Manual delineation of only one image in unseen databases is sufficient for accurate performance in automated multiple sclerosis lesion segmentation

Sergi Valverde[1], Mostafa Salem[1], Mariano Cabezas[1], Deborah Pareto[2], Joan C. Vilanova[3], Lluís Ramió-Torrentà[4], Àlex Rovira[2], Joaquim Salvi[1], Arnau Oliver[1] and Xavier Lladó[1]

1. Computer Vision and Robotics Research Institute (VICOROB), University of Girona, Girona (Spain)
2. Magnetic Resonance Unit, Dept. of Radiology, Vall d'Hebron University Hospital, Barcelona (Spain)
3. Girona Magnetic Resonance Center, Girona (Spain)
4. Multiple Sclerosis and Neuro-immunology Unit, Dr. Josep Trueta University Hospital, Girona (Spain)

1 INTRODUCTION and PURPOSE:

Convolutional Neural Network (CNN) methods are being proposed for automated white matter lesion segmentation, increasing the performance of typical state-of-the-art methods. However, their accuracy decreases significantly when using them on other image domains that those used for training, showing lack of adaptability to unseen imaging data and limiting its applicability in non-specialized hospitals. Here, we analyze the effect of domain adaptation on multiple sclerosis (MS) lesion segmentation, investigating how transferable a CNN model is when applied to other unseen image domains.

2 METHOD:

2.1 Source model:

A 11-layer CNN classifier[1] was firstly fully-trained using 35 T1-w and FLAIR scans from the MS lesion segmentation challenges (MICCAI 2008[2] and MICCAI 2016[3]).

2.2 Domain adaptation:

Using a **single example image** from the *target* dataset, the same source model is fine-tuned re-training only part or all the connected (FC) layers, analyzing the performance of the classifier under different lesion loads (**Figure 1**)

2.3 Target datasets:

Two independent datasets from **different hardware vendors and image protocols** are used to evaluate the transfer knowledge capability:
- 30 T1w and FLAIR images from a clinical MS dataset (Vall d'Hebron, Barcelona)
- 61 T1-w and FLAIR images from the public ISBI2015[4] MS lesion challenge.

3 RESULTS:

We compute the ***Dice*** overlap (DSC), ***sensitivity*** and ***precision*** coefficients between the resulting segmentations and manual lesion annotations. Domain adapted segmentations are also compared with respect to the same model fully trained on the target domain and with respect to other methods such as LST[5], SLS[6] or other CNN methods.

.

lesion vol (num lesions)	DSC	sensitivity	precision
1 layer (FC3)			
0.5 ml (9 lesions)	0.30 (0.19)	0.44 (0.23)	0.49 (0.30)
1.2 ml (11 lesions)	0.39 (0.19)	0.44 (0.19)	0.67 (0.23)
3.1 ml (17 lesions)	0.38 (0.22)	0.46 (0.20)	0.54 (0.25)
8.3 ml (90 lesions)	0.44 (0.17)	0.58 (0.19)	0.58 (0.26)
18 ml (78 lesions)	0.47 (0.18)	0.59 (0.18)	0.58 (0.23)
2 layers (FC2 + FC3)			
0.5 ml (9 lesions)	0.30 (0.17)	0.52 (0.23)	0.54 (0.28)
1.2 ml (11 lesions)	0.39 (0.18)	0.49 (0.21)	0.72 (0.29)
3.1 ml (17 lesions)	0.36 (0.22)	0.42 (0.20)	0.54 (0.27)
8.3 ml (90 lesions)	0.45 (0.15)	0.55 (0.18)	0.66 (0.24)
18 ml (78 lesions)	0.44 (0.19)	0.62 (0.20)	0.52 (0.25)
3 layers (FC1 + FC2 + FC3)			
0.5 ml (9 lesions)	0.28 (0.17)	0.48 (0.22)	0.48 (0.28)
1.2 ml (11 lesions)	0.38 (0.17)	0.52 (0.22)	0.72 (0.26)
3.1 ml (17 lesions)	0.38 (0.21)	0.46 (0.21)	0.55 (0.25)
8.3 ml (90 lesions)	0.44 (0.17)	0.61 (0.17)	0.57 (0.26)
18 ml (78 lesions)	0.45 (0.18)	0.60 (0.21)	0.55 (0.23)
Source (0 lesions)	0.23 (0.22)	0.42 (0.43)	0.45 (0.34)
SLS	0.25 (0.17)	0.34 (0.25)	0.51 (0.30)
LST	0.28 (0.23)	0.31 (0.21)	0.59 (0.27)
CNN	0.53 (0.16)	0.60 (0.21)	0.75 (0.21)

Table 1: ***Clinical MS dataset****: DSC, sensitivity and precision coefficients for each of the models re-trained using a single case with varying degree of lesion load. For comparison, the obtained values for SLS, LST and the same cascaded CNN method fully trained using the 30 available training cases are also shown. For each coefficient, the reported values are the mean (standard deviation) when evaluated on the 30 testing case.*

Method	DSC	sensitivity	precision	score	
Andermatt et al. (2017)	0.63 (0.14)	0.54 (0.19)	0.84 (0.10)	92.07	
Hashemi et al. (2018)	0.66 (0.11)	0.67 (0.20)	0.71 (0.16)	91.52	
Valverde et al. (2017)	0.64 (0.12)	0.57 (0.17)	0.79 (0.15)	91.44	full train
Birenbaum and Greenspan (2017)	0.63 (0.14)	0.55 (0.18)	0.80 (0.15)	91.26	
Roy et al. (2018)*	0.52 (- -)	- - (- -)	0.86 (- -)	90.48	
Deshpande et al. (2015)	0.60 (0.13)	0.55 (0.17)	0.73 (0.18)	89.81	
Jain et al. (2015)	0.55 (0.14)	0.47 (0.15)	0.73 (0.20)	88.74	
Shiee et al. (2010)	0.55 (0.19)	0.54 (0.15)	0.70 (0.29)	88.46	
Valcarcel et al. (2018)	0.57 (0.13)	0.57 (0.18)	0.61 (0.16)	87.71	
Sudre et al. (2015)	0.52 (0.14)	0.46 (0.15)	0.66 (0.18)	86.44	
one-shot (3 layers, 26.8 ml.)	0.58 (0.16)	0.48 (0.19)	0.84 (0.13)	90.32	single subject

(*) Obtained results for Roy et al. (2018) were extracted from the related publication.

Table 2: ***ISBI challenge:*** *DSC, sensitivity, precision and overall score coefficients for the best domain adapted model (ISBI02 with 3 layers) after submitting the segmentation masks for blind evaluation. The obtained results are compared with different top rank participant strategies. For each method, the reported values are extracted from the challenge results board. The reported values are the mean (standard deviation) when evaluated on the 61 testing images.* ***The performance of the methods with an overall score>=90 is considered to be similar to human performance****.*

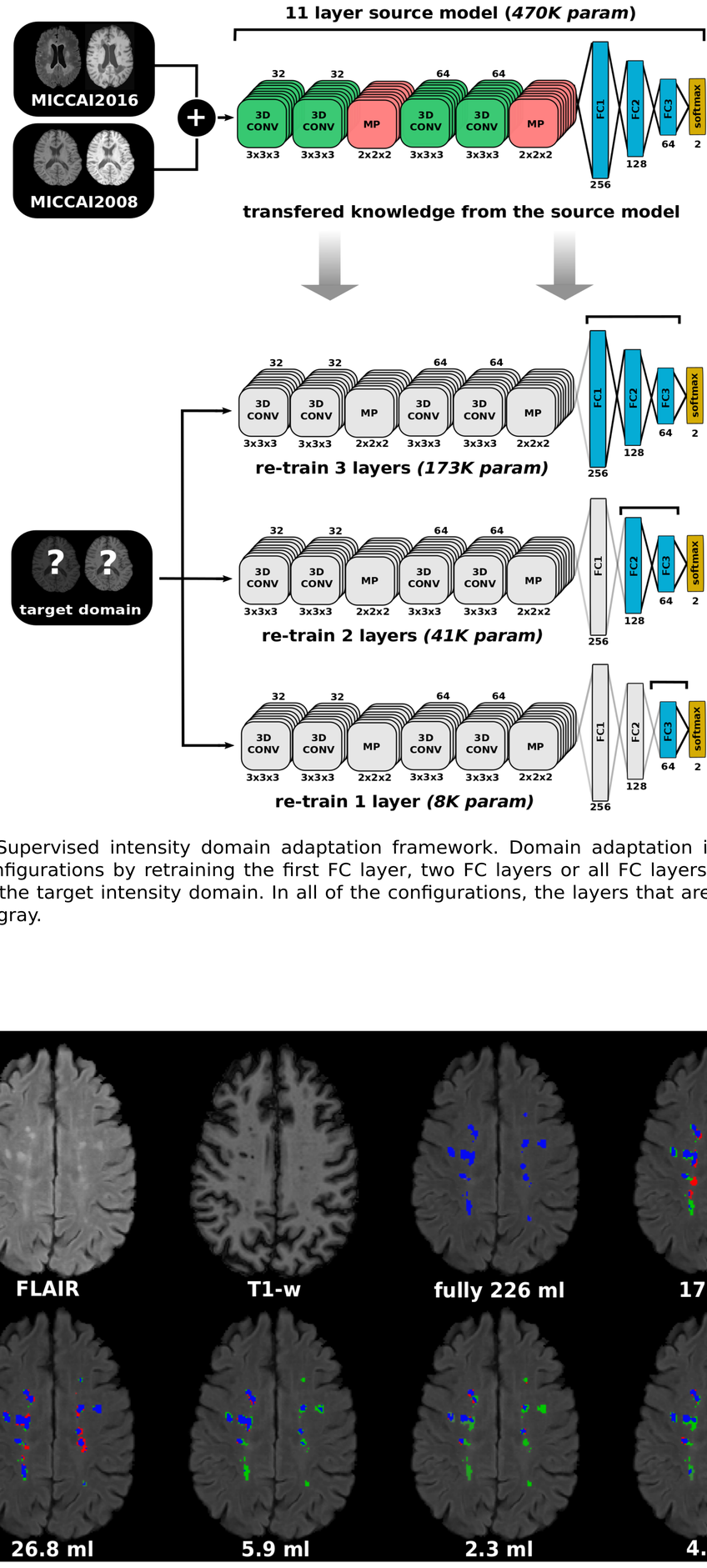

Figure 1: Supervised intensity domain adaptation framework. Domain adaptation is performed via 3 possible configurations by retraining the first FC layer, two FC layers or all FC layers using images and labels from the target intensity domain. In all of the configurations, the layers that are not re-trained are depicted in gray.

Figure 2: *Lesion volume comparison on training performance between different single subject training sets. The blue regions depict the overlapped lesions between the fully trained and each of the single subject models. The red and green regions depict false-positive and false-negative lesions, respectively, with respect to the full trained model.*

4. CONCLUSIONS:

Domain adaptation allows to use pre-trained CNNs on unseen clinical settings. A manual delineation of the lesions in only one image is sufficient to obtain accurate automated lesion segmentation performances.

REFERENCES:

1. *Valverde S. et al (2017) Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. NeuroImage, 155, 159-168.*
2. *MICCAI2008 MS lesion segmentation challenge: http://www.ia.unc.edu/MSseg/about.htm*
3. *MICCAI2016 MS lesion segmentation challenge: https://portal.fli-iam.irisa.fr/msseg-challenge/workshop-day*
4. *ISBI2015 MS lesion segmentation challenge: https://biomedicalimaging.org/2015/3d-segmentation-of-neurites-in-em-images/*
5. *Schmidt, P. et al. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. NeuroImage, 59(4), 3774–83.*
6. *Roura, E. et al. (2015). A toolbox for multiple sclerosis lesion segmentation. Neuroradiology, 57(10), 1031–1043.*

ACKNOWLEDGMENTS:

This work has been supported by "La Fundació la Marató de TV3", by Retos de Investigación TIN2014-55710-R, MPC UdG 2016/022, TIN2015-73563-JIN and DPI2017-86696-R grants from the Ministerio de Ciencia y Tecnología. The authors gratefully acknowledge the support of the NVIDIA Corporation with their donation of the TITAN-X used in this research.

CONTACT INFORMATION:

Sergi Valverde, PhD. University of Girona (UdG). Av. Lluís Santaló s/n (P4-building). 17071, Girona (Spain)
Phone: (+34) 972419812 E-mail: svalverde@eia.udg.edu

Download!

The proposed method is freely available here: