



**HAL**  
open science

## **CNN for multiple sclerosis lesion segmentation: How many patients for a fully supervised method?**

Alexandre Fenneteau, Pascal Bourdon, David Helbert, Christine Fernandez-Maloigne, Christophe N Habas, Rémy Guillevin

► **To cite this version:**

Alexandre Fenneteau, Pascal Bourdon, David Helbert, Christine Fernandez-Maloigne, Christophe N Habas, et al.. CNN for multiple sclerosis lesion segmentation: How many patients for a fully supervised method?. International Conference on Advances in Biomedical Engineering, Oct 2021, Wardanyeh, Lebanon. pp.30-33, 10.1109/ICABME53305.2021.9604859 . hal-03335735

**HAL Id: hal-03335735**

**<https://hal.science/hal-03335735v1>**

Submitted on 19 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# CNN for multiple sclerosis lesion segmentation: How many patients for a fully supervised method? \*

Alexandre Fenneteau<sup>1,2,3,4</sup>, Pascal Bourdon<sup>2,3</sup>, David Helbert<sup>2,3</sup>,  
Pascal Bourdon<sup>2,3</sup>, Christine Fernandez-Maloigne<sup>2,3</sup>, Christophe  
Habas<sup>3,4,6</sup>, and Rémy Guillevin<sup>3,5,7</sup>

<sup>1</sup>Siemens Healthcare, Saint Denis, France

<sup>2</sup>XLIM Laboratory, University of Poitiers, UMR CNRS 7252; Poitiers,  
France

<sup>3</sup>I3M, Common Laboratory CNRS-Siemens, University and Hospital  
of Poitiers ; Poitiers, France

<sup>4</sup>Neuroimaging Department, Quinze Vingts Hospital; Paris, France

<sup>5</sup>Poitiers University Hospital, CHU; Poitiers, France

<sup>6</sup>University of Versailles Saint-Quentin; Versailles, France

<sup>7</sup>DACTIM-MIS/LMA Laboratory University of Poitiers, UMR CNRS  
7348; Poitiers, France

November 19, 2021

## Abstract

In this study we propose to improve an existing artificial neural network architecture, the MPU-net, which is designed for having very few parameters for multiple sclerosis lesion segmentation on magnetic resonance images. With this improved architecture we conducted a study to assess the influence of the number of training examples on the model performance and generalization. The question behind this study is: “With an appropriate architecture, how many patients do we need?”. We evaluated 9 different adaptations of the MPU-net architecture. Then, after the selection of the best architecture we learned the model multiple times with different numbers of patients and assessed its performances. The addition of deep supervision, the reduction of number of convolutional layers and the addition of regularization layers produced a more stable and performant architecture. Learnings of selected model with only 10 exams delivered performances equivalent to learnings with 23 exams. So, in our experimental setup, it is possible to learn a performant model with only 10 fully annotated examples.

**Keywords** - Segmentation, Deep Learning, Few examples, Multiple Sclerosis

---

\*corresponding author: alexandre.fenneteau@siemens-healthineers.com

# 1 Introduction

Multiple Sclerosis (MS) is an auto-immune disease which affects the central nervous system (brain and spinal cord). It causes lesions and demyelination of nerves. These reactions can induce visual, mental, motor disorders and physical disability.

In Magnetic Resonance (MR) images the radiologist can detect, classify and quantify lesions induced by MS and orient patient care and treatment accordingly. The lesion screening is time-consuming and subject to intra and inter observer variability[1]. The automatic segmentation of MS lesion has the potential to bring more reproducibility and to help radiologists in their daily tasks.

There are different types of algorithms for MS lesion segmentation from brain MR images, including statistical models, classic machine learning models and more recently deep learning models. Several deep learning methods for MS lesion segmentation have been proposed [2]. Those methods are highly supervised, tend to outperform other approaches and reach a segmentation performance comparable to human [3]. Currently, the trainings of supervised methods are performed on public or private segmentation data sets. The clinical or private data sets used have a number of annotated subject from a few dozen to more than 1000 [2]. The most known public data sets (MICCAI 2008 [4], MICCAI 2016 [5] and ISBI 2015 [3, 6]) contain not much than 21 available annotated exams each. Public data sets are extensively used but are their size sufficient? For this study we used the ISBI 2015 data set to stay consistent with previous work and the Lesjak-2018 data set which is the public data set that contains the biggest number of annotated exams. We chose not to assemble data sets to avoid issues on working with multi-center MR images.

In a previous study, we did not improve the performance of a Convolutional Neural Network (CNN) by using self-supervision for MS lesion segmentation [7]. One hypothesis advanced was that the data set size was already sufficient and, therefore, pre-training would not improve segmentation predictions. Driven by previous work on finding an efficient CNN architecture for MS lesion segmentation in terms of number of learnable parameters the Minimally Parameterized U-net (MPU-net) [8], we propose 1) to improve the MPU-net and 2) to assess the needed number of training patients for this light architecture. We did not find published architectures with a comparable low number of learnable parameters. Therefore, we did not compare our results to another method. We show that the learning of a MS lesion segmentation task with a modified MPU-net only requires a few fully annotated samples.

## 2 Material and methods

The experiment is split in two successive steps:

1. Refinement of the MPU-net architecture
2. Evaluation of the required number of training patient

### 2.1 Architecture refinement

We started with the MPU-net architecture see Fig. 1 but as said in [8], it can be improved especially by adding regularizers such as dropout [9] and batch normalization [10] layers. Inspired by the U-net++ [11] we wanted to add deep supervision and to extend the architecture as shown in Fig. 1.

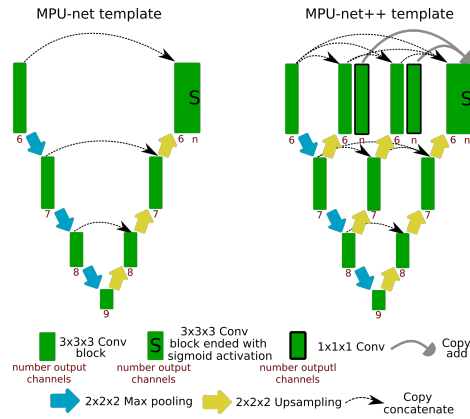


Figure 1: The architecture template for the MPU-net and the MPU-net++. The content of conv block is detailed in Fig. 2.

We evaluated 9 different variations of the MPU-net based on the architecture template of Fig. 1. The variation consists in adding batch normalization, dropout layers and including residual blocks [12] in the encoder part of the model. For the MPU-net++ template, the number of consecutive convolutions in each block has also been evaluated since the addition of multiple decoders increased the number of convolutional layers from 21 in the original MPU-net to 32 in the MPU-net++. The evaluated variations are detailed and named in Fig. 2.

### 2.2 Training patient number

The evaluation of required training patients was performed on the Lesjak-2018 data set [13] described in 2.3.1 with 30 segmented patients. During

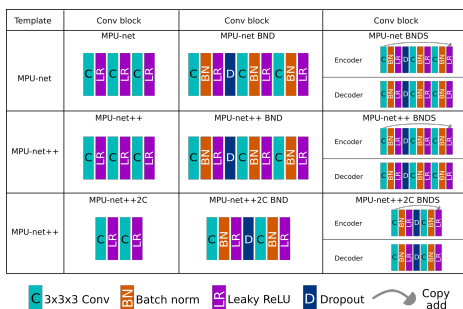


Figure 2: The different variations of MPU-net tested. For each variation the template see Fig. 1 and convolutional block change

this study we kept the same 6 patients for the testing set and the remaining for training set.

We evaluated 7 different sizes of sub-training set with a number of patients of: 1;3;5;8;10;15;19;23. For each sub-training set, the subjects were randomly selected among the 24 patients of the training set. Each training was repeated 5 times with new random selections to see the influence of the number of patients on training robustness and to lower the impact of patient selection (a set can be more instructive than another one).

## 2.3 Data

During the study, following [8], only T2-fluid-attenuated inversion recovery (FLAIR) images were used among all MR images available in each exam since they are the most effective in practice [14].

### 2.3.1 Data sets

For this study we used two different data sets: one for the architecture refinement and the other one for the evaluation of the training patient number.

The MPU-net was designed and evaluated on the ISBI 2015 data set. We decided to continue its improvement with the same set to stay consistent with the first study. But, this data set contains multiple exams from only 5 different patients. To avoid this limitation, we opt to select another data set for the evaluation of training patients number, the Lesjak-2018 data set with 30 exams from different patients.

**ISBI 2015** The data set contains 21 MR exams belonging to 5 different patients. Each patient has 4 to 5 consecutive exams within a one-year span.

Images were preprocessed, skull-stripped and resampled to a  $1 \text{ mm}^3$  isometric resolution. The ground truth segmentation of MS lesion from 4 different radiologists and 1 consensus are included. The split between training and validation sets have been performed accordingly to [8] with exams from one patient as the testing set and the exams from all remaining patients as training set.

**Lesjak-2018** The data set consists of 30 MR exams from different patients. Images were preprocessed and skull-stripped. The resolution of images is  $0.8 \times 0.47 \times 0.47 \text{ mm}^3$ . We chose to resample images to a resolution of  $0.8 \times 0.8 \times 0.8 \text{ mm}^3$  to have 3D convolutions spatially consistent. The ground truth segmentation provided is the consensus of 3 different radiologists' segmentation. 6 patients were kept for the testing set. The training set and subsets were selected among the 24 remaining patients.

### 2.3.2 Pre-processing

For each data set we performed an additional intensity preprocessing including histogram normalization [15] and z-normalization consisting in subtracting to each image its mean intensity and dividing by its standard deviation.

## 2.4 Learning and testing specifications

The trainings were performed on  $32 \times 32 \times 32$  patches randomly extracted from the brain area with a batch size of 256 for the ISBI 2015 data-set and of 161 for the Lesjak-2018 data set. The model was trained 30 epochs containing 26,100 patches each on 4GPUS. The Adam [16] optimizer with a learning rate of 0.004 and the Dice loss were used for the training as in [8].

For the testing, the patches are extracted all along the volumes with an overlap of 8 in each dimension. The segmentation result is spatially averaged and thresholded to set to background values  $< 0.6$  and to foreground remaining values.

## 3 Result and analysis

### 3.1 Architecture refinement

Following method described in 2.1, we evaluated 9 different adaptation of the MPU-net 3 times each.

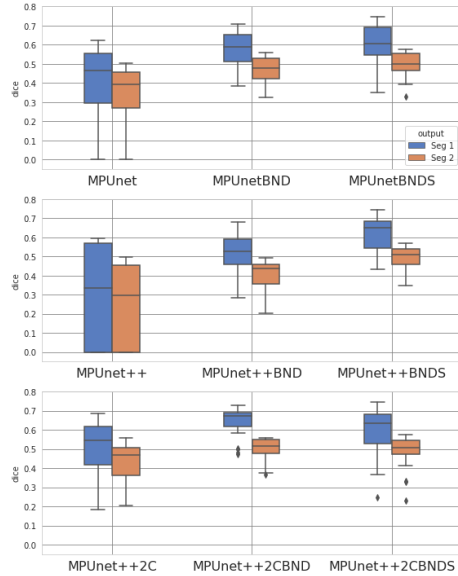


Figure 3: Box plot of the test results in terms of Dice score for each architecture variation. The box shows the quartiles of Dice for each variation while the whiskers show the rest of the distribution, except for outliers showed as points.

The box plot of test results on Fig. 3 shows that adding batch normalization and dropout layers improved the performances over the testing set. Those regularizers also improved learning reproducibility since they reduced the range of results quartiles and distribution. The residual implementation (3<sup>rd</sup> column) improved also regularized architectures (2<sup>nd</sup> column) except the MPU-net++2CBND model. We make the assumption that residuality in encoder might help segmentation prediction but can also brings instability when the model is sufficiently regularized and well-designed. Note that the MPU-net++ confidence intervals are very large, this is in accord with the fact that by adding multiple convolutional layers without regularizers, it drastically decreases the performances and the robustness of the model.

We notice that adding deep supervision while reducing the number of convolutional layers gave improvements in training robustness and testing performance. Adding batch normalization and dropout layers improved again model training and performance. Residual blocks in encoder might help but in our case, the best results were obtained with the MPU-net++2CBND architecture. We continue the next experiments with the MPU-net++2CBND architecture, this architecture is constituted by 22 convolution layers and has a total of 37,935 learnable parameters versus 21 convolutional layers and 33,332 learnable parameters for the original MPU-net. In terms of com-

plexity the improvement observed only cost 1 convolutional layer and  $\tilde{4000}$  learnable parameters which is acceptable.

### 3.2 Training patient number

To evaluate the required number of training patients we repeated experiments 5 times following the 2.2.

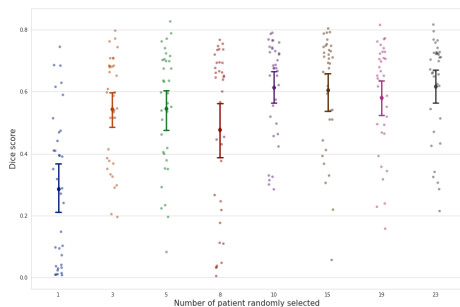


Figure 4: Performances in Dice score depending on the training patient number. Each dot represents a tested patient. The whiskers represent standard deviation centered around the average.

In Fig. 4 we can see that, best average Dice scores (near 0.6) are obtained from 10 training patients. Learning from a single patient does not seem to be a good strategy since the mean Dice score is low (0.3) but is almost doubled with 3 learning patients. We did not see significant difference between learning with 3 or 5 patients. Surprisingly, learning with 8 patients reduces performances compared to 5 patients which is something we did not expect.

This shows that, with our experimental setup, learning with only 10 training patients is equivalent to learning with 23 patients. Note that learning with 3 patients gave lower segmentation performances than with 10 patients but still delivered honest segmentation quality. The fact that learning with 8 patients is worst than with 5 patients in average was not expected. The random choice of learning patients may have been a bad combination in some trainings or composed by cases particularly different to those in the testing set.

## Conclusion

Many algorithms for the automatic segmentation of MS lesion in brain MR images have been proposed. This has the potential to bring more reproducibility in a daily radiologist task and to improve the quality of patient



care.

In this study we focused on improving the MPU-net, a “minimal architecture” in terms of number of parameters since we did not find lighter published architecture. As a low number of parameters should require a low number of training examples to be able to generalize, we study the quality of trained models with different number of learning patients.

We show that the MPU-net++2CBND architecture was the better evaluated for our task and experimentally observed that best performances on our testing set were achieved with only 10 learning patients.

This study should be extended to other data sets, with more repetitions to give more reliability to the work performed. This study is constrained by the size of data sets and the random choice of learning patients.

Our results show that with a low number of parameters and a regularized U-net like architecture, the MPU-net++2CBND, we can learn a performant model for MS lesion segmentation with only 10 learning patients and that we can lower this number to 3 to learn a model with reasonable performances. This work is a preliminary work for studying few shot and weakly supervised learning for this particular task and see how many examples and supervision we need to learn an efficient model for medical image segmentation.

## References

- [1] 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.
- [2] A. Shoeibi, M. Khodatars, M. Jafari, P. Moridian, M. Rezaei, R. Alizadehsani, F. Khozeimeh, J. M. Gorriz, J. Heras, M. Panahiazar, *et al.*, “Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review,” *arXiv preprint arXiv:2105.04881*, 2021.
- [3] A. Carass, S. Roy, A. Jog, J. L. Cuzzocreo, E. Magrath, A. Gherman, J. Button, J. Nguyen, F. Prados, C. H. Sudre, *et al.*, “Longitudinal multiple sclerosis lesion segmentation: resource and challenge,” *NeuroImage*, vol. 148, pp. 77–102, 2017.

- [4] M. Styner, J. Lee, B. Chin, M. S Chin, O. Commowick, H.-H. Tran, V. Jewells, and S. Warfield, “3d segmentation in the clinic: A grand challenge ii: Ms lesion segmentation,” *MIDAS Journal*, 11 2007.
- [5] O. Commowick, A. Istace, M. Kain, B. Laurent, F. Leray, M. Simon, S. C. Pop, P. Girard, R. Ameli, J.-C. Ferré, *et al.*, “Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure,” *Scientific reports*, vol. 8, no. 1, p. 13650, 2018.
- [6] A. Carass, S. Roy, A. Gherman, J. C. Reinhold, A. Jesson, T. Arbel, O. Maier, H. Handels, M. Ghafoorian, B. Platel, *et al.*, “Evaluating white matter lesion segmentations with refined sørensen-dice analysis,” *Scientific Reports*, vol. 10, no. 1, pp. 1–19, 2020.
- [7] A. Fenneteau, P. Bourdon, D. Helbert, C. Fernandez-Maloigne, C. Habas, and R. Guillevin, “Learning a CNN on multiple sclerosis lesion segmentation with self-supervision,” in *3D Measurement and Data Processing, IS&T Electronic Imaging 2020 Symposium*, (San Francisco, United States), #Jan# 2020. Best paper award 3D Measurement and Data Processing.
- [8] A. Fenneteau, P. Bourdon, D. Helbert, C. Fernandez-Maloigne, C. Habas, and R. Guillevin, “Investigating efficient cnn architecture for multiple sclerosis lesion segmentation,” *Journal of Medical Imaging*, vol. 8, no. 1, p. 014504, 2021.
- [9] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [10] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.
- [11] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [13] Ž. Lesjak, A. Galimzianova, A. Koren, M. Lukin, F. Pernuš, B. Likar, and Ž. Špiclin, “A novel public mr image dataset of multiple sclero-

- sis patients with lesion segmentations based on multi-rater consensus,” *Neuroinformatics*, vol. 16, pp. 51–63, Jan 2018.
- [14] Y. Feng, H. Pan, C. H. Meyer, and X. Feng, “A self-adaptive network for multiple sclerosis lesion segmentation from multi-contrast mri with various imaging protocols.,” *CoRR*, vol. abs/1811.07491, 2018.
- [15] L. G. Nyúl, J. K. Udupa, and X. Zhang, “New variants of a method of mri scale standardization,” *IEEE transactions on medical imaging*, vol. 19, no. 2, pp. 143–150, 2000.
- [16] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.