol{\psi}$.

$$\begin{aligned}
&\log L_C(\boldsymbol{\psi}) = \log\left(f(\mathbf{Y}, \mathbf{P}, \mathbf{Z}|\boldsymbol{\psi})\right) \\
&\sim \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik} \log\left(\pi_k f(\mathbf{Y}_i|\mathbf{Z}_{ik}, \boldsymbol{\psi}_k) f(\mathbf{Z}_{ik}|\mathbf{P}_{ik}) f(\mathbf{Y}_i|\mathbf{P}_{ik})\right) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} z_i k \log\left(\pi_k \pi_{Pik} \mathcal{N}(\mathbf{Y}_i|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}) \mathcal{N}(\mathbf{Y}_i|\boldsymbol{\mu}_{Pk}, \boldsymbol{\Sigma}_{Pk})\right) \qquad (1)
\end{aligned}$$

Since the underlying tissue segmentation $\mathbf{Z}$ is unknown, the EM algorithm will be used to find the parameters that maximize the complete log-likelihood.

E-Step: The E-step requires the computation of the conditional expectation of $\log\left(L_C(\boldsymbol{\psi})\right)$ given $\mathbf{Y}$ and $\mathbf{P}$, using the current parameter estimate $\boldsymbol{\psi}^{(m)}$.

$$Q(\boldsymbol{\psi}; \boldsymbol{\psi}^{(m)}) = E_{\boldsymbol{\psi}^{(m)}}\{\log L_C(\boldsymbol{\psi})|\mathbf{Y}, \mathbf{P}\}$$

The complete log-likelihood is linear in the hidden labels z_{ij}. Again, the E-step requires the calculation of the current conditional expectation of $\mathbf{Z}i$ given the observed data $\mathbf{Y}$:

$$\begin{aligned}
&E_{\boldsymbol{\psi}^{(m)}}(\mathbf{Z}_i = \mathbf{e}_k|\mathbf{Y}, \mathbf{P}) = f(\mathbf{Z}_i = \mathbf{e}_j|\mathbf{Y}_i, \mathbf{P}_i, \boldsymbol{\psi}_{\boldsymbol{j}}^{(\boldsymbol{m})}) \\
&= \frac{\pi_k \pi_{Pik} \mathcal{N}(\mathbf{Y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \mathcal{N}(\mathbf{Y}_i|\boldsymbol{\mu}_{Pk}, \boldsymbol{\Sigma}_{Pk})}{\sum_{k'=1}^{K} \pi'_k \pi_{Pik'} \mathcal{N}(\mathbf{Y}_i|\boldsymbol{\mu}'_k, \boldsymbol{\Sigma}'_k) \mathcal{N}(\mathbf{Y}_i|\boldsymbol{\mu}_{Pk'}, \boldsymbol{\Sigma}_{Pk'})} \qquad (2)
\end{aligned}$$

M-Step: Because the local reference population model parameter $\boldsymbol{\xi}$ is constant, the Maximization step will consist of the maximization of $Q(\boldsymbol{\psi}; \boldsymbol{\psi}^{(m)})$ with respect to $\boldsymbol{\psi}$, which results in:

$$f(\mathbf{Z}_i = \mathbf{e}_k | \boldsymbol{\psi}_k^{(m+1)}) = \frac{1}{N} \sum_{i=1}^{N} f\left(\mathbf{Z}_i = \mathbf{e}_k | \mathbf{Y}_i, \mathbf{P}_{ik}, \boldsymbol{\psi}^{(m)}\right) \tag{3}$$

$$\boldsymbol{\mu}_k^{(m+1)} = \frac{\sum_{i=1}^{N} \mathbf{Y}_i f(\mathbf{Z}_i = \mathbf{e}_k | \mathbf{Y}, \mathbf{P}_{ik}, \boldsymbol{\psi}^{(m)})}{\sum_{i=1}^{N} f(\mathbf{Z}_i = \mathbf{e}_k | \mathbf{Y}, \mathbf{P}_{ik}, \boldsymbol{\psi}^{(m)})} \tag{4}$$

$$\boldsymbol{\Sigma}_k^{(m+1)} = \frac{\sum_{i=1}^{N} f(\mathbf{Z}_i = \mathbf{e}_k | \mathbf{Y}, \mathbf{P}_{ik}, \boldsymbol{\psi}^{(m)}) \left(\mathbf{Y}_i - \boldsymbol{\mu}_k\right)^T \left(\mathbf{Y}_i - \boldsymbol{\mu}_k\right)}{\sum_{i=1}^{N} f(\mathbf{Z}_i = \mathbf{e}_k | \mathbf{Y}, \mathbf{P}_{ik}, \boldsymbol{\psi}^{(m)})} \tag{5}$$

It follows intuitively that the local intensity model downweights the likelihood of those voxels having an abnormal intensity given the reference population. Since MRI structural abnormalities will show an abnormal intensity level compared to similarly located brain tissues in healthy subjects, we seek to identify MS lesions by searching for areas with low likelihood $L_C(\boldsymbol{\psi})$.

We define a voxels lesion probability as $1 - f(Y_i, P_i, Z_i | \boldsymbol{\psi})$. We initialize a subject specific MS lesion model by selecting those voxels with a lesion probability greater than a given decision threshold $(1 - f(Y_i, P_i, Z_i | \boldsymbol{\psi}) > 0.5)$.

3 Implementation Details

This section explains the implementation and distribution details of the proposed brain tissue segmentation algorithm. First, we describe the reference template library of healthy subjects which the MOPS algorithm uses to initialize the local intensity distribution. Following, we outline the implementation details of the segmentation pipeline.

3.1 Reference Population

We collected data from 15 volunteers on a 3T clinical MR scanner from GE Medical Systems (Waukesha, WI, USA) using an 8-channel receiver head coil and three different pulse sequences: a T1-weighted MPRAGE (Magnetization Prepared Rapid Acquisition Gradient Echo) sequence; a T2-weighted scan from an FSE (Fast Spin Echo) sequence; and a FLAIR scan, also run with an FSE sequence. We acquired the T1w sequence axially; the T2w and FLAIR sequences were sagitally acquired. All sequences were acquired with a matrix size of 256x256 and a field of view of 28 cm. Slice thickness was 1.3 mm for the T1w-MPRAGE sequence; 1 mm for the T2w-FSE sequence; and 2 mm for the FLAIR-FSE sequence. The MPRAGE parameters were TR 10/TE 6 ms with a flip angle of 8. For the FSE, the paramenters were TR 3000/TE 140 ms with an echo train length of 15.

After image acquisition, we aligned the T2w and FLAIR images to the T1w scan. Last, a trained expert manually segmented the intra-cranial volume, CSF, GM and WM tissues [3].

3.2 Segmentation Pipeline Implementation

1. Inputs
   - RAW T1w scan.
   - RAW T2w scan.
   - RAW FLAIR scan.
2. Pre-processing
   - Intra-subject rigid registration: The T2w and FLAIR scans were rigidly aligned to the reference space of the T1w scan.
   - Intra cranial cavity segmentation: To exclude non-brain tissues from the posterior segmentation steps, an intra-cranial mask was estimated based using the T1w and T2w scans based in the strategy described in [7].
   - Intensity inhomogeneity artifact correction [8].
   - Healthy Reference Population registration: To achieve accurate alignment between healthy volunteers and a patient with MS, we used a block-matching nonlinear registration algorithm proposed by [4], which, although not intrinsic to our method, was selected because it is robust in the presence of WM lesions.
3. MOPS Brain Tissue Segmentation
   - Spatially varying estimator sliding window radius was set to 2 voxels.
   - To initialize the subject specific MS lesion model, we thresholded the MOPS Lesion probability map to keep those voxels with a probability higher than 0.5.

References

1. Cotton, F., Kremer, S., Hannoun, S., Vukusic, S., Dousset, V., Imaging Working Group of the Observatoire Français de la Sclérose en Plaques: OFSEP, a nationwide cohort of people with multiple sclerosis: Consensus minimal MRI protocol. J. Neuroradiol. 42(3), 133–40 (jun 2015), http://www.ncbi.nlm.nih.gov/pubmed/25660217
2. Kwan, R.K., Evans, A.C., Pike, G.B.: MRI simulation-based evaluation of image-processing and classification methods. IEEE Trans. Med. Imaging 18(11), 1085–97 (nov 1999)
3. Makris, N., Meyer, J.W., Bates, J.F., Yeterian, E.H., Kennedy, D.N., Caviness, V.S.: MRI-Based topographic parcellation of human cerebral white matter and nuclei II. Rationale and applications with systematics of cerebral connectivity. Neuroimage 9(1), 18–45 (jan 1999)
4. Suarez, R.O., Commowick, O., Prabhu, S.P., Warfield, S.K.: Automated delineation of white matter fiber tracts with a multiple region-of-interest approach. Neuroimage 59(4), 3690–700 (feb 2012)
5. Tomas-Fernandez, X., Warfield, S.K.: A Model of Population and Subject (MOPS) Intensities With Application to Multiple Sclerosis Lesion Segmentation. IEEE Trans. Med. Imaging 34(6), 1349–1361 (jun 2015)
6. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Trans. Med. Imaging 18(10), 897–908 (oct 1999)
7. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain MRI. Neuroimage 47(2), 564–72 (aug 2009)
8. Wells, W.M., Grimson, W.L., Kikinis, R., Jolesz, F.a.: Adaptive segmentation of MRI data. IEEE Trans. Med. Imaging 15(4), 429–42 (jan 1996)

A 3D hierarchical multimodal detection and segmentation method for multiple sclerosis lesions in MRI

Hélène Urien[⋆†], Irène Buvat[⋆], Nicolas Rougon[†], and Isabelle Bloch[⋆†]

[⋆†] LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France
[⋆] IMIV, Inserm, CEA, Université Paris-Sud, CNRS, Université Paris-Saclay, 91400, Orsay
[†] MAP5, CNRS, Télécom SudParis, Université Paris-Saclay, 91011, Evry, France

Abstract. In this paper, we propose a novel 3D method for multiple sclerosis segmentation on FLAIR Magnetic Resonance images (MRI), based on a lesion context-based criterion performed on a max-tree representation. The detection criterion is refined using prior information from other available MRI acquisitions (T2, T1, T1 enhanced with Gadolinium and DP). The method has been tested on fifteen patients suffering from multiple sclerosis. The results show the ability of the method to detect almost all lesions. However, the algorithm also provides false detections.

Keywords: Max-tree, multiple sclerosis, MRI multimodal segmentation.

1 Introduction

In this paper, we present a new method to detect and segment lesions from MRI images of patients suffering from multiple sclerosis. The method is based on a hierarchical approach, where the image is represented by a tree depicting the relationship between all the connected components of all the thresholded versions of the image. The nodes of the tree represent the potential lesions, and are selected only if they satisfy a criterion based on the difference between the lesion and the surrounding intensities. Our method was tested on the fifteen training datasets given by the MICCAI MSSEG challenge, providing different MRI acquisitions for each patient (FLAIR, T2, T1, T1 enhanced with Gadolinium and DP). The detection is first performed on FLAIR, and then refined using the other available modalities. Parameter setting is given and results are discussed.

2 A 3D multiple sclerosis lesion detection method using a hierarchical approach

The lesion detection is based on the hypothesis that the FLAIR intensity of a lesion is higher in a lesion than in the surrounding region. This assumption

was verified on the lesions extracted from the manual consensus segmentations performed using the LOP STAPLE algorithm [1], considering the average FLAIR intensity in the segmented region and in a neighborhood. The detection relies on a max-tree representation of the image [2], which highlights regions of high intensities. This method, based on the study of inclusion relationships between connected components obtained by thresholding an image at all its intensity values, has several advantages. It allows searching for potential lesions in the whole image, without any prior information about the number or localizations of lesions. Connected components are selected according to a given criterion. An example is depicted in Figure 1, where homogeneous regions, called flat zones, are illustrated by a letter identifiying a specific region, and a number corresponding to the signal intensity in each voxel of the region. The tree is oriented from the root corresponding to the flat zone of lowest intensity, to the leaves, which correspond to the flat zones of highest intensities in the image. A node N is said to be a descendant of a node M if it is higher in the tree. As an example, D is a descendant of C in Figure 1.

(a) (b)

Fig. 1. Example of a synthetic image (a) and its corresponding max-tree (b). The root node is depicted in red, while the leaves are represented in green. Regions are identified by a letter and a number corresponding to the grey level in the region.

The first step is the selection of relevant nodes according to a given criterion, composed of two terms and applied to each current node N:

$$\chi_I(N) = X1_I(N) - X2(N) \tag{1}$$

The first term $X1_I$ follows our hypothesis of difference in intensities between the lesion and its surroundings:

$$X1_I(N) = \frac{\frac{\sum_{x \in \widehat{N}} I(x)}{n_{\widehat{N}}}}{\frac{\sum_{x \in C_{\widehat{N}}} I(x)}{n_{C_{\widehat{N}}}}} \tag{2}$$

where $I(x) = I_{\text{FLAIR}}(x)$ is the pre-processed FLAIR intensity in a voxel x, $n_{\widehat{N}}$ the number of voxels within the region $\widehat{N}$, representing the current node N and all its descendants, and $n_{C_{\widehat{N}}}$ the number of voxels within a contextual volume surrounding node N and all its descendants. The maximal Euclidian distance

from a voxel of the contextual volume to the nearest voxel to $\widehat{N}$ is set to α voxels. If a node N is representative of a lesion, the criterion value $X_I(N)$ is superior to 1.

The second term $X2$ depends on the location of current node N compared to the z axis. If the node is located in slices of the FLAIR image containing ventricles, we assume that the lesion tends to be near the ventricles:

$$X2(N) = \frac{\min\limits_{x \in \widehat{N}, v \in V} d(x, v)}{\max\limits_{b \in B, v \in V} d(b, v)} \tag{3}$$

where x is a voxel from the volume $\widehat{N}$ composed of the current node N and all its descendants, v a voxel from the binary mask of the ventricles V, d the Euclidian distance and b a voxel from the binary mask of the brain provided in the unprocessed dataset B. Otsu's thresholding [3] is applied on the pre-processed T1 image to segment the ventricles.

If the node is not located in slices containing ventricles, we penalize small nodes:

$$X2(N) = \frac{1}{n_{\widehat{N}}} \tag{4}$$

where $n_{\widehat{N}}$ is the number of voxels within region $\widehat{N}$, representing the current node N and all its descendants.

The last step is the reconstruction of the image. Only the nodes with criterion value χ superior to 1 are taken into account. The first node to be selected is the one with the highest criterion value. All its descendants and ancestors with criterion value superior to 1 are removed from the searching list. The remaining node with the highest criterion value superior to 1 is then selected. This process is repeated until there is no node with criterion value superior to 1 in the searching list. The final image is the union of all the selected nodes filled with their descendants.

To improve the detection performance, additional constraints are used, depending on the location of the lesions with respect to brain structures, and with adaptative ranges of admissible χ values. These ranges are defined in a supervised way. The position of a lesion compared to brain structures is assigned to one of three categories: near the ventricles, inside the white matter, and near the cortex. Otsu's thresholding [3] is applied on the pre-processed FLAIR image to segment the white matter. Using an automatic thresholding is enough to roughly localize the structures of interest. The manually segmented lesions were also assigned to these three regions, according to their position. For each region, and for each available modality I, the maximal and minimal values of the criterion χ_I were computed, as illustrated in Table 1. Moreover, the maximal and minimal values of the volume of each lesion were computed according to their position, as illustrated in Table 2. Finally, a node N is selected only if the value of the criterion χ_I computed for all the modalities I and depending on the position of

the node with respect to the brain structures falls within the intervals defined in Tables 1 and 2. If so, the criterion χ_{FLAIR} is assigned to this node N, which is taken into account in the image reconstruction step only if its value is superior to 1.

3 Results and discussion

The algorithm uses images from both pre-processed (FLAIR, T1 enhanced with Gadolinium, T2, T1, DP) and unprocessed (FLAIR and brain binary mask) datasets.

3.1 Parameter setting

In order to limit the size of the tree, while preserving contour information, the input image is first quantified on k levels. Here we used $k = 20$ for FLAIR images. Then the tree is created using Salembier's algorithm [2] on the quantized FLAIR image.

The maximal Euclidian distance from a voxel of the contextual volume to the nearest voxel to node $\widehat{N}$ filled with all its descendants is set to $\alpha = 1$ voxel.

The connectivity was set to 26 to compute the max-tree. Tables 1 and 2 show the minimal and maximal values of the criterion and of the volumes of manually segmented lesions according to the position of the lesion.

	Lesion near the ventricles		Lesion inside the white matter		Other lesion	
	Minimal value	Maximal value	Minimal value	Maximal value	Minimal value	Maximal value
Original FLAIR	1.00	1.50	1.00	1.50	1.00	1.50
Pre-processed FLAIR	1.00	1.50	1.00	1.50	1.00	1.50
Pre-processed T2	0.90	2.00	0.90	2.00	1.00	2.00
Pre-processed T1	0.50	1.50	0.80	1.50	0.70	1.00
Pre-processed T1 Gado	0.50	1.50	0.80	1.50	0.70	1.00
Pre-processed DP	0.90	1.50	0.90	1.50	1.00	1.50

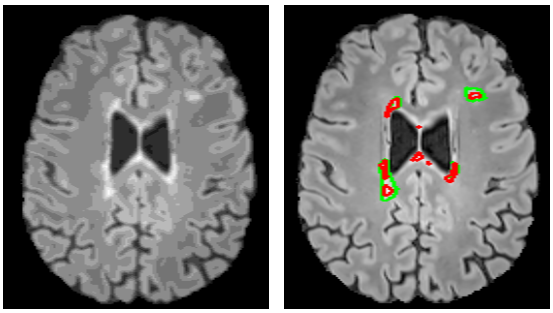
Table 1. Ranges of values for criterion χ for a node, depending on the modality and the position of the node.

	Lesions near the ventricles		Lesions inside the white matter		Other lesions	
	Minimal value	Maximal value	Minimal value	Maximal value	Minimal value	Maximal value
Volume (mm^3)	3.85	$6.44\ 10^4$	3.02	$2.28\ 10^3$	9.07	807

Table 2. Ranges of values of node volume (mm^3) depending on the position of the node.

3.2 Preliminary results

A preliminary result is shown for one of the patients in Figure 2. The algorithm manages to detect almost all lesions, but leads to an over-detection in areas of the brain which do not contain ventricles.



(a) (b)

Fig. 2. Visual results. (a) Pre-processed FLAIR image requantized to 20 grey levels. (b) Contours of manual (in green) and automatic (in red) segmentations.

4 Conclusion

We proposed a 3D method for multiple sclerosis detection embedded in a hierarchical approach. The results show that the lesions from the manual segmentation are well included in the segmentation, but at the price of a number of false detections. The method could be improved using other contextual criteria, such as the one from [4], based on the difference of variance between the object of interest and its surroundings.

References

1. Akhondi-Asl, A., and Hoyte, L., and Lockhart, M. E., and Warfield, S.K. A Logarithmic Opinion Pool Base STAPLE Algorithm for the Fusion of Segmentations With Associated Reliability Weights. IEEE Transactions on Medical Imaging, 33(10):1997–2009 (2014).
2. Salembier, P., and Oliveras, A., and Garrido, L. Antiextensive connected operators for image and sequence processing. IEEE Transactions on Image Processing. 7, 555–570 (1998).
3. Otsu, N.: A threshold selection method from gray-level histograms. Automatica, 11, 285–296 (1975).
4. Xu, Y., and Géraud, T., and Najman, L. Context-based energy estimator: Application to object segmentation on the tree of shapes. 19th IEEE International Conference on Image Processing (ICIP).1577-1580 (2012).

Multiple sclerosis lesion detection and segmentation using a convolutional neural network of 3D patches

Sergi Valverde, Mariano Cabezas, Eloy Roura, Sandra González-Villà, Joaquim Salvi, Arnau Oliver, and Xavier Lladó

Department of Computer Architecture and Technology, University of Girona. Spain.
{svalverde,mcabezas,eroura,sgonzalez,qsalvi,aoliver,llado}@atc.udg.edu

Abstract. Automatic multiple sclerosis (MS) lesion segmentation in magnetic resonance (MR) imaging is a challenging task due to the small size of the lesions, its heterogeneous shape and distribution, overlapping tissue intensity distributions, and the inherent artifacts of MR images. In this paper we propose a convolutional neural network trained with 3D patches of candidate lesion voxels. The method uses 4 anatomical MR images: T1-weigthed, T2-weighted, PD-weighted and T2-FLAIR-weighted.

1 Introduction

Magnetic resonance (MR) imaging of the brain has been widely used during the last years in clinical practice. This image modality presents a high contrast for soft tissues, including white matter lesions (WML). Expert tracing of these lesions is a time-consuming task prone to observer errors. On the other hand, intensity inhomogeneities and image artifacts render difficult the task of obtaining an automatic and reliable segmentation of these lesions based only on intensity features.

Common supervised approaches rely on the use of a classification algorithm. These algorithms involve a first stage in which a model is estimated on training data composed by a set of features and their corresponding ground truth, and a second stage in which the model is tested on a new dataset to provide the desired classification. These features can include information from an atlas, context, spatial coordinates or even texture. However, classic machine learning methods require hand-crafting these feature vectors to extract appearance information using, for instance, Gaussian or Haar-like kernels. In contrast, convolutional neural networks (CNNs) learn sets of convolution kernels that are specifically created for the task at hand.

Currently, CNNs have demonstrated a superior performance in several computer vision tasks including handwriting recognition [10], classification of 2D images in 1000 classes [9], segmentation of crowds in surveillance videos [7] or the application of a painting's style to other pictures [5]. Recently, CNNs have also gained popularity in medical imaging in general [2,6], and brain imaging specifically [12,14].

Different architectures have been published in the literature to tackle the abovementioned problems. For instance, Zhang et al [14] proposed a deep convolutional neural network for segmenting isointense brain tissues using multi-modality MR images. Their multiple intermediate layers applied convolution, pooling, normalization and other operations on 2D patches to capture highly nonlinear mappings between the inputs and the outputs. Moreover, Moeskops et al [12] also presented a CNN architecture based on 2D patches of a single anatomical MR image. However, in their work they used different patch sizes to which they applied different convolutional layers and average pooling that were finally combined using a fully connected layer with softmax to obtain a 9 class segmentation including background. On the other hand, Brosch et al [1] defined a multiscale fully convolutional encoder network with shortcuts to segment lesions using the whole brain image. However, as pointed out in their work, this kind of network requires a large number of cases in order to train a deep network of more than 1 layer for the convolutional and deconvolutional pathways.

In this paper we present a 3D CNN that uses 3D candidate voxel patches to train an architecture that incorporates convolutional layers, max pooling and dense layers to obtain the probability for each candidate voxel of being lesion. This map is then post-processed to obtain a final lesion mask.

2 Methods

2.1 Preprocessing

We decided to use the pre-processed dataset to focus exclusively on the CNNs implementation. This dataset has been denoised with the NL-means algorithm [4] and rigidly registered [3] of each image towards the FLAIR image. Moreover, the skull has been stripped using the volBrain platform [11] on the T1 image and applied on the other modalities with sinc interpolation, and, finally, bias correction was applied using the N4 algorithm [13].

In order to train our CNN architecture, we also normalised the intensities for each image using the mean and standard deviation of the brain voxels.

2.2 CNN architecture

Most of the CNNs from the literature use 2D images or patches to segment tissues or lesions. However, when using MR images, such approaches are prone to false positives on some slices, due to the similarity between lesions and artifacts in some slices. By using 3D patches we can discard those false positives that are clearly not lesions when analysed in 3D. Furthermore, we decided to use patches instead of the whole image as input to obtain a higher number of training samples (positive and negative) while also reducing the amount of parameters to optimise as network weights. For each candidate voxel, we define a patch of size $15 \times 15 \times 15$ for each image T1,T2,PD,FLAIR. Therefore, our input vector has a size of $N \times 4 \times 15 \times 15 \times 15$, where N is the number of training patches.

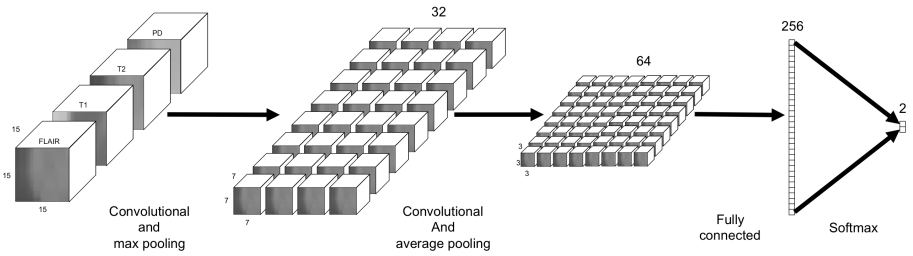


Fig. 1. Architecture of the convolutional network taking candidate patches of size $15 \times 15 \times 15$ as inputs.

With this input vector, we designed our CNN architecture detailed in Figure 1. The first convolutional layer contained 32 filters of size $5 \times 5 \times 5$, followed by a max pooling of size 2 with stride 2. The following convolutional layer had 64 filters of the same size, also followed by a max pooling with the previous parameters. Afterwards we applied a dropout on probabilities lower than 0.5 to reduce overfitting, and we finished the architecture with 2 dense layers. The first one had 256 outputs and the last one was a 2-way softmax to obtain the probabilities for the 2 possible classes (lesion and not lesion).

2.3 Training

Since the dataset is unbalanced, we have a larger number of voxels that belong to normal appearing tissues than lesion voxels, we decided to apply an iterative process during training. This process had two main steps that trained the same CNN architecture with different data.

During the initial step, we selected a random number of negative voxels. First, we applied an empirical threshold of 1.5 (deviations) on the normalised FLAIR image to obtain all hyperintense candidate voxels. From this set of voxels, we used all the voxels defined as lesion in the consensus and a random sampling of the same size of all the candidates that are not lesion, in order to balance the dataset. Since this initial selection is suboptimal, some of the tissue voxels from each image have a high probability of belonging to lesion and would be classified as false positives. Therefore, we trained again our network with a new subset of negative voxels to better classify these voxels.

During the second and final step, we tested the training images with the first network to obtain a probabilistic map. In order to select challenging false positive, we applied a threshold of 0.5 to this map and we randomly selected a sample of the negative voxels inside this mask of the same size as the lesion voxel dataset. Finally, we trained again the same CNN architecture with these new training set.

Since both networks use the same positive voxels, when testing both networks, all lesions are correctly classified. However, since we used different negative voxels, the probabilistic maps present different false positive detections that

do not overlap. Therefore, we decided to multiply the output of both networks to maximise the true positives and minimise the false positives.

2.4 CNN parameter tuning

Our CNN architecture was developed in Python using the nolearn[1] and Lasagne[2] modules for Theano[3]. The batch size of the net for training was set to 4096 and the maximum number of epochs was set to 50 for the initial iteration and 2000 for the final iteration (even though it automatically stops if there is no improvement after 50 iterations). To update the weights we used the ADAM learning algorithm [8].

In order to evaluate the performance of the net, we used the training dataset with a leave-one-out strategy. Since we had more than million of patches per training, we decided to use a Tesla K40 GPU device with CUDA and cuDNN in a server with 56 cores and 256GB of RAM memory to accelerate the process.

3 Discussion

We presented a multichannel approach based on deep learning of 3D patches from candidate lesion voxels previously segmented in T2-FLAIR-weighted images.

Acknowledgments. This work has been partially supported by "La Fundació la Marató de TV3", by Retos de Investigación TIN2014-55710-R, and by MPC UdG 2016/022 grant. The authors gratefully acknowledge the support of the NVIDIA Corporation with their donation of the Tesla K40 GPU used in this research.

References

1. Brosch, T., Tang, L.Y.W., Yoo, Y., Li, D.K.B., Traboulsee, A., Tam, R.: Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. IEEE Trans. Med. Imag. 35(5), 1229 – 1239 (2016)
2. Ciresan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013 - 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II. pp. 411 – 418 (2013)
3. Commowick, O., Wiest-Daesslé, N., Prima, S.: Block-matching strategies for rigid registration of multimodal medical images. In: 9th IEEE International Symposium on Biomedical Imaging (ISBI'2012). pp. 700 – 703 (2012)

[1] `https://pythonhosted.org/nolearn/`
[2] `http://lasagne.readthedocs.io/en/latest/index.html`
[3] `http://deeplearning.net/software/theano/index.html`

4. Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans. Med. Imag. 27(4), 425 – 41 (2008)
5. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. ArXiv 1508.06576 (2015), http://arxiv.org/abs/1508.06576
6. Jain, V., Seung, S.: Natural image denoising with convolutional networks. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (eds.) Advances in Neural Information Processing Systems 21. pp. 769 – 776 (2009)
7. Kang, K., Wang, X.: Fully convolutional neural networks for crowd segmentation. ArXiv 1411.4464 (2014), http://arxiv.org/abs/1411.4464
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. ArXiv 1412.6980 (2014), http://arxiv.org/abs/1412.6980
9. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1106 – 1114 (2012)
10. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278 – 2324 (1998)
11. Manjon, J.V., Coupé, P.: volBrain: An online MRI brain volumetry system. In: Organization for Human Brain Mapping 15 (2015)
12. Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J.N.L., Isgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans. Med. Imag. 35(5), 1252 – 1261 (2016)
13. Tustison, N., Avants, B., Cook, P., Zheng, Y., Egan, A., Yushkevich, P., Gee, J.: N4ITK: Improved N3 bias correction. IEEE Trans. Med. Imag. 29(6), 1310 –1320 (2010)
14. Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D.: Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage 108, 214 – 224 (2015)

Random Forest for Multiple Sclerosis Lesion Segmentation

F.J. Vera-Olmos, H. Melero and N.Malpica

Universidad Rey Juan Carlos, Madrid, Spain
{javier.vera,norberto.malpica,helena.melero}@urjc.es

Abstract. Multiple Sclerosis (MS) is a chronic, inflammatory and demyelinating disease that primarily affects the white matter of the central nervous system. Automatic segmentation of MS lesions in brain MRI has been widely investigated in recent years with the goal of helping MS diagnosis and patient follow-up. It offers an attractive alternative to manual segmentation, which remains a time-consuming task and suffers from intra- and inter-expert variability. We propose a new approach that uses a Random Forest (RF) classifier. Its input has been filtered with a threshold based on the gray matter (GM) distribution and uses several features that take into account voxel and context information. A Markov Random Field (MRF) post processing algorithm has been applied to make lesions grow through probable neighborhoods. In order to train and test the method we use the database provided for the MICCAI 2016 challenge, which consists in 15 subject from 4 different centers, imaged on 1.5T or 3T scanners.The provided MR sequences include: 3D FLAIR, 3D T1-w, 3D T1-w GADO, 2D DP and 2D T2. In this data set each patient has being manually annotated by seven experts. We found that segmentation results are maximized by using all available sequences, but the FLAIR volume provides most of the information. Thus our method can work only with the T1-w volume to segment the tissues and the FLAIR volume to extract the most important features, still, its performance improves slightly if more types of sequences are available.

1 Introduction

Our approach is inspired in the work of O.Maier et al.[4] in which a RF classifier is used to classify longitudinal MS lesions with several features, Roura et al.[5] in which lesions are segmented as outliers to the normal apparent GM brain tissue on the FLAIR image and P. Smith et al.[7] where a MRF is used to makes lesions grow from an initial segmentation based on tissue thresholding. We use the RF to obtain an initial lesion segmentation instead of the tissue segmentation thresholding used in P. Smith et al.[7]. The RF used by O.Maier et al.[4] is trained with random sampled voxels. In contrast, we propose to use a FLAIR thresholding based in the GM distribution to reduce the number of points 3-4 times; this eliminates background voxels while preserving most of the lesion voxels. Finally, in a similar approach as P. Smith et al.[7] a MRF is used to make lesions grow through probable neighborhoods.

2 Materials and Method

We use the data provided by the MSSEG Challenge that will take place during MICCAI 2016[1]. This data consist in 15 subjects, of which the raw data, the bias-corrected and the skull-stripped versions of the following MRI sequences are provided: 3D FLAIR, 3D T1-w, 3D T1-w GADO, 2D DP and 2D T2. We work with the volumes already pre-processed by the challenge organizer. In order to train the algorithm, a manual segmentation annotated by 7 experts and a unique mask for every subject obtained using LOP STAPLE is also provided. We train RF with all sequence types provided by the challenge, in order to obtain the best result, but we also train the model using just the T1 and FLAIR sequences, to contrast the performance using a variable number of input sequences.

2.1 Pre-processing

The different training images of the challenge display large intensity differences, which usually occurs in MRI images acquired in different scanners. Using a learning-based intensity standardization method implemented in MedPy [3], we harmonize every sequence intensity profile [4].

2.2 Data filtering for Training

In order to reduce the training time of the classifier, the number of candidate voxels that are input to the classifier must be reduced. This is achieved by detecting the possible lesions as outliers of the apparently normal GM brain tissue on the FLAIR volume. The GM mask was used to compute the intensity distribution of the GM in the FLAIR image. To detect the outliers, we used the full width at half maximum (FWHM) of the main peak to determine the standard deviation [5]. The threshold is computed as follows:

$$Thr = \mu + \alpha \times \theta \quad (1)$$

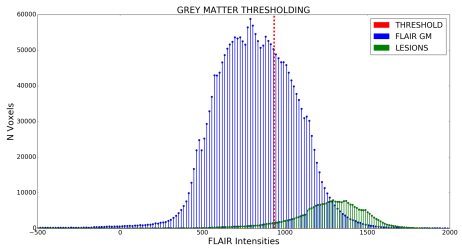


Fig. 1: Gray matter threshold

where μ is the mean intensity of the GM distribution and θ is the standard deviation determined using the FWHM. The parameter α is used to adjust the threshold, in order to obtain the maximum reduction of lesion candidates

while maximizing the true positive rate of the filter. This thresholding makes the training 3-4 times faster, it makes the data more manageable, and we found this method gives better results than extracting random voxels from the images, which is the approach used in [4]. We tested different values for α and selected 0.2 for the rest of our experiments, as it maintains a good TPR while reducing the voxels to a manageable amount.

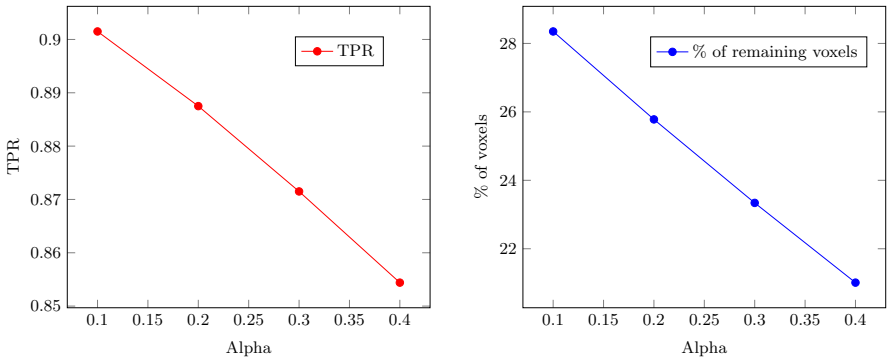


Fig. 2: True positive rate and % of remaining voxels for different α values

2.3 The Features

From each MRI sequence, a number of intensity-based features are extracted: voxel intensity value, voxel intensity value after smoothing (Gaussian filter with σ = 3; 5; 7mm) and the difference between the voxel and its neighborhood. These features provide information about gray-values at different scales as well as the mean intensity distribution in small areas around each voxel. To provide the classifier with a rough estimation of spatial location, we additionally compute the distance of every voxel to the image center. Since MS lesions appear primarily in white matter (WM), a probability based tissue segmentation is obtained with SPM12 on the MPRAGE/T1w sequence, separating the brain tissue into WM, gray-matter (GM) and cerebral spinal fluid (CSF). From the resulting tissue probability maps, we extract voxel gray-value and voxel gray value after smoothing (Gaussian filter at = 3; 7; 15 mm) [2]. Since it is easy to find false positive lesions in the external CSF, we also extract the external CSF in the FLAIR volume and calculate the distance from every voxel to this region.

2.4 The classifier

We use the RF implemented in Sckit-Learn[2]. We train it with the data filtered from the GM thresholding, using 200 trees with the maximum possible depth, to obtain pure leaves. The number of background voxels are 4-5 times greater

than the lesion voxels so we need to balance the weight of the classes. Therefore, we design the weights to be inversely proportional to the class frequencies of the input data of the RF.

With this classifier we obtain an initial lesion mask and a lesion probability mask. The initial lesion mask is obtained from the probability lesion mask using a threshold θ, with a value selected using cross validation. These two masks will be used to generate the final lesion segmentation.

2.5 Markov Random Field for lesion growing

Each voxel in the neighborhood of the initialized lesions obtained from the RF is labeled as lesion (Les) or non-lesion ($Non-Les$). This latter class consists of the three main tissue classes CSF, GM and WM. Thus, the discrete label for a voxel can be either Les or $Non-Les$. We approximate the distribution of Les by a gamma distribution with shape and scale parameters α and β, respectively, and the distribution of $Non-Les$ by a mixture of three Gaussians with means μ_k. Using the probability mask obtained from the Random Forest classifier, we compute the average of the neighbourhood of the voxel, to take into account if it is surrounded by probable lesion voxels. With these probability distributions and the initial lesion mask, we apply the MRF algorithm in every voxel i conected to a Les voxel using the following equation:

$$\text{If} \quad \frac{P_{les}(y_i|\alpha,\beta) * \frac{1}{N_i} * \sum_{j\in N_i} \pi_j^{les}}{P_{GMM}(y_i|\mu_k) * \frac{1}{N_i} * \sum_{j\in N_i}(1-\pi_j^{les})} \quad > \ \lambda \ \text{ then voxel } i \text{ is labeled as } Les \quad (2)$$

The λ value was selected using cross validation in order to obtain the best results .

3 Cross validation

We use the cross-validation to select the best values for the θ to obtain the initial lesion mask and the λ value used to stop the lesion growing. We use always one subject to test and the remaining subjects for training.

Table 1 show the different results for different values of λ and θ, we have selected 1.5 for the λ and 0.3 for the θ.

4 Results

Table 2 shows the inter-variability between the experts, while Table 3 summarizes the results of our algorithm using all sequences and just using the FLAIR and the T1-w. To compare our method with the state of the art, we chose the second algorithm of the Lesion Segment Tool (LST) LPA proposed by Paul Schmidt [6].

$\lambda \setminus \theta$	0.15	0.2	0.25	0.3	0.4
1	64.01	65.52	66.33	66.65	65.78
1.5	64.47	65.96	66.59	66.6	65.55
2	64.48	66.02	66.52	66.42	65.15
2.5	64.46	65.94	66.45	66.28	64.77

Table 1: Cross validation measuring the DICE metric in %

Experts	1	2	3	4	5	6	7	Average
DICE	0.649	0.644	0.596	0.664	0.687	0.595	0.588	0.632
TPR	0.681	0.739	0.587	0.786	0.788	0.566	0.508	0.665
FPR	0.641	0.661	0.423	0.831	0.605	0.303	0.182	0.498

Table 2: Expert inter-variability

	DICE	TPR	FPR
RF	0.651	0.689	0.437
RF FLAIR-T1	0.638	0.683	0.483
LST-LPA	0.610	0.633	0.460

Table 3: Evaluation of results

As we can see in the tables, our proposal improves the state of the art and it provides results which are similar to the inter-variability of the experts. Moreover, our method can achieve good results using just the FLAIR and T1-w sequences.

References

1. Msseg challenge. https://portal.fli-iam.irisa.fr/msseg-challenge/overview. Accessed: 2016-17-05.
2. L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, et al. Api design for machine learning software: experiences from the scikit-learn project. *arXiv preprint arXiv:1309.0238*, 2013.
3. O. Maier. Medpy. https://pypi.python. org/pypi/MedPy. Accessed: 2016-17-05.
4. O. Maier and H. Handels. Ms lesion segmentation in mri with random forests. *Proc. 2015 Longitudinal Multiple Sclerosis Lesion Segmentation Challenge*, pages 1–2, 2015.
5. E. Roura, A. Oliver, M. Cabezas, S. Valverde, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, and X. Lladó. A toolbox for multiple sclerosis lesion segmentation. *Neuroradiology*, 57(10):1031–1043, 2015.
6. P. Schmidt. Lesions were segmented by the lesion prediction algorithm as implemented in the lst toolbox version 2.0.15 for spm. www.statistical-modelling.de/lst.html. Accessed: 2016-17-05.
7. P. Schmidt, C. Gaser, M. Arsic, D. Buck, A. Förschler, A. Berthele, M. Hoshi, R. Ilg, V. J. Schmid, C. Zimmer, et al. An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. *Neuroimage*, 59(4):3774–3783, 2012.