



HAL
open science

Segmentation of new multiple sclerosis lesions on FLAIR MRI using online hard example mining

Marius Schmidt-Mengin, Arya Yazdan-Panah, Théodore Soulier, Mariem Hamzaoui, Nicholas Ayache, Olivier Colliot

► To cite this version:

Marius Schmidt-Mengin, Arya Yazdan-Panah, Théodore Soulier, Mariem Hamzaoui, Nicholas Ayache, et al.. Segmentation of new multiple sclerosis lesions on FLAIR MRI using online hard example mining. MICCAI-MSSEG-2 - 25th International Conference on Medical Image Computing and Computer Assisted Intervention - challenge on multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure, Sep 2021, Strasbourg, France. hal-03826787

HAL Id: hal-03826787

<https://hal.science/hal-03826787>

Submitted on 24 Oct 2022

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.

Segmentation of new multiple sclerosis lesions on FLAIR MRI using online hard example mining

Marius Schmidt-Mengin^{1,2}, Arya Yazdan-Panah¹, Théodore Soulier¹, Mariem Hamzaoui¹, Bruno Stankoff¹, Nicholas Ayache³, and Olivier Colliot^{1,2}[0000-0002-9836-654X]

¹ Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France

² Inria - Aramis project-team, Paris, France

³ Inria - Epione project-team, Sophia-Antipolis, France
`olivier.colliot@sorbonne-universite.fr`

Abstract. This paper summarizes our contribution to the MSSEG-II MICCAI 2021 challenge. The aim is to segment new multiple sclerosis (MS) lesions using pairs of FLAIR MR images. Our approach is based on a 3D U-Net applied patch-wise to the images. In order to take into account both time-points, we simply concatenate the images along the channel axis before passing them to the 3D U-Net. The strong imbalance between positive and negative voxels, exhibited by the challenge data, makes training deep learning model difficult. Instead of using handcrafted priors like brain masks or multi-stage methods, we experiment with a novel modification to online hard example mining (OHEM), where we use an exponential moving average (i.e., its weights are updated with momentum) of our 3D U-Net to mine hard examples. Using a moving average instead of the raw model should allow smoothing its predictions and allowing it to give more consistent feedback for OHEM.

Keywords: Segmentation · Deep learning · Hard example mining · Multiple Sclerosis · MRI.

1 Introduction

The MSSEG-II MICCAI 2021 challenge aims at segmenting new multiple sclerosis (MS) lesions on brain FLAIR MRIs. The present paper describes our contribution to the challenge. We use a patch-wise 3D U-Net [1]. One important characteristic of the dataset is the very strong imbalance between positive and negative voxels (0.005% of positive voxels). We address this problem by performing online hard example mining [4] (OHEM). Notably, we use a moving average of our 3D U-Net to perform inference for hard example mining. Our intuition is that, similar to [2], doing so will provide more stable predictions as training progresses.

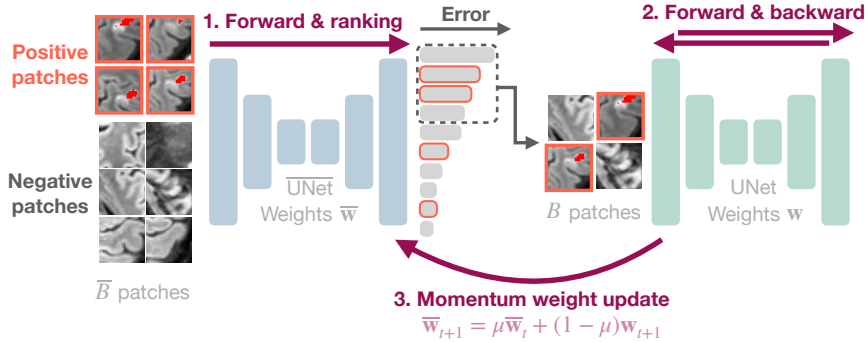


Fig. 1. Illustration of our training strategy. \bar{B} patches are fed to a first 3D U-Net and the segmentation errors are computed for each patch. The patches are ranked according to their errors, and the top B patches are selected to perform a training step with a second 3D U-Net. The weights of the first 3D U-Net are momentum-updated with the weights of the second 3D U-Net.

2 Method description

Preprocessing We resample each FLAIR to a voxel size of 0.5mm and apply z -normalization to each FLAIR individually. As the two consecutive FLAIRs of a patient have been aligned in the halfway space using a rigid transformation by the challenge providers, our method starts by concatenating them along the channel dimension, resulting in a tensor of shape $2 \times D \times H \times W$. This tensor is then subdivided into patches of shape $2 \times 32 \times 32 \times 32$, which are passed through a 3D U-Net to obtain the segmentation.

Model Our model is a simple 3D U-Net, described by the following equations:

$$\begin{aligned}
 B(n) &:= 2 \times \{3DConvolution(n) \rightarrow \text{Group Normalization} \rightarrow \text{ReLU}\} \\
 \text{3D U-Net} &:= B(16)\downarrow \rightarrow B(32)\downarrow \rightarrow B(64) \rightarrow \uparrow B(32) \rightarrow \uparrow B(16) \rightarrow \text{Conv}(1)
 \end{aligned}$$

where the numbers in the parentheses are the number of filters, \downarrow indicates max pooling and \uparrow indicates bilinear upsampling. The model is trained on patches of size 32. For inference, we split the image into a grid of patches of size 32, with a stride of 24. This means that the patches have an overlap of 8 pixels. In these overlapping regions, we average all predictions.

Training As the images contain very few positive voxels, we do not sample the patches uniformly during training. One common strategy is to over-sample patches containing positive regions with a constant ratio. However, this ratio must be fine-tuned by hand. If it is too high, it can result in many false positives.

Instead, our method uses a 3D U-Net with momentum weight updates to perform hard example mining. A training iteration consists of 3 steps, illustrated

in Figure 1 and formalized by Algorithm 1. In the first step, we select a batch of $\bar{B} = 128$ patches which contains 30% of positive patches and 70% of uniformly sampled patches (i.e., mostly negatives due to the class imbalance). We then pass this batch through a first 3D U-Net, denoted by $\overline{\text{UNet}}$, to obtain a prediction for each element of the batch and compute the segmentation errors with respect to the ground truth. Secondly, we select the $B = 32$ patches with highest error and perform a training step on them with a second 3D U-Net, denoted UNet . Lastly, we perform a momentum update of the weights of the first 3D U-Net with the second 3D U-Net. The use of momentum ensures that the predictions given by the first 3D U-Net do not fluctuate too much during training and provide reliable samples for online hard example mining.

Algorithm 1: Training procedure for OHEM

\bar{B} : batch size for hard example mining
 $B < \bar{B}$: batch size for gradient descent
 $\overline{\text{UNet}}$: momentum-updated 3D U-Net
 \mathbf{w} : weights of UNet
 $\overline{\mathbf{w}}$: weights of $\overline{\text{UNet}}$
 Initialization: $\overline{\mathbf{w}}_0 = \mathbf{w}_0$
 μ : momentum coefficient
while training **do**
 $\mathbf{x}_1 \dots \mathbf{x}_{\bar{B}} = \text{sample_patches}(\text{positive_ratio}=0.3)$
 $e_1 \dots e_{\bar{B}} = \text{error}(\overline{\text{UNet}}(\mathbf{x}_1 \dots \mathbf{x}_{\bar{B}}), \mathbf{y}_1 \dots \mathbf{y}_{\bar{B}})$
 $i_1 \dots i_B = \text{top}_B(e_1 \dots e_{\bar{B}})$
 $\ell_1 \dots \ell_B = \text{loss}(\text{UNet}(\mathbf{x}_{i_1} \dots \mathbf{x}_{i_B}), \mathbf{y}_{i_1} \dots \mathbf{y}_{i_B})$
 $\mathbf{w}_{t+1} = \text{Adam}(\mathbf{w}_t, \ell_1 \dots \ell_B)$
 $\overline{\mathbf{w}}_{t+1} = \mu \overline{\mathbf{w}}_t + (1 - \mu) \mathbf{w}_{t+1}$
end

3 Experiments and results

We optimize each network for 3 hours on one NVIDIA Tesla P100 graphic card using Adam [3]. Note that for OHEM, the duration of one iteration is roughly 2 times longer. In the end, 3 hours of training correspond to about 30k iterations with OHEM and 64k without. The initial learning rate is set to 10^{-3} and decayed to 10^{-4} and 10^{-5} after respectively 50% and 80% of the training time.

We split the dataset into 30 patients for training, and 10 for validation. We did not have time to perform a k -fold evaluation before the submission.

Figure 2 presents the evolution of the Dice score as a function of training time. Unfortunately, our experiments indicate that OHEM is not effective, as the results are comparable or worse than those obtained by oversampling positive patches.

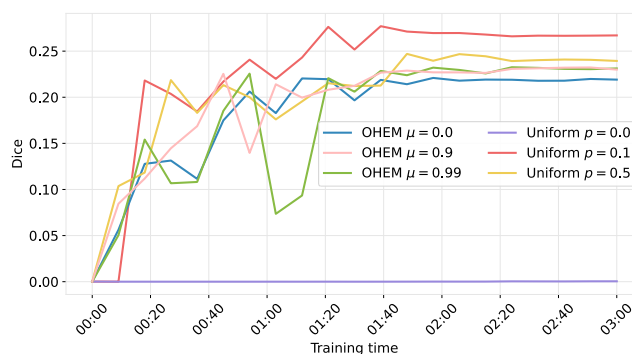


Fig. 2. Dice on the validation set during training. "Uniform" means that positive patches are sampled with probability p , and that with probability $1 - p$ any other patch (not necessarily negative) is selected. In the end, uniform sampling with $p > 0$ ended up performing best, and $\mu > 0$ seems to be beneficial when using OHEM.

4 Conclusion

This paper presented our contribution to the MSSEG-II MICCAI 2021 challenge. In order to deal with the strong imbalance between negative and positive voxels, we proposed a novel modification to online hard example mining (OHEM), where we use an exponential moving average of a 3D U-Net to mine hard examples. Unfortunately, in our experiments, it did not provide an improvement compared to a simple oversampling. We hypothesize that this may be due to the use of the Dice loss. Future work shall focus on using alternative losses to fix this issue.

Acknowledgments

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute), reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6), and reference number ANR-19-P3IA-0002 (3IA Côte d'Azur).

References

1. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: MICCAI 2016
2. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.B.: Momentum contrast for unsupervised visual representation learning. In: CVPR 2020
3. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) ICLR 2015
4. Shrivastava, A., Gupta, A., Girshick, R.B.: Training region-based object detectors with online hard example mining. In: CVPR 2016