Detecting the appearance of new T2-w multiple sclerosis lesions in longitudinal studies using deep convolutional neural networks

Mostafa Salem[1, 3], Sergi Valverde[1], Mariano Cabezas[1], Deborah Pareto[2], Arnau Oliver[1], Joaquim Salvi[1], Àlex Rovira[2] and Xavier Lladó[1]

[1] University of Girona, Department of Computer Architecture and Technology, Girona, Spain.

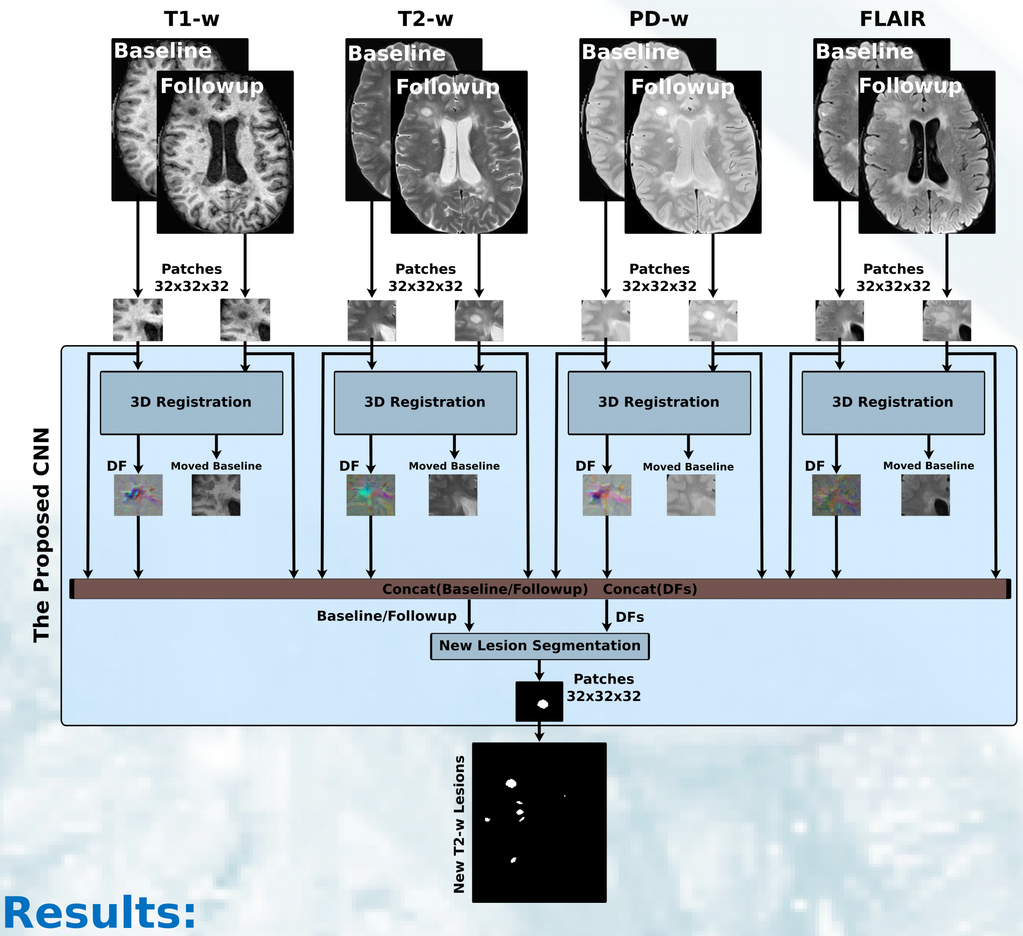
[2] Magnetic Resonance Unit (Department of Radiology). Vall d'Hebron University Hospital. Barcelona, Spain.

[3] Computer Science Department, Faculty of Computers and Information, Assiut University, Egypt

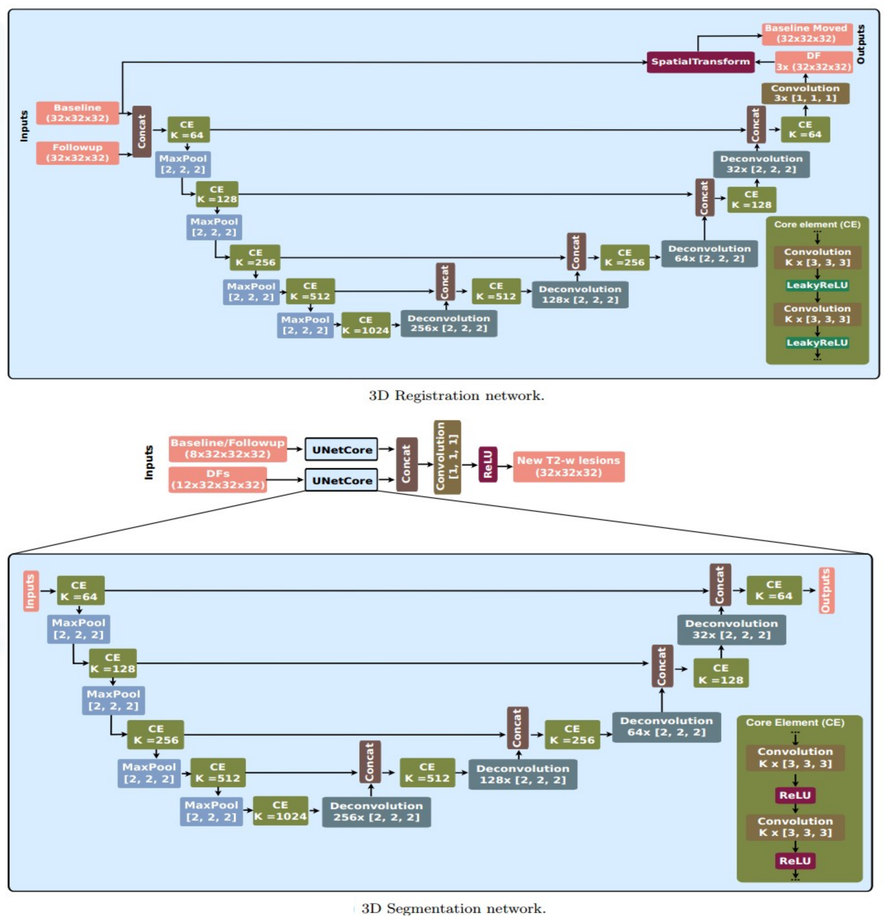
1. Introduction and purpose:

Longitudinal magnetic resonance imaging (MRI) analysis has an important role in multiple sclerosis (MS) diagnosis and follow-up. The presence of new T2-w lesions on brain MRI scans is considered a prognostic and predictive biomarker for the disease. Recently, several deformation-based approaches have been proposed for the detection of newly appearing lesions. In these approaches, the new T2-w lesions detection is performed by analyzing the deformation fields (DF) obtained by nonrigid registration between successive scans. The DF between time-points are fundamental in detecting the appearance of new T2-w MS lesions in longitudinal studies. These DF can either be obtained using classical optimisation approaches or learning-based approaches. In classical registration approaches, registration is defined as an optimization problem which is computationally intensive and extremely slow. In learning-based approaches, a parametrized registration function is learnt from a collection of images during training. Then, on testing time, a registration field can be quickly computed by directly evaluating the function using the learned parameters. We present here a fully convolutional neural network (FCNN) to detect new T2-w lesions in longitudinal brain MRI images. The proposed model is trained end-to-end, learning the DFs and the new T2-w lesions simultaneously using a combined loss function.

2. Methods:



Networks in detail



3D Registration network.

3D Segmentation network.

3. Results:

The database used in this study consists of images (T1-w, T2-w, PD-w, and FLAIR) from 60 different patients with a clinically isolated syndrome (CIS) or early relapsing MS. 36 of the patients confirmed MS with new T2-w lesions, while 24 patients do not present new T2-w lesions. We evaluated the proposed framework in two scenarios. Firstly, we analyzed the accuracy of the detection using a leave-one-out cross-validation strategy with 36 patients with new MS lesions. This strategy was applied per patient on our 36 images from the MS patient dataset. Secondly, we analyzed the specificity of the method with the 24 patients with no new T2-w lesions. To do this, we performed a new training using all the 36 images with new MS lesions. We compared also the obtained results with those of recent state-of-the-art approaches (Sweeney et al., 2013; Cabezas et al., 2016; Salem et al., 2018; Schmidt et al., 2019). Standard measures such as the true positive fraction (TPF), the false positive fraction (FPF), and the Dice similarity coefficient (Detection (DSCd) and Segmentation (DSCs)) were used for the evaluation.

Method	TPF	FPF	DSCd	DSCs
SimLearnedDFs	**83.09 ± 21.06**	**9.36 ± 16.97**	**0.83 ± 0.16**	**0.54 ± 0.18**
SepLearnedDFs	57.77 ± 34.34	13.67 ± 21.99	0.60 ± 0.31	0.39 ± 0.29
DemonsDFs	62.06 ± 32.74	11.98 ± 23.09	0.67 ± 0.29	0.48 ± 0.21
NDFs	53.99 ± 38.01	17.20 ± 26.96	0.55 ± 0.35	0.37 ± 0.28
Sweeney et al. (2014)	59.82 ± 37.59	33.59 ± 33.52	0.57 ± 0.33	0.44 ± 0.26
Cabezas et al. (2016)	70.93 ± 34.48	17.80 ± 27.96	0.68 ± 0.33	0.53 ± 0.24
Salem et al. (2018)	80.0 ± 27.77	21.87 ± 26.26	0.76 ± 0.25	0.55 ± 0.22
Schmidt et al. (2019) (κ = 0.15)	68.66 ± 35.26	31.89 ± 36.10	0.62 ± 0.34	0.40 ± 0.25

NOTE: SimLearnedDFs (proposed model), SepLearnedDFs (registration blocks and the segmentation blocks were trained separately), DemonsDFs (proposed network using the DFs obtained with the Demons), NDFs (proposed network without DFs).

Analyzing the results according to the different lesion size :

Analysis of TPF for different classifiers for different lesion sizes. Lesions between 3 and 10 voxels are considered small; lesions between 11 and 50 voxels, medium; and lesions with 50 voxels, large.

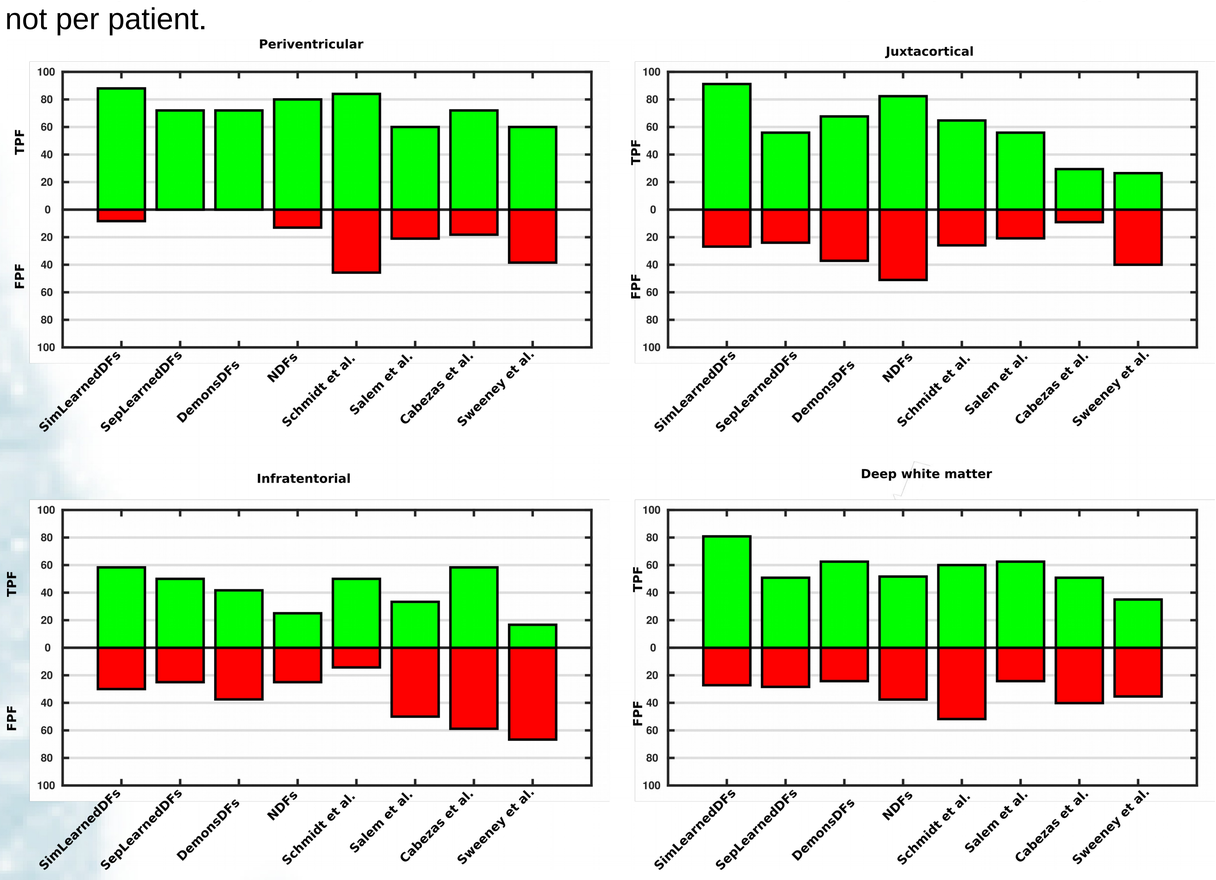
Method	Small (3 - 10)	Medium(11-50)	Large (+50)
SimLearnedDFs	**39.52**	**83.32**	**97.14**
SepLearnedDFs	22.62	49.40	83.14
DemonsDFs	30.0	78.25	90.26
NDFs	14.29	47.71	80.48
Sweeney et al. (2014)	16.67	52.06	78.25
Cabezas et al. (2016)	42.86	48.57	77.42
Salem et al. (2018)	34.40	65.70	91.30
Schmidt et al. (2019) (κ = 0.15)	13.10	71.92	94.08

Analyzing the results of the 24 cases with no new T2-w lesions:

We evaluated the 24 patients with no new T2-w lesions, after training the SimLearnedDFs model with all the 36 patients with new T2-w lesions. The obtained results showed only 2 cases with one FP detection in each, being these results better than those obtained with the other approaches.

Analyzing the results according to its location in the brain:

We studied also the performance of the different approaches on different brain regions according to the lesion location in the brain (periventricular, juxtacortical, infratentorial, and deep white matter). Note that the TPF and FPF were computed per lesion type and not per patient.



Statistical Analysis:

A paired t-test at the 5% level was used to evaluate the significance of the results of the proposed method. As mentioned in previous Table, the SimLearnedDFs model was significantly better than all the other methods except the Salem et al. method ($p < 0.05$). However, the TPF improved by 3%. In terms of FPF, the SimLearnedDFs model was not significantly better than the SepLearnedDFs (4.31% improvement) and the DemonsDFs (2.62% improvement), but it was significantly better than the other methods ($p < 0.05$).

4. Conclusion:

The obtained results indicate that the combination of DFs and supervised classification increases the accuracy when detecting new T2-w lesions. Given the sensitivity and limited number of false positives, we strongly believe that the proposed method may be used in clinical studies in order to monitor the progression of the disease.

Acknowlegments:

M. Salem holds a grant for obtaining the Ph.D. degree from Egyptian Ministry of Higher Education. This work has been supported by by Retos de Investigación TIN2015-73563-JIN and DPI2017-86696-R.