Image Quality Data Augmentation for New MS Lesion Segmentation

Reda Abdellah Kamraoui[1], Vinh-Thong Ta[1], José V Manjon[3], and Pierrick Coupé[1]

[1] Univ. Bordeaux, Bordeaux INP, CNRS, LaBRI, UMR5800, PICTURA, F-33400 Talence, France
[2] ITACA, Universitat Politècnica de València, 46022 Valencia, Spain

Abstract. Recently, deep learning methods achieved remarkable results in highly controlled medical imaging studies. However, these methods fail to generalize to new images coming from different sources than training data. In this work, we use image quality data augmentation for training a robust new multiple sclerosis lesion segmentation model.

Keywords: Deep Learning Generalization · New lesion segmentation · Data Augmentation.

1 Introduction

The detection of new lesions is an important bio-marker in Multiple Sclerosis (MS) that allows clinicians to adapt the patient treatment and assess the evolution of the disease.

Similarly to other medical imaging tasks, Deep Learning (DL) methods will undeniably be extensively explored to automate the segmentation of new MS lesions from two time-points FLAIR. These techniques have already shown remarkable performance for MS lesion segmentation in controlled evaluation conditions (see [2, 3]). However, clinical use of DL based methods is still limited mainly because of their poor generalization on new data coming from medical sites that have not been covered during training [8, 1, 9].

Recent works [11] showed that applying extensive data augmentation during training enhances the robustness of the method. Thus, we use the Image Quality Data Augmentation (IQDA) [6] in order to improve the generalization ability of our proposed new lesion segmentation model.

2 Method and Material

2.1 Image Quality Data Augmentation

The quality of the MRI greatly varies between datasets. In fact, the quality of the images depends on several factors such as signal to noise ratio, contrast to

noise ratio, resolution or slice thickness. To address this issue, we use a data augmentation strategy, originally proposed in [6], which considers image quality disparity. During training, we simulate "on the fly" altered versions of 3D patches. We randomly introduce at each iteration either blur, edge enhancement, or axial subsampling distortion (2D FLAIR are usually acquired along the axial direction). For the blur, a gaussian kernel is used with a randomly selected standard deviation ranging between $[0.5, 1.75]$. For edge enhancement, we use unsharp masking with the inverse of the blur filter. For axial subsampling distortion, we simulate acquisition artifacts that can result from the varying slice thickness. We use a uniform filter (a.k.a mean filter) on the axial direction with a size of $[1 \times 1 \times sz]$ where $sz \in 2, 3, 4$. Ground truth is kept the same as the original version. This process reduces the domain bias when learning to extract relevant features caused by data variability.

2.2 New lesions Segmentation

As input, the network receives a concatenation of FLAIR patches from the two time-points augmented with IQDA. The model output predicts the new lesion mask. The Dice similarity loss is used during training to quantify similarity between the ground truth and the predicted mask.

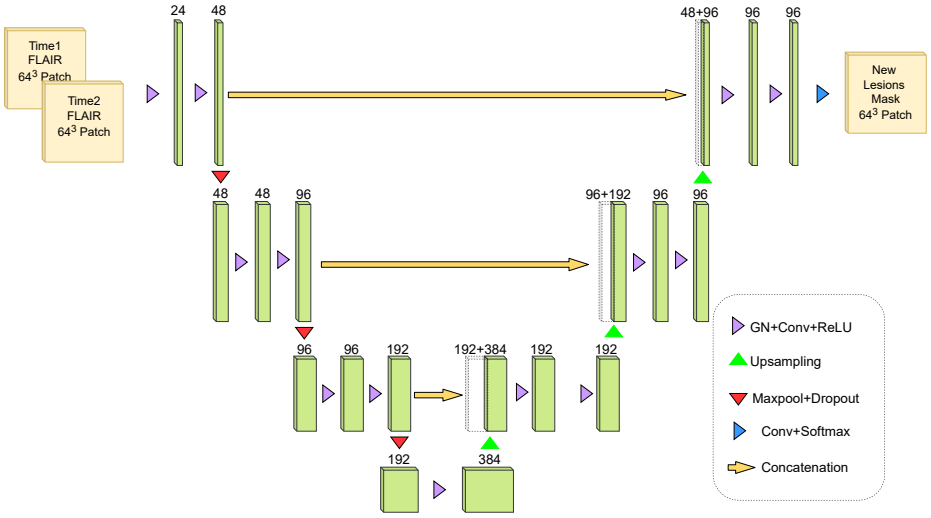

Fig. 1. Illustration of the used U-Net architecture. Each block is composed of group normalization (GN), Convolution (Conv) and Rectified Linear Unit (ReLU) activation.

As shown in Fig. 1, the network architecture is a 3D U-Net composed of a downsampling part and an upsampling one, linked with one another by skip connections at the multiple scales. Dropout with 0.5 rate is used after max-pooling

layers to prevent the overfitting of our model to the training data. We used Group Normalization (GN) [10] with 8 groups before each convolution which is adapted for small batch size. We have chosen Rectified Linear Units (ReLu) to introduce non linearity after convolution layers. The model is optimized with Adam [7] using a learning rate of 0.0001 and a momentum of 0.9.

2.3 Data

The dataset provided by the MICCAI 2021 - Longitudinal Multiple Sclerosis Lesion Segmentation Challenge [4] was used to train our method. The challenge dataset features a total of 100 MS patients. For each patient two 3D FLAIR sequence time-point have been acquired spaced apart by a 1 to 3 years period. Dataset has been split into 40 patients for training and 60 patients for testing. A total of 15 different MRI scanners were used. However, all images from GE scanners have been reserved only to the testing set to see the generalizability of the algorithms. Reference segmentation on these data was defined by a consensus of 4 expert neuroradiologists.

For preprocessing, our strategy used the docker [5] built with the Anima scripts proposed by the challenge organizers. It includes bias correction, denoising and skull stripping.

References

1. Bron, E.E., Klein, S., Papma, J.M., Jiskoot, L.C., Venkatraghavan, V., Linders, J., Aalten, P., De Deyn, P.P., Biessels, G.J., Claassen, J.A., et al.: Cross-cohort generalizability of deep and conventional machine learning for mri-based diagnosis and prediction of alzheimer's disease. arXiv preprint arXiv:2012.08769 (2020)
2. Carass, A., Roy, S., Jog, A., Cuzzocreo, J.L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C.H., et al.: Longitudinal multiple sclerosis lesion segmentation: resource and challenge. NeuroImage **148**, 77–102 (2017)
3. Commowick, O., Cervenansky, F., Ameli, R.: Msseg challenge proceedings: Multiple sclerosis lesions segmentation challenge using a data management and processing infrastructure (2016)
4. the challenge dataset: https://portal.fli-iam.irisa.fr/msseg-2/data/
5. the pre-processing docker: https://github.com/Inria-Empenn/lesion-segmentation-challenge-miccai21/
6. Kamraoui, R.A., Ta, V.T., Tourdias, T., Mansencal, B., Manjon, J.V., Coupé, P.: Towards broader generalization of deep learning methods for multiple sclerosis lesion segmentation. arXiv preprint arXiv:2012.07950 (2020)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Mårtensson, G., Ferreira, D., Granberg, T., Cavallin, L., Oppedal, K., Padovani, A., Rektorova, I., Bonanni, L., Pardini, M., Kramberger, M.G., et al.: The reliability of a deep learning model in clinical out-of-distribution mri data: a multicohort study. Medical Image Analysis p. 101714 (2020)
9. Omoumi, P., Ducarouge, A., Tournier, A., Harvey, H., Kahn, C.E., Louvet-de Verchère, F., Dos Santos, D.P., Kober, T., Richiardi, J.: To buy or not to

buy—evaluating commercial ai solutions in radiology (the eclair guidelines). European Radiology pp. 1–11 (2021)
10. Wu, Y., He, K.: Group normalization. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 3–19 (2018)
11. Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B.J., Roth, H., Myronenko, A., Xu, D., et al.: Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Transactions on Medical Imaging (2020)